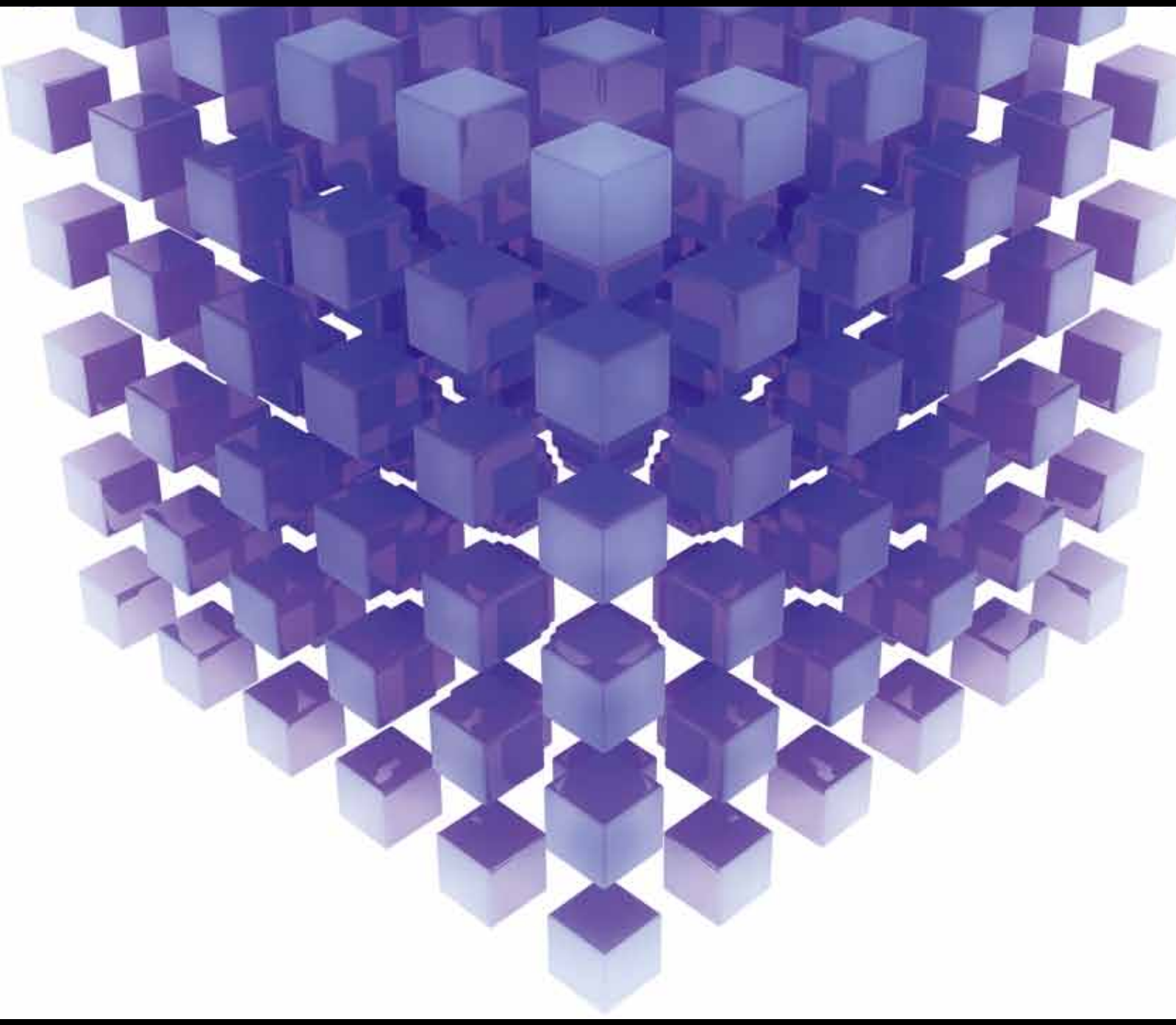
Mathematical Problems in Engineering

Information and Modeling in Complexity 2013

Guest Editors: Carlo Cattani, Shengyong Chen, and Gani Aldashev



Information and Modeling in Complexity 2013

Mathematical Problems in Engineering

Information and Modeling in Complexity 2013

Guest Editors: Carlo Cattani, Shengyong Chen,
and Gani Aldashev

Editorial Board

Contents

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 797368, 2 pages
http://dx.doi.org/10.1155/2013/797368

Editorial
Information and Modeling in Complexity 2013

Carlo Cattani,[1] Shengyong Chen,[2] and Gani Aldashev[3]

[1] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*
[2] *School of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China*
[3] *Department of Economics, University of Namur, Rempart de la Vierge 8, 5000 Namur, Belgium*

Correspondence should be addressed to Carlo Cattani; ccattani@unisa.it

Received 5 August 2013; Accepted 5 August 2013

Recent research in engineering systems builds theories represented by mathematical models that aim at understanding fundamental questions of spatial structure, self-organization, environmental interaction, behavior, and development. Thus, researchers and engineers are increasingly facing the challenge of dealing with complexity in practical applications that are based on efficient mathematical models.

The topics that deal with these issues can be divided into two parts: (i) development of general mathematical methods/models and (ii) specific applications in particular domains. The focus of this issue on topics of information and modeling complexity is both on the mathematical models and the engineering applications in complex systems. Moreover, we are interested in both new theoretical developments and the studies of practical implementation concerning modeling, complexity, fractals, statistics, and signal extraction and transformation.

Complex systems, such as living systems, are often maintained by information flows. Signals extract such information from the complex phenomena being measured. Such signals are typically a time series having both a regular and a random component. Solutions of the arising mathematical problems attempt to map general principles for modeling how the complex systems operate. This area is a burgeoning field of research due to both the practical significance of the applications and the general scientific importance.

Further to the 63 submissions in 2012, the special issue receives 94 new submissions in 2013. After a rigorous review process, 36 papers, that is, a 38.3% acceptance rate, which represent a good panel in the related new development, are selected in the journal, with the topics being connected with modeling of complex systems, computational issues in dynamical systems, signal processing of complex information, complexity and parameters, entropy, multifractality, Hurst exponent, correlation, power spectrum, and practical applications.

In the category of modeling, M. Scalia et al. study the Johnson noise in biological matter, L. Wu et al. propose a contextual hierarchical part-driven conditional random field model for object category detection, and M. Li and W. Zhao investigate on 1 over f noise. L. Zhu et al. propose a bilayer resource model for cloud manufacturing services. L. Wu et al. study the inverse 1-median problem on tree networks with variable real edge lengths. L. Luo et al. present maximum variance hashing via column generation. J. Hu et al. show the adaptive multilevel kernel machine for scenic classification. X. Duan et al. propose a Shannon-Runge-Kutta-Gill method for convection-diffusion equations. H. Ma et al. present strong list edge coloring of subcubic graphs. J. Leng et al. construct bivariate nonseparable compactly supported orthogonal wavelets. Z.-W. Li et al. propose a dynamically weighted clique evolution model in clique networks. B. Nouhi et al. study chaotic charged system search with a feasible-based method for constraint optimization problems. L. Huang et al. propose a dynamics-based stranded-crowd model for evacuation in building bottlenecks. E. G. Bakhoum and C. Toma make modeling transitions in complex systems by multiplicative effect of temporal patterns extracted from signal flows.

On computational issues, M. Aktas and E. Agcakoca study the impact of the HMCFRP ratio on the strengthening of steel composite I-beams. M. Li represents a smoothed spectrum estimate with the Cauchy integral. S. Liu et al. show approximation slack scaling for structural support

vector machines in scene depth analysis. C. Biscarini et al. make detailed simulation of complex hydraulic problems with macroscopic and mesoscopic mathematical methods. D. Sendrescu proposes parameter identification of anaerobic wastewater treatment bioprocesses using particle swarm optimization. M. Chen et al. construct a cloud computing platform for an online model library system. Z. Lu et al. study normality of Ethernet traffic at large time scales. X. Huang et al. propose a marching cubes algorithm for fast modeling of human face by incremental data fusion. P.-G. Popescu et al. present new inequalities between information measures of network information content. X. Xu et al. give a hybrid discrete differential evolution algorithm for lot splitting with capacity constraints in flexible job scheduling. H. Pan et al. review recent modeling for deformable body and motion analysis. K. Guo et al. propose a differential and statistical approach to partial model matching. J. Zhang et al. propose a multiobjective quantum evolutionary algorithm for the vehicle routing problem with customer satisfaction.

On practical applications, Z. Lu et al. give stationarity testing of accumulated Ethernet traffic. Y. Niu et al. work on airborne infrared and visible image fusion for target perception based on target region segmentation and discrete wavelet transform. F. Su et al. identify adaptive color features in image for object tracking. D. Zhang et al. show human model adaptation for multiview markerless motion capture. Y.-S. Juang et al. make fast texture synthesis in adaptive wavelet packet trees. F.-H. Cheng and T.-Y. Sung merge static and dynamic depth cues with optical-flow recovery for creating stereo videos. A. Marucci et al. study energy efficiency of a greenhouse for the conservation of forestry biodiversity. H. Cai et al. propose entropy-based maximally stable extremal regions for robust feature detection. C. B. Guure and N. A. Ibrahim use Bayesian analysis of the survival function and failure rate of Weibull distribution with censored data.

Nevertheless, this special issue is not intended to constitute a complete compendium on the topic, but it does offer the reader an opportunity to appreciate the possibilities and the progress in the research area. It can be seen that although some models and solutions become available, most problems remain open and research is highly active in this field. We are aware of the present and upcoming research progress in the field.

Acknowledgments

As guest editors of this special issue, we would like to thank all the authors for their valuable contributions and the referees for their thorough reviews and helpful comments on the papers.

Carlo Cattani
Shengyong Chen
Gani Aldashev

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 103748, 8 pages
http://dx.doi.org/10.1155/2013/103748

Research Article

Parameter Identification of Anaerobic Wastewater Treatment Bioprocesses Using Particle Swarm Optimization

Dorin Sendrescu

Department of Automatic Control, University of Craiova, A.I. Cuza 13, 200585 Craiova, Romania

Correspondence should be addressed to Dorin Sendrescu; dorins@automation.ucv.ro

Received 25 January 2013; Accepted 30 June 2013

Academic Editor: Carlo Cattani

This paper deals with the offline parameters identification for a class of wastewater treatment bioprocesses using particle swarm optimization (PSO) techniques. Particle swarm optimization is a relatively new heuristic method that has produced promising results for solving complex optimization problems. In this paper one uses some variants of the PSO algorithm for parameter estimation of an anaerobic wastewater treatment process that is a complex biotechnological system. The identification scheme is based on a multimodal numerical optimization problem with high dimension. The performances of the method are analyzed by numerical simulations.

1. Introduction

It is well known that the biotechnology is one of the fields that over the last decades have a high development. Therefore, due to its advantages, the control of industrial bioprocesses has been an important practical problem attracting wide attention. Biotechnology applications can be found especially in agriculture, in food industry, in medicine and pharmaceutical processes, in waste treatment processes, and so forth. A frequent and important challenge in control of such living processes is finding an accurate model of the system. The bioprocesses are highly nonlinear, and their kinetic parameters are usually badly or inadequately known [1]. This problem becomes of great importance in complex systems where critical instability of the process must be avoided. Parameters characterizing the internal behavior of biotechnological systems are usually not directly accessible to measurement. Their measurement is usually approached indirectly as a parameter estimation problem [2]. In this paper a dynamic model describing the internal structure of the system is formulated, and an algorithm based on PSO for parameter estimation is designed.

In recent years, a progress has been made in the area of continuous-time system identification [3]. Even if the most physical systems are naturally continuous, a much more attention has been paid to parameter estimation of discrete-time systems, mainly because they are better suited for numerical implementations. Continuous-time identification makes possible a more direct link to the physical properties and operation of the underlying systems and the direct estimation of physical parameters which have a clear significance. The most common approach for parameter estimation of linear or nonlinear systems is the use of prediction-error identification methods (PEM) [4]. In this category falls the well-known least squares methods or the maximum likelihood methods. In this approach, identification consists in minimization of a scalar-valued function of the model parameter. In general, this function cannot be minimized by analytical methods so the solution has to be found by iterative, numerical techniques. There is an extensive literature on such numerical problems. In classical approach the most used procedures are the quasi-Newton methods and interior point algorithms. The main drawback of these nonlinear parameter optimization techniques is that they are often unreliable; for example, they give no guarantee of converging to a true minimum. The increasing computational power of personal computers and microcontrollers allowed the implementation of several optimization algorithms inspired from natural phenomena. Examples of these algorithms include the Simulated Annealing [5], Genetics Algorithms (GA) [6], or Ant Colony

Optimization [7] algorithms. Particle Swarm Optimization (PSO) [8] is among these nature inspired algorithms. It is inspired by the ability of birds flocking to find food that they have no previous knowledge of its location. Every member of the swarm is affected by its own experience and its neighbors' experiences. Although the idea behind PSO is simple and can be implemented by two lines of programming code, the emergent behavior is complex and hard to completely understand [9, 10].

The most important approaches for the yield and kinetic coefficients estimation of biotechnological systems make use of the state transformations based on the general structure [11]. In this paper we propose an identification method based on particle swarm optimization techniques for these classes of biotechnological systems considering that the unknown parameters can appear in rational relations with measured variables. The paper is organized in the following way. The nonlinear dynamical model of an anaerobic wastewater treatment bioprocess is given in Section 2. Section 3 presents the identification algorithm using the particle swarm optimization techniques. Some numerical simulations are presented in Section 4 and conclusions in Section 5.

2. Nonlinear Dynamical Model of Anaerobic Wastewater Treatment Bioprocesses

A process that takes place in a bioreactor can be described as a set of m biochemical reactions involving n components (with $n \geq m$). The global dynamics can be represented by the following dynamical state-space model [1]:

$$\frac{d\xi}{dt} = K \cdot \varphi(\xi, t) - D\xi + F - Q, \tag{1}$$

where $\xi \in \Re^{n\times 1}$.

This model describes the behavior of an entire class of biotechnological processes and is referred to as the general dynamical state-space model of this class of bioprocesses [9]. In (1), the term $K \cdot \varphi(\xi, t)$ is the rate of consumption and/or production of the components in the reactor; that is, the reaction kinetics and the term $-D\xi + F - Q$ represents the exchange with the environment. The strongly nonlinear character of the model (1) is given by the reaction kinetics. In many situations, the yield coefficients, the structure, and the parameters of the reaction rates are partially known or unknown. Many of the evolved control methods for these kinds of systems—like model predictive control and robust or adaptive control—are based on good initial estimates of the yield and kinetic parameters.

Anaerobic digestion is a multistage biological wastewater treatment process whereby bacteria, in the absence of oxygen, decompose organic matter to carbon dioxide CO_2, methane CH_4, and water [12]. Four metabolic paths can be identified in this process: two for acidogenesis and two for methanation (see Figure 1). In the first acidogenic path, glucose (or another complex substrate) from the wastewater is decomposed into volatile fatty acids (acetates, propionic acid), hydrogen H_2, and inorganic carbon by acidogenic bacteria.

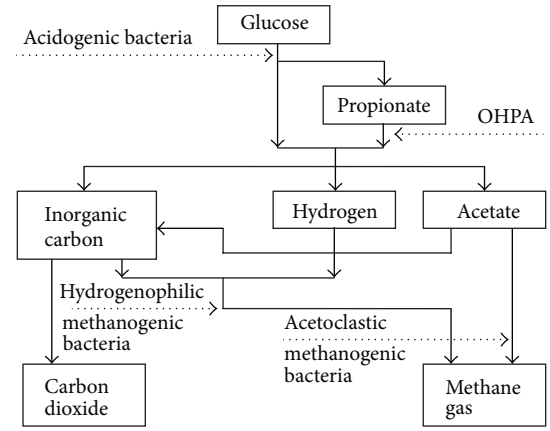


Figure 1: A schematic view of a typical anaerobic digestion process.

In the second acidogenic path, Obligate Hydrogen Producing Acetogens (OHPA) decompose propionate into acetate, H_2, and inorganic carbon. In the first methanation path, acetate is transformed into CH_4 and CO_2 by acetoclastic methanogenic bacteria, while, in the second methanation path, H_2 combines inorganic carbon to produce CH_4 under the action of hydrogenophilic methanogenic bacteria. The reaction scheme of this complex bioprocess involves 4 reactions and 10 components.

Since the anaerobic digestion is a complex bioprocess, this dynamical model being described by ten differential equations, for control purpose an appropriately reduced order model can be used. Using the singular perturbation method, the following reduced order dynamical model can be obtained:

$$\dot{X}_1 = \varphi_1 - DX_1, \tag{2}$$

$$\dot{S}_1 = -k_1\varphi_1 - DS_1 + DS_{\text{in}}, \tag{3}$$

$$\dot{X}_2 = \varphi_2 - DX_2, \tag{4}$$

$$\dot{S}_2 = k_3\varphi_1 - k_2\varphi_2 - DS_2, \tag{5}$$

$$\dot{S}_5 = k_4\varphi_1 + k_5\varphi_2 - DS_5 - Q_{CO_2}, \tag{6}$$

$$\dot{P} = k_6\varphi_2 + k_7\varphi_1 - DP - Q_P, \tag{7}$$

where X_1, X_2 are acidogenic and acetoclastic methanogenic bacteria, respectively, and S_1, S_2, S_5 are glucose, acetate and inorganic carbon, respectively, and P is methane; φ_1, φ_2 are the rates of first acidogenic reaction and methanation reaction respectively, $Q_P = c_pP$ with $c_p > 0$ and Q_{CO_2} represent gaseous outflow rates of CH_4 and CO_2, respectively, S_{in} is the influent substrate concentration, D the dilution rate, and k_i $(i = 0, \ldots, 7)$ are yield coefficients. Each reaction rate is a growth rate and may be written as $\varphi_i = \mu_i X_i$, $i = 1, 2$, where μ_i, $i = 1, 2$, is the specific growth rate of reaction i.

Defining the state vector as $\xi = [X_1 \ S_1 \ X_2 \ S_2 \ S_5 \ P]^T$, the model (2)–(7) can be written in matrix form as

$$\begin{bmatrix} \dot{X}_1 \\ \dot{S}_1 \\ \dot{X}_2 \\ \dot{S}_2 \\ \dot{S}_5 \\ \dot{P} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -k_1 & 0 \\ 0 & 1 \\ k_2 & -k_3 \\ k_4 & k_5 \\ k_6 & k_7 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} - D \begin{bmatrix} X_1 \\ S_1 \\ X_2 \\ S_2 \\ S_5 \\ P \end{bmatrix} + \begin{bmatrix} 0 \\ DS_{\text{in}} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ Q_{CO_2} \\ Q_P \end{bmatrix}, \tag{8}$$

or in compact form as

$$\dot{\xi} = K\varphi(\xi) - D\xi + F - Q, \tag{9}$$

where $F = [0 \ DS_{\text{in}} \ 0 \ 0 \ 0 \ 0]^T$ is the vector of inflow rates, $Q = [0 \ 0 \ 0 \ 0 \ Q_{CO_2} \ Q_P]^T$ is the vector of gaseous outflow rates, $\varphi = [\varphi_1 \ \varphi_2]^T$ is the vector of reaction rates, which can be written as $\varphi(\xi) = G(\xi)\alpha(\xi)$, with $G(\xi)$ being a diagonal matrix whose entries are products of the component concentrations involved in each reaction and $\alpha = [\alpha_1 \ \alpha_2]^T$ the vector of specific reaction rates, and K is the yield coefficient's matrix. The matrices K and G have the following structure:

$$K = \begin{bmatrix} 1 & -k_1 & 0 & k_2 & k_4 & k_6 \\ 0 & 0 & 1 & -k_3 & k_5 & k_7 \end{bmatrix}^T, \qquad G = \begin{bmatrix} X_1 S_1 & 0 \\ 0 & X_2 S_2 \end{bmatrix}. \tag{10}$$

The most difficult task for the construction of the dynamical model is the modeling of the reaction kinetics [13]. The form of kinetics is complex, nonlinear, and in many cases unknown. In our study one considers that reaction rates are given by the Monod law

$$\varphi_1(\xi) = \mu_1^* \frac{S_1 \cdot X_1}{K_{M_1} + S_1}, \tag{11}$$

and the Haldane kinetic model

$$\varphi_2(\xi) = \mu_2^* \frac{S_2 \cdot X_2}{K_{M_2} + S_2 + S_2^2/K_i}, \tag{12}$$

where K_{M_1}, K_{M_2} are Michaelis-Menten constants, μ_1^*, μ_2^* represent specific growth rate coefficients, and K_i is the inhibition constant.

For simplicity, we will denote the unknown plant parameters by the vector

$$\theta = [\theta_1 \ \theta_2 \cdots \theta_{12}]^T, \tag{13}$$

where

$$\begin{gathered} \theta_1 = k_1; \quad \theta_2 = k_2; \quad \theta_3 = k_3; \\ \theta_4 = k_4; \quad \theta_5 = k_5 \quad \theta_6 = k_6 \quad \theta_7 = k_7 \\ \theta_8 = \mu_1^*; \quad \theta_9 = \mu_2^*; \quad \theta_{10} = K_{M_1}; \\ \theta_{11} = K_{M_2}; \quad \theta_{12} = K_i. \end{gathered} \tag{14}$$

3. Parameter Estimation Using PSO

At the beginning of parameter estimation, the input and output data are known and the real system parameters are assumed as unknown. The identification problem is formulated in terms of an optimization problem in which the error between an actual physical measured response of the system and the simulated response of a parameterized model is minimized. The estimation of the system parameters is achieved as a result of minimizing the error function by the PSO algorithm.

3.1. Problem Statement. Consider the following n-dimensional nonlinear system:

$$\frac{d\xi(t)}{dt} = f(\xi, t; \theta), \tag{15}$$

where $\xi \in R^n$ is the state vector, $\theta \in R^m$ is the unknown parameters vector, and f is a given nonlinear vector function.

To estimate the unknown parameters in (15), a parameter identification system is defined as follows:

$$\frac{d\hat{\xi}(t)}{dt} = f\left(\hat{\xi}, t; \hat{\theta}\right), \tag{16}$$

where $\hat{\xi} \in R^n$ is the estimated state vector and $\hat{\theta} \in R^m$ is the estimated parameters vector.

The objective function defined as the mean squared errors between real and estimated responses for a number N of given samples is considered as fitness of estimated model parameters:

$$V = \frac{1}{N+D} \sum_{j=1}^{D} \sum_{k=1}^{N} \left(\xi_j^k - \hat{\xi}_j^k\right)^2, \tag{17}$$

where D is the number of measurable states, N is the data length used for parameter identification, whereas ξ_j^k and $\hat{\xi}_j^k$ are the real and estimated values of state j at time k, respectively.

This objective function is a difficult function to minimize because there are many local minima, and the global minimum has a very narrow domain of attraction. Our goal is to determine the system parameters, using particle swarm optimization algorithms in such a way that the value of V is minimized, approaching zero as much as possible.

3.2. Overview of Basic PSO Algorithms. During the last decade, PSO algorithms have gained much attention and wide

applications in different fields due to their effectiveness in performing difficult optimization issues, as well as simplicity of implementation and ability to fast converge to a reasonably good solution [14–16]. PSO is a population-based heuristic global optimization technique, first introduced by Kennedy and Eberhart [8] and referred to as a swarm-intelligence technique. It is motivated from the simulation of social behavior of animals such as bird flocking, fish schooling, and swarm. In this algorithm, the population is called a swarm, and the trajectory of each particle in the search space is controlled through the medium of a term called "velocity," according to its own flying experience and swarm experience in the search space. Mathematical description of basic PSO and some important variants is presented in the following.

Candidate solutions of a population called particles coexist and evolve simultaneously based on knowledge sharing with neighboring particles. Each particle represents a potential solution to the optimization problem, and it has a fitness value decided by optimal function. Supposing that search space is D-dimensional, each individual is treated as a particle in the D-dimensional search space. The position and rate of position change for ith particle can be represented by D-dimensional vector, $x_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$ and $v_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$, respectively. The best position previously visited by the ith particle is recorded and represented as $p_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$, called *pbest*. The swarm best position previously visited by all the particles in the population is represented as $p_g = (p_{g1}, p_{g2}, \ldots, p_{gD})$, called *gbest*. Then particles search their best position, which are guided by swarm information p_g and their own information p_i. Each particle modifies its velocity to find a better solution (position) by applying its own flying experience (i.e., memory of the best position found in earlier flights) and the experience of neighboring particles (i.e., the best solution found by the population). Each particle position is evaluated by using fitness function and updates its position and velocity according to the following equations:

$$v_i^{k+1} = \omega \cdot v_i^k + c_1 r_1 \left(pbest_i^k - x_i^t\right) + c_2 r_2 \left(gbest_i^k - x_i^k\right),$$
$$x_i^{k+1} = x_i^k + v_i^{k+1}, \tag{18}$$

where k is iteration number, ω is inertia weight, c_1 and c_2 are two acceleration coefficients regulating the relative velocity toward local and global best positions, and r_1 and r_2 are two random numbers from interval $[0, 1]$. Many effects have been made over the last decade to determinate the inertia weight. Various studies has shown that under certain conditions convergence is guaranteed to a stable equilibrium point. These conditions include $\omega > (c_1+c_2)/2-1$ and $0 < \omega < 1$. The technique originally proposed was to bound velocities so that each component of v_i is kept within the range $[-V_{\max}, +V_{\max}]$.

Unfortunately, this simple form of PSO suffers from the premature convergence problem, which is particularly true in complex problems since the interacted information among particles in PSO is too simple to encourage a global search. Many efforts have been made to avoid the premature convergence. One solution is the use of a constriction factor to insure convergence of the PSO, introduced in [17]. Thus, the expression for velocity has been modified as

$$v_i^{k+1} = h \cdot \left[v_i^k + c_1 r_1 \left(pbest_i^k - x_i^t\right) + c_2 r_2 \left(gbest_i^k - x_i^k\right)\right],$$
$$x_i^{k+1} = x_i^k + v_i^{k+1}, \tag{19}$$

where h represents the constriction factor and is defined as

$$h = \frac{2}{\left|2 - \alpha - \sqrt{\alpha^2 - 4\alpha}\right|}, \tag{20}$$

$$\alpha = c_1 + c_2 > 4. \tag{21}$$

In this variant of the PSO algorithm, h controls the magnitude of the particle velocity and can be seen as a dampening factor. It provides the algorithm with two important features. First, it usually leads to faster convergence than standard PSO. Second, the swarm maintains the ability to perform wide movements in the search space, even if convergence is already advanced, but a new optimum is found. Therefore, the constriction PSO has the potential to avoid being trapped in local optima while possessing a fast convergence. It was shown to have superior performance compared to a standard PSO.

It is shown that a larger inertia weight tends to facilitate the global exploration and a smaller inertia weight achieves the local exploration to fine-tune the current search area [18]. The best performance could be obtained by initially setting ω to some relatively high value (e.g., 0.9), which corresponds to a system where particles perform extensive exploration, and gradually reducing ω to a much lower value (e.g., 0.4), where the system would be more dissipative and exploitative and would be better at homing into local optima. In [19], a linearly decreased inertia weight ω over time is proposed, where ω is given by the following equation:

$$\omega = \left(\omega_i - \omega_f\right) \cdot \frac{k_{\max} - k}{k_{\max}} + \omega_f, \tag{22}$$

where ω_i, ω_f are starting and final values of inertia weight, respectively; $k_{\max}$ is the maximum number of the iteration, and k is the current iteration number. It is generally taken that starting value $\omega_i = 0.9$ and final value $\omega_f = 0.4$ [20].

On the other hand, in [21] was introduced PSO with time-varying acceleration coefficients. The improvement has the same motivation and the similar techniques as the adaptation of inertia weight. In this case, the cognitive coefficient c_1 is decreased linearly and the social coefficient c_2 is increased linearly over time as follows:

$$c_1 = \left(c_{1f} - c_{1i}\right) \cdot \frac{k_{\max} - k}{k_{\max}} + c_{1i},$$
$$c_2 = \left(c_{2f} - c_{2i}\right) \cdot \frac{k_{\max} - k}{k_{\max}} + c_{2i}, \tag{23}$$

where c_{1i} and c_{2i} are the initial values of the acceleration coefficients c_1 and c_2; c_{1f} and c_{2f} are the final values of the acceleration coefficients c_1 and c_2, respectively. Usually, $c_{1i} = 2.5$, $c_{2i} = 0.5$, $c_{1f} = 0.5$, and $c_{2f} = 2.5$.

3.3. Identification Algorithm. Considering all the states of the nonlinear system (8) at the sampling moments $k * Ts$ (Ts = sampling period) known, the identification algorithm has the following steps.

Step 1. Initialize a population of particles with random positions and velocities on D dimensions in search space.

Step 2. For each particle, evaluate the desired optimization fitness function (17) in D variables.

Step 3. Compare particle's fitness evaluation with its *pbest*. If current value is better than *pbest*, then set *pbest* equal to the current value x_i in D-dimensional space.

Step 4. Identify the particle in swarm with the best success so far, and assign its index to the variable *gbest*.

Step 5. Change the velocity and position of the particle according to (19).

Step 6. *if* a criterion is met (usually a sufficiently good fitness or a maximum number of iterations)

then Stop;

else

go to Step 2.

3.4. "Classical" Identification Procedure. The "classical" approach for identification of yield coefficients is a two-step procedure under the assumption that full state measurements are available [1]. This method is based on a state transformation that allows reformulating the dynamical model into separate submodels. The first submodel depends only on the reaction structure and is independent of the kinetics. It can be linearly reparametrized and used for the identification of the yield coefficients by means of linear regressions, provided suitable identifiability conditions are satisfied. We present briefly this method for estimating the yield coefficients, and we use it for comparison in the next section.

The general dynamical model given in (2) represents a particular class of nonlinear state-space models. The nonlinearity lies in the reaction rates $\varphi_i(\xi)$ that are (nonlinear) functions of the state variables. These functions enter the model in the form $K\varphi_i(\xi)$ (where K is a constant matrix), which is a set of linear combinations of the same nonlinear functions $\varphi_i(\xi)$. This particular feature can be exploited to separate the nonlinear part from the linear part of the model by an adequate linear state transformation. More precisely, one chooses a nonsingular partition

$$\begin{pmatrix} K_a \\ K_b \end{pmatrix} = TK, \tag{24}$$

with $K_a \in R^{p\times m}$ of full row rank matrix (i.e., $p = \text{rank}(K)$), $K_b \in R^{(n-p)\times m}$, and T a permutation matrix. The induced partitions of the vectors ξ and u are

$$\begin{pmatrix} \xi_a \\ \xi_b \end{pmatrix} = T\xi, \qquad \begin{pmatrix} u_a \\ u_b \end{pmatrix} = Tu. \tag{25}$$

Model (2) is then partitioned into two submodels

$$\begin{aligned} \frac{d\xi_a}{dt} &= K \cdot \varphi(\xi_a, t) - D\xi_a + u_a, \\ \frac{d\xi_b}{dt} &= K \cdot \varphi(\xi_b, t) - D\xi_b + u_b. \end{aligned} \tag{26}$$

Then with the state transformation

$$\begin{aligned} \xi_a &= \xi_a, \\ z &= C\xi_a + \xi_b, \end{aligned} \tag{27}$$

one transforms the initial model into

$$\begin{aligned} \dot{\xi}_a &= K_a\varphi(\xi_a, z - C\xi_a) - D\xi_a + u_a, \\ \dot{z} &= -Dz + Cu_a + u_b, \end{aligned} \tag{28}$$

where the $(n - p) \times p$ matrix C is the unique solution of

$$CK_a + K_b = 0. \tag{29}$$

That is,

$$C = -K_bK_a^*, \tag{30}$$

where K_a^* is a generalized inverse or pseudoinverse of K_a. The subsystem (29) can be augmented with an equation derived from (8) as follows:

$$\begin{aligned} \dot{z} &= -Dz + Cu_a + u_b, \\ \xi_b &= z - C\xi_a. \end{aligned} \tag{31}$$

It can be considered as a linear time-varying (if d varies in the course of time) model with state z, input (ξ_a, u_a, and u_b), and output ξ_b. It is nonlinearly parametrized by the yield coefficients but linearly reparametrized by the nonzero entries of C. When data of the signals ξ_a, u_a, u_b, and ξ_b are available, the auxiliary model (31) can be used to identify the yield coefficients independently of the knowledge of the reaction rates. The model (31) can be used to perform the identification of the nonzero entries of C by a linear regression technique, with the yield coefficients k_i recovered afterwards from (29).

For the estimation of kinetic parameters the main approach is the use of a parameter observers (high gain observers, regressive parameter estimator, sliding mode observers, etc.). All these techniques have numerous tuning parameters and are difficult to implement.

4. Simulation Results and Discussion

The efficacy of our approach is shown by numerical simulations on an interval of 30 hours. The model given by relation (9) was integrated using a fourth-order Runge-Kutta routine with a sampling period of 1 minute and with initial conditions: IC = $[2.5 \;\; 6 \;\; 0.2 \;\; 10 \;\; 0.5 \;\; 0]$.

The influence of sampling period and type of the optimization algorithm and of noisy measurements are analyzed. To compare statistical performances of the different

Table 1: Influence of the sampling period.

Estimated parameter	"Real" value	Sampling period		
		1 min	10 min	30 min
θ_1	5.5	5.5009	5.4811	5.4798
θ_2	1	0.9982	1.0111	1.0637
θ_3	16	15.9110	15.9328	14.5629
θ_4	10	10.0016	9.9707	9.9571
θ_5	1.5	1.4938	1.4626	1.3078
θ_6	0.2	0.2003	0.2011	0.1873
θ_7	3	2.9823	2.9581	2.6947
θ_8	0.2	0.2002	0.1974	0.1949
θ_9	0.6	0.5994	0.6010	0.6188
θ_{10}	0.75	0.7539	0.6754	0.6483
θ_{11}	4	3.9731	3.9007	3.0265
θ_{12}	20	20.3973	19.4738	20.3947
NMSE	**0**	**3.0717*e* − 05**	**5.6623*e* − 04**	**1.2143*e* − 01**

Table 2: Influence of the algorithm type.

Estimated parameter	"Real" value	Type of optimization algorithm		
		PSO-1	PSO-2	PSO-3
θ_1	5.5	5.5613	5.5849	5.4338
θ_2	1	1.0405	1.0083	0.8406
θ_3	16	15.2050	14.9971	15.6475
θ_4	10	10.0848	10.1252	9.9468
θ_5	1.5	1.5721	1.6422	1.1698
θ_6	0.2	0.2022	0.2040	0.3005
θ_7	3	2.8006	2.7916	2.7569
θ_8	0.2	0.1990	0.1975	0.2286
θ_9	0.6	0.5993	0.6311	0.6594
θ_{10}	0.75	0.9991	1.1153	0.4728
θ_{11}	4	3.5221	3.3470	4.0707
θ_{12}	20	20.9257	20.5048	20.0971
NMSE	**0**	**3.0717*e* − 04**	**0.0283**	**0.3845**

approaches the empirical normalized mean square error (NMSE) was used, that is defined as

$$\text{NMSE} = \frac{1}{N}\sum_{j=1}^{N} \text{NMSE}\left(\widehat{\theta}_j\right), \tag{32}$$

with $\text{NMSE}(\widehat{\theta}_j) = ((\widehat{\theta}_j - \theta_j^*)/\theta_j^*)^2$, where N is the number of estimated parameters, $\widehat{\theta}_j$ is the jth element of the estimated parameter vector while the "$*$" superscript denotes the true value of the parameter.

In order to study the sensitivity of the estimation method to the sampling period and to the type of PSO algorithm, and to the noise, the following parameters were used.

Sampling period: $T_s \in \{1 \text{ min}, 10 \text{ min}, 30 \text{ min}\}$.

Types of the optimization algorithm are as follows:

PSO_1: algorithm based on relation (19) with h defined by relation (20), and $c_1 = c_2 = 2.1$;

Table 3: Influence of the noise level.

Estimated parameter	"Real" value	Noise level		
		50 dB	40 dB	30 dB
θ_1	5.5	5.4001	5.3697	5.3362
θ_2	1	1.0066	0.6628	0.4180
θ_3	16	15.5721	16.6508	16.3867
θ_4	10	9.8502	9.8463	9.7771
θ_5	1.5	1.1817	1.1428	0.9150
θ_6	0.2	0.1867	0.2769	0.2711
θ_7	3	2.9626	2.9178	3.1112
θ_8	0.2	0.2042	0.2350	0.3020
θ_9	0.6	0.6030	0.6254	0.5744
θ_{10}	0.75	0.3306	0.2379	0.0058
θ_{11}	4	3.7860	4.7660	4.5629
θ_{12}	20	20.8802	20.9267	19.7113
NMSE	**0**	**3.6827*e* − 3**	**0.0827**	**0.1625**

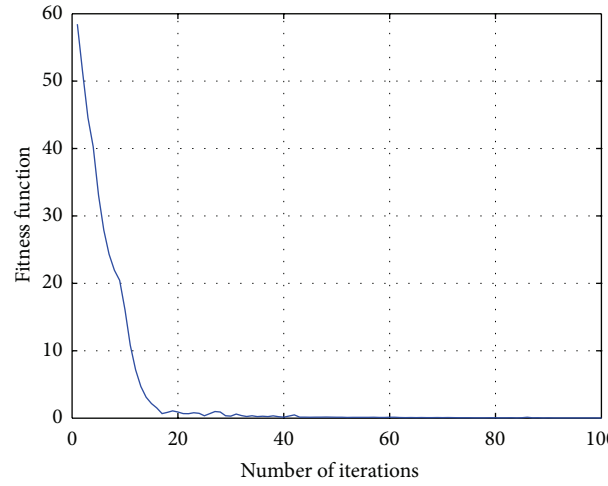


Figure 2: Convergence rate for PSO_1 algorithm.

PSO_2: algorithm based on relation (18) with ω, c_1, and c_2 defined by relations (22) and (23) and $\omega_i = 0.9$, $\omega_f = 0.4$, $c_{1i} = 2.5; c_{2i} = 0.5; c_{1f} = 0.5, c_{2f} = 2.5$;

PSO_3: algorithm based on relation (18) with ω, c_1, andc_2 constants: $\omega = 0.75, c_1 = 1.8, c_2 = 2.2$, and $V_{\max} = 10$.

Noise: {zero-mean white: SNR = 50 dB, 40 dB, 30 dB}.

The results of these simulations are presented in Tables 1, 2, and 3. The simulations were performed with a number of particles between 50 and 120. All the presented results are obtained for a number of 80 particles. For a greater number of particles the accuracy of the estimates was not better. In Figure 2 is presented the convergence rate for PSO_1 algorithm with a sampling period of 1 min.

The results presented in Tables 1–3 suggest that our proposed parameter estimation technique yields consistent results. The method is very simple, easily completed and needs fewer parameters, which made it fully developed. However, the research on the PSO is still at the beginning,

and a lot of problems are to be resolved. Research on the topology of the new pattern particle swarm which has a better function can be carried out. The neighbouring topology of the different particle swarms is based on the imitation of the different societies. It is meaningful to the use and spread of the algorithm to select the proper topology to enable PSO have the best property and do the research on the suitable ranges of different topologies. Blending PSO with the other intelligent optimization algorithms means combining the advantages of the PSO with the advantages of the other intelligent optimization algorithms to create the compound algorithm that has practical value. For example, the particle swarm optimization algorithm can be improved by the Simulated Annealing approach. Rapid swarm convergence is one of the main advantages of PSO, but this can also be problematic since, if an early solution is suboptimal, the swarm can easily stagnate around it without any pressure to continue exploration. Overall the results indicate that PSO algorithms can be used in the optimization of parameters during model identification.

5. Conclusions

The paper presents a particle swarm optimization based identification procedure for offline estimation of yield and kinetic coefficients in an anaerobic wastewater treatment bioprocess. The identification scheme is formulated in terms of an optimization problem where the error between an actual physical measured response of the system and the simulated response of a parameterized model is minimized. This function is multimodal, and classical iterative methods fail to find the global optimum. The estimation of the system parameters is achieved as a result of minimizing the error function by the PSO algorithm.

The simulations evaluated the effects of sampling period and some basic variants of PSO algorithm and of the noisy measurements. The proposed strategy can still converge to accurate results even in the presence of measurement noise, as illustrated by the numerical study. The PSO algorithm has a simpler procedure and higher computational efficiency than other optimization techniques.

Nomenclature

c_1, c_2: Acceleration coefficients
D: Dilution rate (h^{-1})
F: Vector of feeding rates
F_{in}: Input feed rate (Lh^{-1})
K: Matrix of yield coefficients
$k_i, i = \overline{1,7}$: Yield coefficients
Q: Vector of rates of removal of the components in gaseous form
S_1: Glucose concentration (g/L)
S_2: Acetate concentration (g/L)
S_5: Inorganic carbon concentration (g/L)
X_1: Acidogenic concentration (g/L)
X_2: Acetoclastic methanogenic concentration (g/L)
P: Methane concentration (g/L)
S_{in}: Glucose concentration on the feed (g/L)
Φ: Vector of reaction rates (reaction kinetics)
μ_1^*, μ_2^*: Maximal specific growth rates (h^{-1})
ξ: State vector
$\mathfrak{R}$: The set of real numbers
Ω: Inertia weight
θ: The vector of unknown parameters
$\theta_i, i = \overline{1,9}$: The unknown parameters.

Acknowledgments

This work was supported by CNCSIS-UEFISCSU, Project no. PN II-RU TE 106, 9/04.08.2010.

References

[1] G. Bastin and D. Dochain, *On-Line Estimation and Adaptive Control of Bioreactors*, Elsevier, New York, NY, USA, 1990.

[2] L. Chen, *Modelling, identifiability and control of complex biotechnological systems [Ph.D. thesis]*, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, 1992.

[3] L. M. Li and S. A. Billings, "Continuous time non-linear system identification in the frequency domain," *International Journal of Control*, vol. 74, no. 11, pp. 1052–1061, 2001.

[4] L. Ljung, *System Identification—Theory for the User*, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition, 1999.

[5] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[6] J. M. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Mich, USA, 1975.

[7] A. Colorni, M. Dorigo, and V. Maniezzo, "Distributed optimization by ant colonies," in *Proceedings of the European Conference on Artificial Life (ECAL '91)*, Elsevier, 1991.

[8] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, 1995.

[9] J. Kennedy, "Particle swarms: optimization based on sociocognition," in *Recent Developments in Biologically Inspired Computing*, The Idea Group, 2004.

[10] H. Li, H. Zhong, Z. Yan, and X. Zhang, "Particle swarm optimization algorithm coupled with finite element limit equilibrium method for geotechnical practices," *Mathematical Problems in Engineering*, vol. 2012, Article ID 498690, 14 pages, 2012.

[11] O. Bernard and G. Bastin, "Identification of reaction networks for bioprocesses: determination of a partially unknown pseudo-stoichiometric matrix," *Bioprocess and Biosystems Engineering*, vol. 27, no. 5, pp. 293–301, 2004.

[12] D. Selişteanu, E. Petre, and V. B. Răsvan, "Sliding mode and adaptive sliding-mode control of a class of nonlinear bioprocesses," *International Journal of Adaptive Control and Signal Processing*, vol. 21, no. 8-9, pp. 795–822, 2007.

[13] S. F. Azevedo, P. Ascencio, and D. Sbarbaro, "An adaptive fuzzy hybrid state observer for bioprocesses," *IEEE Transactions on Fuzzy Systems*, vol. 12, pp. 641–651, 2004.

[14] I. Hassanzadeh and S. Mobayen, "Controller design for rotary inverted pendulum system using evolutionary algorithms," *Mathematical Problems in Engineering*, vol. 2011, Article ID 572424, 17 pages, 2011.

[15] T.-S. Zhan and C.-C. Kao, "Modified PSO method for robust control of 3RPS parallel manipulators," *Mathematical Problems in Engineering*, vol. 2010, Article ID 302430, 25 pages, 2010.

[16] P. Umapathy, C. Venkataseshaiah, and M. S. Arumugam, "Particle swarm optimization with various inertia weight variants for optimal power flow solution," *Discrete Dynamics in Nature and Society*, vol. 2010, Article ID 462145, 15 pages, 2010.

[17] M. Clerc and J. Kennedy, "The particle swarm: explosion, stability, and convergence in a multi-dimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.

[18] X. Qian, M. Cao, Z. Su, and J. Chen, "A hybrid particle swarm optimization (PSO)-simplex algorithm for damage identification of delaminated beams," *Mathematical Problems in Engineering*, vol. 2012, Article ID 607418, 11 pages, 2012.

[19] R. C. Eberhart and Y. Shi, "Comparing inertia weights and constriction factors in Particle Swarm Optimization," in *Proceedings of the Congress on Evolutionary Computating*, pp. 84–88, July 2000.

[20] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 1945–1950, IEEE Press, 1999.

[21] A. Ratnaweera, S. K. Halgamuge, and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 240–255, 2004.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 725301, 12 pages
http://dx.doi.org/10.1155/2013/725301

Research Article

Merging Static and Dynamic Depth Cues with Optical-Flow Recovery for Creating Stereo Videos

Fang-Hsuan Cheng[1] and Tze-Yun Sung[2]

[1] *Department of Computer Science and Information Engineering, Chung Hua University, Hsinchu City 300-12, Taiwan*
[2] *Department of Electronics Engineering, Chung Hua University, Hsinchu City 300-12, Taiwan*

Correspondence should be addressed to Tze-Yun Sung; bobsung@chu.edu.tw

Received 20 February 2013; Revised 18 April 2013; Accepted 27 April 2013

Academic Editor: Shengyong Chen

A method for estimating the depth information of a general monocular image sequence and then creating a 3D stereo video is proposed. Distinguishing between foreground and background is possible without additional information, and then foreground pixels are moved to create the binocular image. The proposed depth estimation method is based on coarse-to-fine strategy. By applying the CID method in the spatial domain, the sharpness and the contrast of an image can be improved by the distance of the region based on its color. Then a coarse depth map of the image can be generated. An optical-flow method based on temporal information is then used to search and compare the block motion status between previous and current frames, and then the distance of the block can be estimated according to the amount of block motion. Finally, the static and motion depth information is integrated to create the fine depth map. By shifting foreground pixels based on the depth information, a binocular image pair can be created. A sense of 3D stereo can be obtained without glasses by an autostereoscopic 3D display.

1. Introduction

For more than 100 years, the concept of depth in 3D images and videos has existed. Wheatstone [1] first created a stereoscopic picture pair by using the binocular parallax theory. Wheatstone makes the first prism stereoscope based on visual parallax theory. David Brewster (1781–1868) [2] uses two lenses to build a prism stereoscope. Both of them used two cameras. A picture would be taken for both the right and the left eye. The end result is a 3D effect when looking through prism stereoscope. George Swan Nottag, a London merchant, established a company for 3D glasses in 1845. More than one million 3D glasses and stereoscopic images were sold within four years. Brewsterz developed lenticular stereoscopic in 1858. Anderton [3] proposed a method to make 3D projectors by using polarized light in 1895. The inventor of TV, John Baird, showed 3D picture on his TV in 1942. Half a century later, the Japanese company SONY was trying to sell their 3D TV. At the same time, NHK was trying to provide a 3D TV service. Anaglyph 3D movies were very popular in 1950. The popularity of the 3D application caused the development of new technology which has been making rapid progress.

As the technology progresses, vivid 3D stereoscopic vision is drawing more and more attention. Since 3D stereoscopic movies and films are becoming more and more popular, 3D stereo application will lead to the evolution of the next generation TV system [4–8]. The 3D stereoscopic display technology has developed from red-blue glasses in the early days to 3D LCD display [9] without glasses. All these stereo equipments allow people to perceive a 3D stereo effect by feeding parallax images into the left and right eyes. Therefore, how to derive the binocular image pairing from a 2D image sequence has become a focus in research.

When looking at an object, the differences between the left and right eye are referred to as parallax. The distance in depth between objects is picked up because of the parallax of images. A pair of images which has parallax information can be used to give the perception of depth between objects, just like when the human brain perceives something. Using this theorem, an image which has parallax information can be created. This then gives the impression of 3D stereo when using a lenticular auto-stereoscopic 3D display [9].

It is very easy to generate a stereo image by using two cameras shooting the same image simultaneously [10]. The image

pair can be fed into the left and right eyes and a stereo image effect forms in the mind of the viewer. In order to make the 3D stereo image more readily available, a 3D stereo image sequence can be derived from a 2D image sequence. The most important 3D stereo display technology can be classified into two categories. One is obtaining the depth information from a single 2D image sequence [11, 12] and the other is deriving a binocular image pair from the image together with its depth information [13, 14]. In this study, the computed image depth (CID) method was used to estimate the coarse depth of a single image, and then the binocular image pair can be created by an image shifting operation based on the temporal motion information estimated with optical-flow method.

2. Related Works

Much attention has been paid to 3D stereo technology in recent years. 3D stereo products are becoming more and more popular. Getting the correct depth map is significant. The most popular and easy way to get 3D stereo images is to use two cameras set in a horizontal line simulating the human eye. The stereo image made with two cameras is the most reliable and direct way when compared to others because it does not need a complex calculation. All that is needed is controlling the cameras and making sure that the object and the cameras are set on the points of an isosceles triangle. The images made by this way already include the depth information. This method is already being used to make 3D stereo movies. Since two cameras are used to get the stereo images, this method cannot be used to get the depth information from the image or video of a single camera.

There are several ways to get the depth information from the image or video of a single camera. The utilization of database is one of them [15–17]. The depth map in the database is checked after the object is identified. The depth information of this method will be correct, if the match between the object and the database is correct. Using the vanishing point and the vanishing line to estimate the depth is another way [18]. When the edges of the scenery converge at the furthest point, the depth information can be defined by the point and the line. These are called the vanishing point and the vanishing line. Another way to estimate the depth information is using computed image depth (CID) method [19]. This method extracts information such as color, shadow, layers, clearness, contrast, object size, object overlap, and camera focus from the image to estimate image depth. The area of the image captured by the camera with correct focus will be clear in high contrast. In this study, the coarse depth information was first obtained by using CID method and then combined with the optical-flow method to estimate the fine depth information for the image sequence.

3. Coarse Depth from Single Image

Depth information can be determined from the arrangement of objects and the scene in the image shot by cameras. 2D display equipments such as the TV and movies are the most common display equipments recording only 2D information. However, there are some depth clues contained in the 2D image. By analyzing the entire depth cue, the depth of objects in the scene can be estimated.

3.1. Computed Image Depth (CID). The computed image depth (CID) method is often used to estimate the depth information in a static image. The method utilizes image-specific information such as color, shadow, layer, clarity, contrast, object size, object overlap, and camera focus to analyze depth structure of the image properties. Although the accurate distance of objects in an image is difficult to derive from only a 2D image, the relative distance can be estimated from image clues, and coarse depth information can be generated. The image is divided into several image blocks, and the distance of the image block is estimated from the image clarity and contrast.

3.1.1. Clarity. In general, the clarity of an object provides an important clue of the distance from the viewer. This depends on several factors such as camera focus and illumination. The clearer the image object is, the nearer it is to the viewer. The image clarity can be computed by a *Laplacian* 3×3 matrix operation as in (1). The higher the magnitude of the *Laplacian* value is, the clearer the image block and the nearer it is. Then the image block can be classified into three distance levels according to clarity with (2),

$$L = \frac{1}{6}\begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}, \tag{1}$$

$$f_s = \begin{cases} 1\,(\text{Far}), & 0 \le S < \dfrac{1}{3}S_{\max}, \\ 2\,(\text{Middle}), & \dfrac{1}{3}S_{\max} \le S \le \dfrac{2}{3}S_{\max}, \\ 3\,(\text{Near}), & \dfrac{2}{3}S_{\max} < S \le S_{\max}, \end{cases} \tag{2}$$

where f_s denotes the depth value, S is the clarity of image block, and $S_{\max}$ is the maximum clarity of all image blocks.

3.1.2. Contrast. When the statistical luminance distribution of image is centralized, the information content contained in the image is low. The image contrast measured by the amount of information gives an indication of how far the image object is from the viewer. The higher the contrast of the image object is, the nearer it is to the viewer. The image contrast can be computed by calculating the variance of image blocks as in (3), and the distance of each image block from the viewer can be estimated. Then the image block can be classified into three distance levels according to contrast with (4). Consider the following:

$$C_m(x, y) = \sqrt{\frac{1}{9}\left(\sum_{k=1}^{3}\sum_{l=1}^{3}\left(I(x+k-2, y+l-2) - \mu(x, y)\right)^2\right)}, \tag{3}$$

Table 1: Relationship between depth and distance.

Value	Distance
6	Near
4~5	Middle
2~3	Far

where $\mu(x, y) = (1/9)(\sum_{k=1}^{3} \sum_{l=1}^{3} I(x + k - 2, y + l - 2))$,

$$f_c = \begin{cases} 1\,(\text{Far}), & 0 \le C < \dfrac{1}{3}C_{\max}, \\ 2\,(\text{Middle}), & \dfrac{1}{3}C_{\max} \le C \le \dfrac{2}{3}C_{\max}, \\ 3\,(\text{Near}), & \dfrac{2}{3}C_{\max} < C \le C_{\max}, \end{cases} \tag{4}$$

where f_c denotes the depth value, C is the contrast of image block, and $C_{\max}$ is the maximum contrast of all image blocks.

3.1.3. Color. The depth estimated from the clearness and contrast is not so accurate that some errors will occur. Therefore, other information is added to make sure that the depth estimated with the CID method is closer to the true depth when combining clarity and contrast. Because the color of an image provides abundant information, this information can be used to correct errors in depth estimation. First, the color space is transformed from RGB to YCbCr, and then the objects are segmented by using Cb and Cr channels. Errors in the estimation of depth can be corrected by color segmentation, because the same object will have similar color at the same depth. The image blocks with the similar colors are merged from top to bottom until there are no new blocks to be merged. Then adjacent colors are merged if they are similar. Besides, general color concepts such as blue sky or green grassland are also used to correct errors in the estimation of depth.

3.2. Coarse Depth Map Generation. By applying the CID method, a coarse depth map can be roughly estimated from a single image. The overall operation is described as follows. First, the image is divided into many blocks (16 × 16 pixels), and the clarity and contrast of each block are calculated. Then each image block is classified according to three distance levels by the summation of the depth values calculated from clarity and contrast. Figure 1 is an example of estimated depth by combining the clarity and contrast. The relationship between depth and distance is shown in Table 1. Finally, the error in distance level is updated using color segmentation and general color as shown in Figure 2.

4. Fine Depth from Image Sequence

Image sequence contains more depth information than a single image [20]. Assuming that an object moves from left to right in image, the object with large motion distance appears to be closer than the object with small motion distance. This concept can be applied to image sequence as shown in Figure 3. The object's motion distance between previous and current frame is used to estimate depth. The motion direction and luminance change of corresponding pixels between previous and current frame can be defined as optical flow [21, 22] or image flow.

4.1. Image Flow. Image flow is defined as the movement of pixels in the image plane when either the object or camera is moving. The difference between previous and current frame can be computed to generate instantaneous speed and motion. Figure 4 shows a point P with coordinate (X, Y, Z) project onto image plane to get pixel p with coordinate (x, y). When P is moving with instantaneous speed $\vec{S}$, we can get the image flow $\vec{V} = (u, v) = (\Delta x/\Delta t, \Delta y/\Delta t)$ of pixel p in image plane.

4.2. Optical Flow. Optical flow is defined as the luminance variations of pixel due to illumination change or object movement. A static object projected on the image plane has no image-flow vector, but a nonzero optical-flow vector is obtained if illumination is changed. When an object is moving in image sequence under stable illumination, the image flow can be regarded as the optical flow.

Assume that an object is moving slowly while the illumination of the environment does not change, and then the luminance of pixels will be unchanged after being projected on the image plane. So the conservation equation preface of the luminance can be expressed as below:

$$E(x + \Delta x, y + \Delta y, t + \Delta t) = E(x, y, t), \tag{5}$$

where $E(x, y, t)$ denotes the luminance of pixel p of (x, y) on the image plane, Δx and Δy are displacements that p moves in the image plane after Δt time. Neglecting high-order items of the previous equation, we obtain

$$E(x, y, t) + E_x \Delta x + E_y \Delta y + E_t \Delta t = E(x, y, t), \tag{6}$$

where E_x, E_y, and E_t, denote the luminance of pixel $p(x, y)$ with deviation to x, y direction and time t. By eliminating the term $E(x, y, t)$, we obtain

$$E_x \Delta x + E_y \Delta y + E_t \Delta t = 0. \tag{7}$$

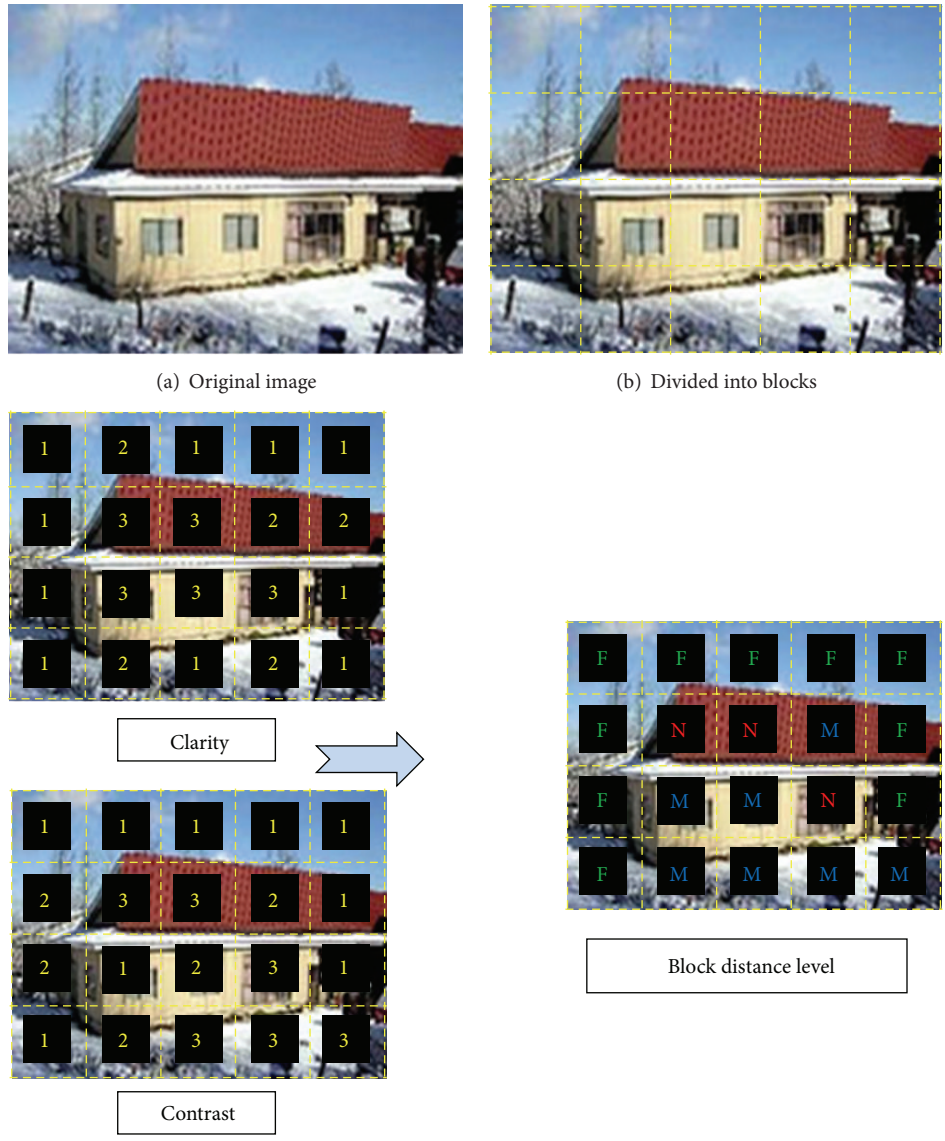
Then the following luminance change equation is obtained by dividing it by Δt,

$$E_x u + E_y v + E_t = 0, \tag{8}$$

where $u = \Delta x/\Delta t$ and $v = \Delta y/\Delta t$.

Horn and Schunck [23] proposed a first-order differential calculation method for E_x, E_y, and E_t without recursive calculations as follows:

$$E_x \approx \frac{1}{4}\left(E_{i,j+1,k} - E_{i,j,k} + E_{i+1,j+1,k} - E_{i+1,j,k} + E_{i,j+1,k+1} - E_{i,j,k+1} + E_{i+1,j+1,k+1} - E_{i+1,j,k+1}\right),$$



(a) Original image

(b) Divided into blocks

(c) Combining clarity and contrast to estimate block distance level

Figure 1: Distance of image block by calculating depth value based on clarity and contrast.

$$E_y \approx \frac{1}{4}\left(E_{i,j+1,k} - E_{i+1,j+1,k} + E_{i,j,k} - E_{i+1,j,k} + E_{i,j+1,k+1} - E_{i+1,j+1,k+1} + E_{i,j,k+1} - E_{i+1,j,k+1}\right),$$
$$E_t \approx \frac{1}{4}\left(E_{i,j,k+1} - E_{i,j,k} + E_{i+1,j,k+1} - E_{i+1,j,k} + E_{i,j+1,k+1} - E_{i,j+1,k} + E_{i+1,j+1,k+1} - E_{i+1,j+1,k}\right), \quad (9)$$

where E_x and E_y are the first-order partial differential of the x and y directions. The gradient ∇ is the sum of the absolute values of E_x and E_y. The difference image $D(x, y)$ is the absolute value of E_t, where E_t is the difference between two consecutive images. Figures 5(a) and 5(b) show the gradient and difference image calculated by

$$\nabla = |E_x| + |E_y|,$$
$$D(x, y) = |E_t(x, y)|. \quad (10)$$

Because the optical-flow difference image as shown in Figure 5(b) is calculated from both the previous and the current frame, it can be enhanced for further segmentation. In order to enhance the difference image, the motion history image (MHI) is calculated by adding a half of the motion history image previously made to the difference image, see (11). Figure 5(c) shows the MHI image obtained by (11). Finally, the search region of object image can be defined by

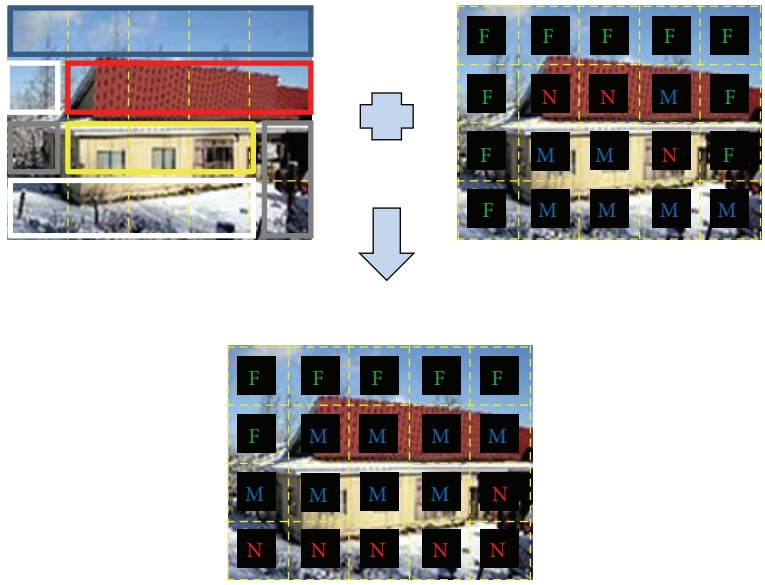


Figure 2: Distance error updated using segmentation and general color.

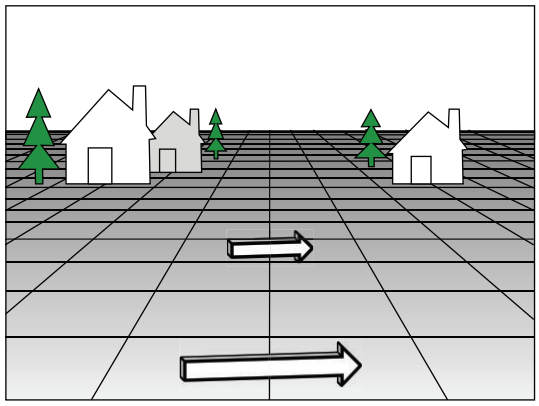


Figure 3: Relationship between distance and motion.

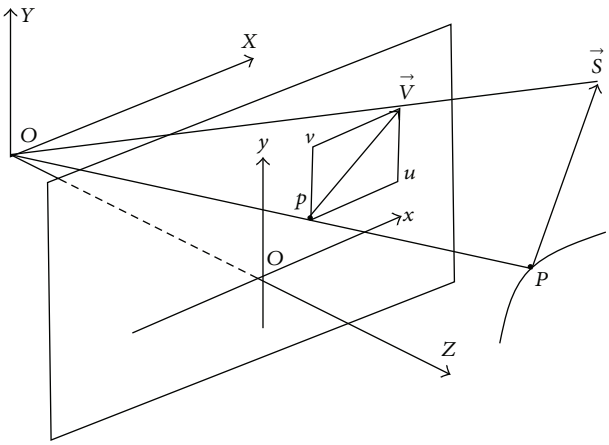


Figure 4: Illustration of image flow.

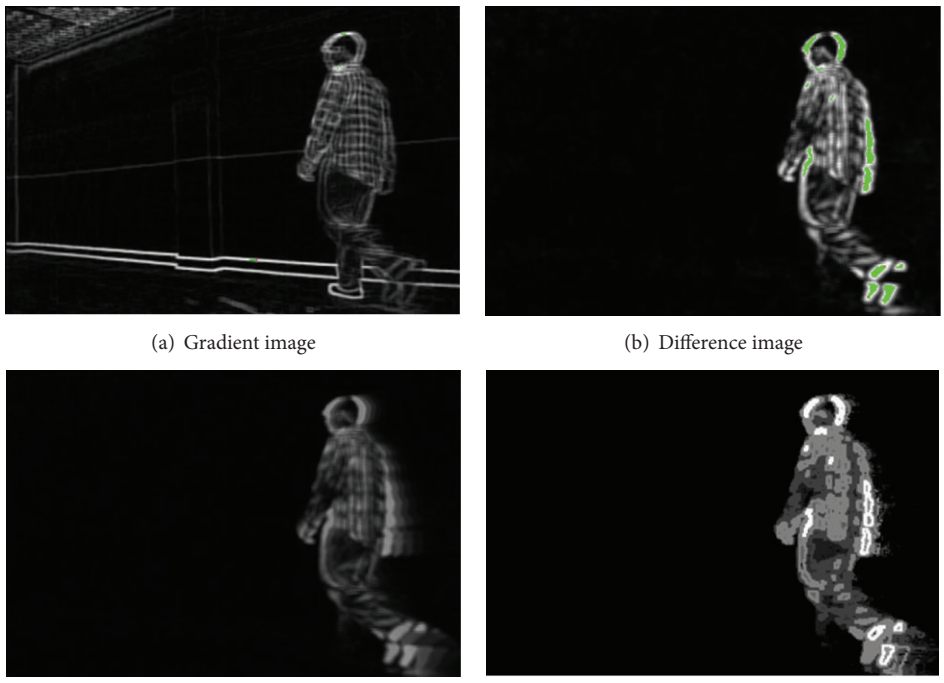
AND operation of the MHI and the gradient image as shown in Figure 5(d). Consider the following:

$$M(x,y) = D(x,y) + \frac{M(x,y)}{2}. \tag{11}$$

4.3. Object Segmentation. After determining the search region of an object, the binary image of an object can first be obtained by a simple threshold method; see Figure 6(a). Because the coarse depth estimated with the CID method is calculated on a block basis, the block image of an object can be obtained from a binary image as shown in Figure 6(b). The calculation of block distance level might be wrong in CID due to the object's texture. The block distance should be the same as or similar to one adjacent to it because they belong to the same object. So distance level in the CID method has to be corrected by using object segmentation. Because the same object usually has the same optical-flow value, the distance level can be updated by using the highest frequency in the optical-flow blocks. Figure 7 shows the experimental results after the distance level has been corrected according to object segmentation by the optical-flow method.

4.4. Background Segmentation. Background is the deepest area in the image. If a segment of the background area can be effectively detected, then the depth information can be evaluated more accurately. A true and more complete depth map can be acquired with accurate background information.

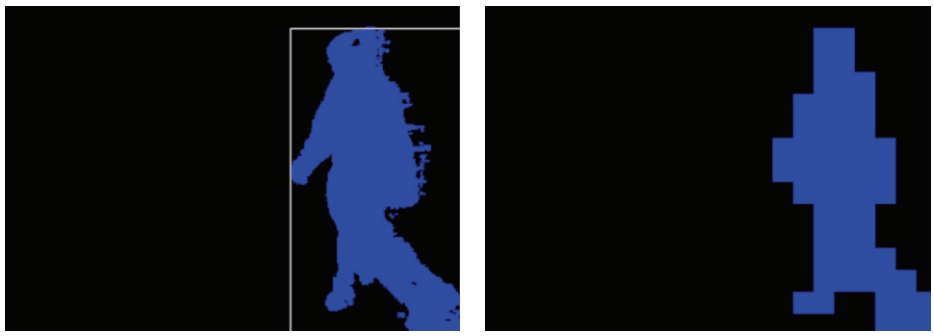
First, the RGB color information is down sampled to 3 bins per each channel, see (12). Then RGB color space is down sampled to 27 bins. Figure 8(a) shows the experimental results after down sampling. Because the background area usually appears on the top of the image, for example, the sky, the color histogram for top of the image can be recorded,



(a) Gradient image (b) Difference image

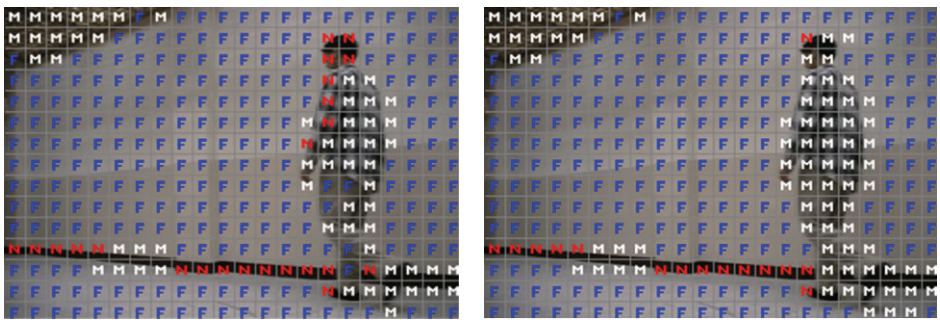
(c) Motion history image (d) Search region of object

Figure 5: Moving object detection with optical-flow method.



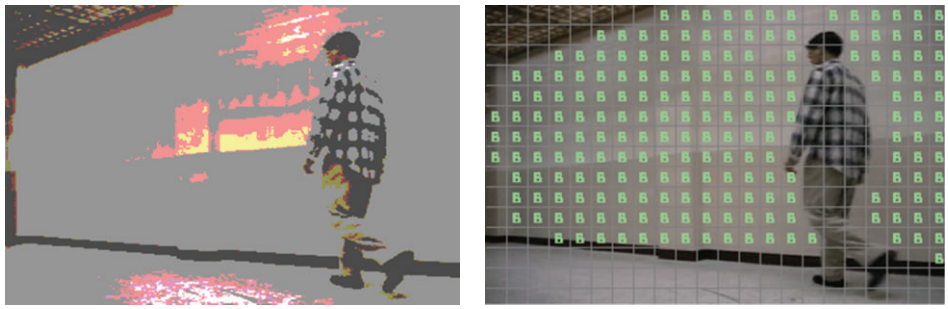
(a) Binary image (b) Block image

Figure 6: Object segmentation with the optical-flow method.



(a) Distance level with CID (b) Distance level correction

Figure 7: Distance level correction according to object segmentation using the optical flow method.



(a) Color downsampling (b) Background segmentation

Figure 8: Illustration of background segmentation.

and then the color with the highest frequency can be set as the background color. Then the same colors are merged from top to bottom until there are no new blocks to be merged. The colors adjacent to each other are checked and merge if they have the same color information. Figure 8(b) shows experimental results after background segmentation

$$R = \begin{cases} 75, & 0 < R \le 75 \\ 150, & 75 < R \le 150 \\ 255, & 150 < R \le 255 \end{cases} \quad \text{same for } G \text{ and } B \text{ channel.} \tag{12}$$

After background segmentation, the background is set to the deepest area in the image when creating the 3D stereo image.

4.5. Binocular Image from Optical-Flow Vector. Considering a point in spatial space, one can see it as point A with the left eye but as point B with the right eye due to the different viewing angles. In Figure 9, if the point is moving in space, then point A seems to be moving to C using the left eye and B moving to D using the right eye. While A is moving to C through B, A and B will have the time difference Δt and the same with C and D. If point A is at time t, and C at time $t + 1$, then B and D will be at time $t + \Delta t$ and $t + \Delta t + 1$. Projecting A and C on the image plane, pixels a and c are obtained in the left eye image. By moving pixel a to b and c to d in the left eye image, pixel b' and d' are obtained in the right-eye image. With this procedure, a binocular image can be created.

Two frames are used to substantiate this idea. First, the previous frame is used as the left-eye image, and then the optical-flow vector of the same point in the current frame is calculated. The shift of the pixel can be defined from the optical-flow vector according to the depth ratio. Then the right-eye image can be created according to the pixel shift in the current frame. When the motion is not large enough, the pixel shift calculated by the depth ratio will be small. This causes an unacceptable stereo effect. In order to improve the stereo effect, the pixel shift is subtracted from the previous frame to create the left-eye frame, and the pixel shift from the current frame is added to create a double depth distance in the right-eye frame as shown in Figure 10.

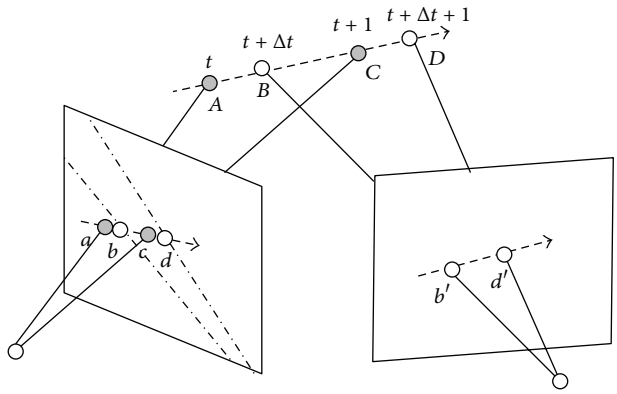


Figure 9: Demonstration of time difference of moving point.

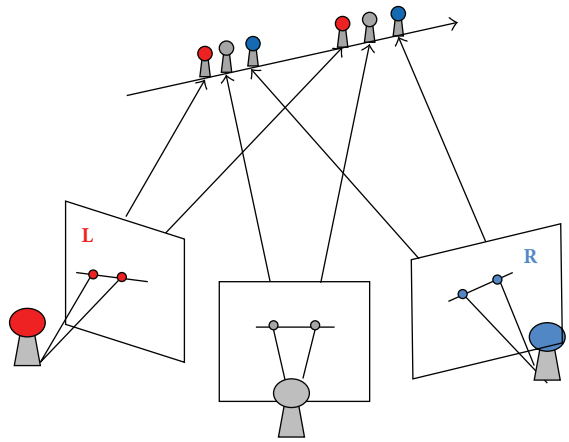


Figure 10: Illustration of binocular image pair creation.

The optical-flow method cannot fully define the distance between objects. So the CID method and the optical-flow method are combined to improve the depth estimation and recovery. The pixel shift value Δs is calculated as in Table 2. Because of the horizontal parallax, the 3D stereo effect can only be perceived with a horizontal pixel shift based on the depth information, so the vertical parallax is fixed to $(y_1 - y_2)/4$.

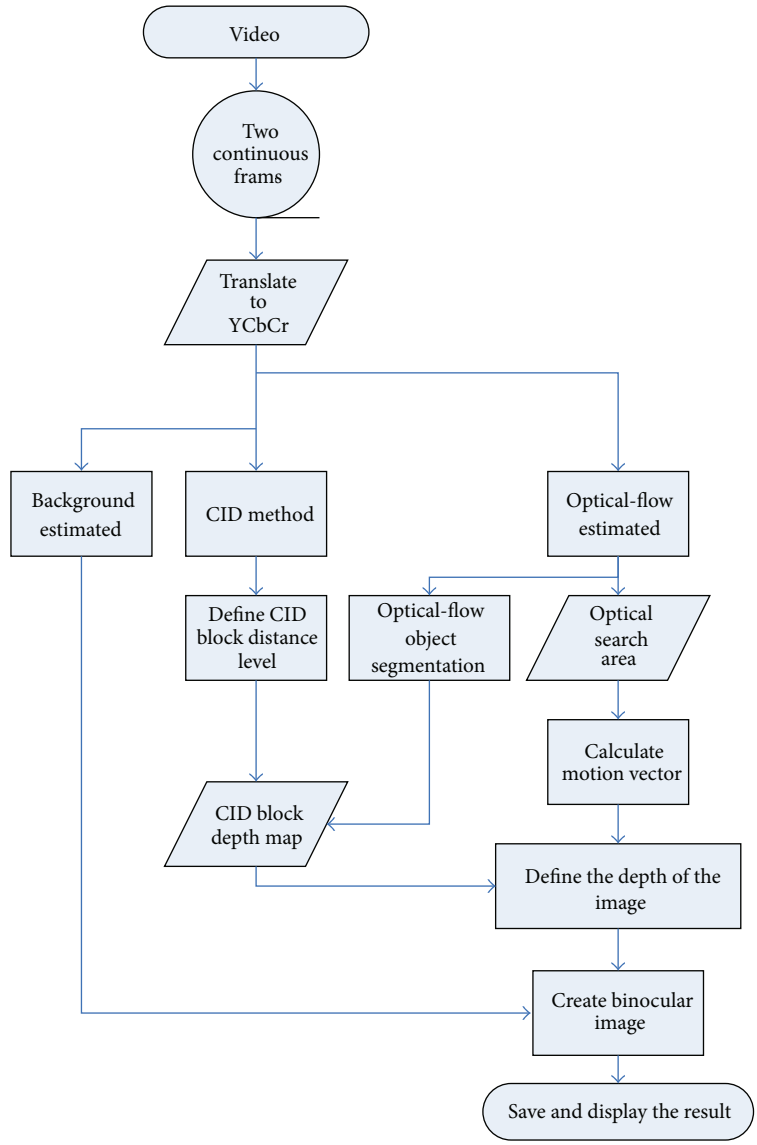


Figure 11: Flow chart of the proposed system.

Table 2: Relationship between distance level and pixel shift value.

CID distance level	Δs
Far	$(x_1 - x_2)/8$
Middle	$(x_1 - x_2)/4$
Near	$(x_1 - x_2)/2$

5. Experimental Results

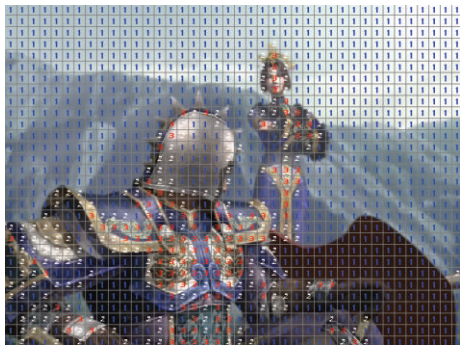
The proposed overall system flow diagram is shown in Figure 11. Figure 12 is an example of two frames of an image sequence. The image is first converted to Y, C_r, and C_b spaces and then the luminance channel is used to compute the image depth and optical-flow vector. Figure 13 shows the experimental results by computing image depth according to clarity and contrast of the image. Figure 14 depicts the estimated depth (map) computed from the optical-flow vector and modified with the CID method. Before creating the binocular image, the background area is set to the furthest from the foreground in order to create the best depth structure in the stereo image. To create the 3D stereo binocular image, the pixel shift is subtracted from the previous frame to create the left-eye image, and the pixel shift from the current frame is added to create the right-eye image based on the estimated depth map. Figure 15 shows the left- and right-eye images of the binocular image. By integrating the left- and right-eye images into an interlaced image and displaying it in a 3D display system, the 3D stereo effect can be perceived without wearing glasses.

In order to prove the feasibility of the proposed method, many famous 2D movies have been converted into 3D

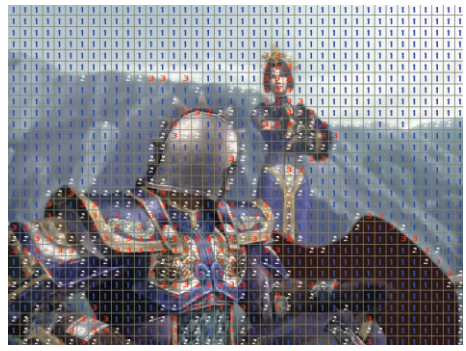
(a) Previous frame

(b) Current frame

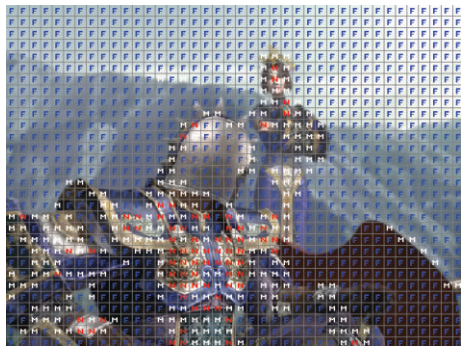
Figure 12: Two consecutive frames from image sequence.



(a) Calculated based on clarity



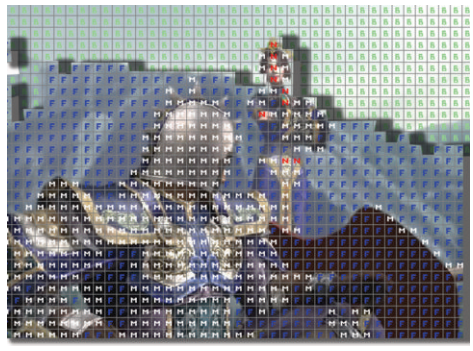
(b) Calculated based on contrast



(c) Estimated image depth based on distance level

Figure 13: Experimental result using the computed image depth method.

(a) Optical-flow vector



(b) Estimated depth map

Figure 14: Estimated depth map computed from the optical-flow vector and modified with the CID method.

(a) Left-eye image (b) Right-eye image

Figure 15: Binocular image creation.

(a) Original image (b) Depth map using CID (c) Optical-flow vector

(d) Estimated depth map (e) Left-eye image (f) Right-eye image

(g) Interlaced image

Figure 16: Example of 3D stereo binocular image creation from a 2D movie.

stereo movies. By viewing it on 3D auto-stereoscopic display equipment, 3D stereo can be experienced without wearing 3D stereo glasses. Figure 16 shows an example of 3D stereo binocular image creation from a 2D movie.

6. Conclusions and Future Works

This paper presents a method for creating a 3D stereo video by using an enhanced CID method combined with the optical-flow method. The static depth structure is analyzed; then the optical-flow method is used to determine the motion distance between current and previous frames. The image depth can be estimated by merging static and dynamic depth cues and using the optical-flow recovery method to generate a binocular view. We do not need any prior information of camera parameter and depth map as with other methods. All we need is a monocular image sequence. After generating binocular image sequences, they can be viewed on 3D stereo display equipment and have a 3D experience.

In the experimental result, the effect was one of a background area being displayed behind the screen and the foreground area in front of screen. There is depth between

the foreground and background. With large moving areas, a multilayer depth is perceived, not just a single depth. The whole scene gives a satisfactory 3D experience. However, the 3D effect is not as good as the image sequences shot with a two-camera system. In order to improve the effect of the 3D stereo experience, further research need to be done. The following three directions are suggested for this.

(1) Some image denoising algorithms [24] can be used to improve the accuracy of estimated depth information.

(2) Because there is such a variety in the types of videos, it is difficult to obtain a common structure of the depth map. For true depth information, all research in scene analysis needs to be investigated. Creating several different scene modes is the most common approach. Some scene features such as affine moment [25] can be used to classify the different scene modes. By using different models to match the different scenes, the optimum depth information can be obtained.

(3) Each kind of 3D display equipment has its own display theory and hardware specification with the result that there are different effects on different display equipment. The display effect is restricted by the software provided by the vendor. For better display quality and 3D stereo effect, the specifications of the display equipment need to be addressed, and the software must be matched for the proposed 3D stereo binocular image creation algorithm.

Acknowledgment

The authors would like to thank the National Science Council of Taiwan for the support of this research with Grant no. NSC101-2221-E-216-034.

References

[1] C. Wheatstone, "Contributions to the physiology of vision. Part the first. On some remarkable and hitherto unobserved phenomena of binocular vision," *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371–394, 1838.

[2] D. Brewster, *The Stereoscope*, Rudolf Kingslake, London, UK, 1971.

[3] Anderton, U.S. Patent No. 542321, 1895.

[4] L. Blondé, "3D stereo rendering challenges and techniques," in *Proceedings of the Conference on Information Sciences and Systems*, pp. 17–19, Princeton, NJ, USA, March 2010.

[5] P. Merkle, K. Müller, and T. Wiegand, "3D video: acquisition, coding, and display," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 946–950, 2010.

[6] Y. Yan, "A novel method for automatic 2D-to-3D video," in *Proceedings of the 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pp. 7–9, June 2010.

[7] L. Zhang, C. Vázquez, and S. Knorr, "3D-TV content creation: automatic 2D-to-3D video conversion," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 372–383, 2011.

[8] Z. Zhang, "A compact stereoscopic video representation for 3D video generation and coding," in *Proceedings of the Data Compression Conference (DCC '12)*, pp. 189–198, Beijing, China, April 2012.

[9] A. H. Wang, Q. H. Wang, X. F. Li, and D. H. Li, "Combined lenticular Lens for autostereoscopic three dimensional display," *Optik—International Journal for Light and Electron Optics*, vol. 123, no. 9, pp. 827–830, 2012.

[10] J. Lee, "The 3D video processing method in the stereoscopic camera for mobile devices," in *Proceedings of the IEEE International Conference on Emerging Signal Processing Applications (ESPA)*, pp. 139–142, January 2012.

[11] G. S. Lin, C. Y. Yeh, W. C. Chen, and W. N. Lie, "A 2D to 3D conversion scheme based on depth cues analysis for MPEG videos," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '10)*, pp. 1141–1145, Suntec City, Singapore, July 2010.

[12] H. M. Wang, C. H. Huang, and J. F. Yang, "Depth maps interpolation from existing pairs of keyframes and depth maps for 3D video generation," in *Proceedings of the IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems (ISCAS '10)*, pp. 3248–3251, June 2010.

[13] P. J. Lee and E. Effendi, "Nongeometric distortion smoothing approach for depth map preprocessing," *IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 246–254, 2011.

[14] J. Choi, "Sparsity based depth estimation and hole-filling algorithm for 2D to 3D video conversion," in *Proceedings of the International Conference on Signals and Electronic Systems (ICSES '12)*, pp. 1–4, Wroclaw, Poland, September 2012.

[15] T. Hassner and R. Basri, "Example based 3D reconstruction from single 2D images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, 2006.

[16] K. C. Liu, Q. Wu, W. C. Chen, C. F. Wu, F. C. Jan, and T. Chen, "Example-based depth generation from single image for 3D content," in *Proceedings of the 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON '08)*, pp. 333–336, Istanbul, Turkey, May 2008.

[17] Z. Wang, "Depth template based 2D-to-3D video conversion and coding system," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '12)*, pp. 308–313, Melbourne, VIC, USA, July 2012.

[18] S. Battiato, S. Curti, M. la Cascia, M. Tortora, and E. Scordato, "Depth-map generation by image classification," in *Proceedings of the Three-Dimensional Image Capture and Applications VI*, vol. 5302, pp. 95–104, January 2004.

[19] S. Okada, "32. 2: a real-time 2D to 3D conversion technique using computed image depth," *SID Digest of Technique Papers*, vol. 29, no. 1, pp. 919–922, 1998.

[20] S. Chen, Y. Wang, and C. Cattani, "Key issues in modeling of complex 3D structures from video sequences," *Mathematical Problems in Engineering*, vol. 2012, Article ID 856523, 17 pages, 2012.

[21] M. B. Holte, T. B. Moeslund, and P. Fihl, "View-invariant gesture recognition using 3D optical flow and harmonic motion context," *Computer Vision and Image Understanding*, vol. 114, no. 12, pp. 1353–1361, 2010.

[22] Y. Yang, Q. Liu, R. Ji, and Y. Gao, "Dynamic 3D scene depth reconstruction via optical flow field rectification," *Plos One*, vol. 7, no. 11, article e47041, 2012.

[23] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.

[24] X. Zheng, Z. Liao, S. Hu, M. Li, and J. L. Zhou, "Improving spatial adaptivity of nonlocal means in low-dosed CT imaging

using pointwise fractal dimension," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 902148, 8 pages, 2013.

[25] J. Yang, M. Li, Z. Chen, and Y. Chen, "Cutting affine moment invariants," *Mathematical Problems in Engineering*, vol. 2012, Article ID 928161, 12 pages, 2012.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 379718, 10 pages
http://dx.doi.org/10.1155/2013/379718

Research Article

Maximum Variance Hashing via Column Generation

Lei Luo,[1] Chao Zhang,[2] Yongrui Qin,[3] and Chunyuan Zhang[1]

[1] *College of Computer, National University of Defense Technology, Changsha, Hunan 410073, China*
[2] *School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*
[3] *School of Computer Science, The University of Adelaide, Adelaide, SA 5005, Australia*

Correspondence should be addressed to Lei Luo; l.luo@nudt.edu.cn

Received 23 January 2013; Accepted 6 March 2013

Academic Editor: Shengyong Chen

With the explosive growth of the data volume in modern applications such as web search and multimedia retrieval, hashing is becoming increasingly important for efficient nearest neighbor (similar item) search. Recently, a number of data-dependent methods have been developed, reflecting the great potential of learning for hashing. Inspired by the classic nonlinear dimensionality reduction algorithm—maximum variance unfolding, we propose a novel unsupervised hashing method, named maximum variance hashing, in this work. The idea is to maximize the total variance of the hash codes while preserving the local structure of the training data. To solve the derived optimization problem, we propose a column generation algorithm, which directly learns the binary-valued hash functions. We then extend it using anchor graphs to reduce the computational cost. Experiments on large-scale image datasets demonstrate that the proposed method outperforms state-of-the-art hashing methods in many cases.

1. Introduction

Nearest neighbor search is a fundamental problem in many applications concerned with information retrieval, including content-based multimedia retrieval [1–3], object and scene recognition [4], and image matching [5]. Due to the exciting advancement of data acquisition techniques, more and more data have been produced in recent years, leading these applications to suffer from the expensive time and storage demand. Recently, hashing has become a popular method to address this issue in terms of storage and speed. These methods convert a high-dimensional data item, for example, an image, into a compact binary code so that more items can be loaded into the main memory and the distance between two items can be computed efficiently by using bit XOR operation of their binary codes, and therefore they have great potential to solve complex problems.

Seminal work of hashing, such as locality-sensitive hashing (LSH) [6], focuses on using random projection to generate random binary codes in the Hamming space. It is then extended to accommodate more distance metrics [7, 8] or kernelized to capture nonlinear relationships in the data space [9, 10]. Without using any training data, LSH and its variances can map close data samples to similar binary codes, and it is theoretically guaranteed that the original metrics are asymptotically preserved in the Hamming space as the code length increases. However, because of the random projection, they need very long codes to achieve good precision in practice.

Data-dependent hashing methods, instead, take advantage of the available data to learn more impact codes for specific tasks, leading to the recent endeavors in hashing. For instance, PCA hashing (PCAH) [11] generates linear hash functions through PCA projections of the training data and is suggested for producing more impact codes rather than random projections. PCAH can be considered as the simplest data-dependent hashing method and can not capture nonlinear similarity information that is available in the training data. Alternatively, spectral hashing (SH) [12] and self-taught hashing (STH) generate hash codes from the low-energy spectrums of data neighborhood graphs to seek nonlinear data representations. The difficulty is how to compute the code of an unseen item, which is known as the problem of *out-of-sample extension*. As a result, SH has to assume that the training data are uniformly distributed in a hyper rectangle, which limits its practicabilities. STH

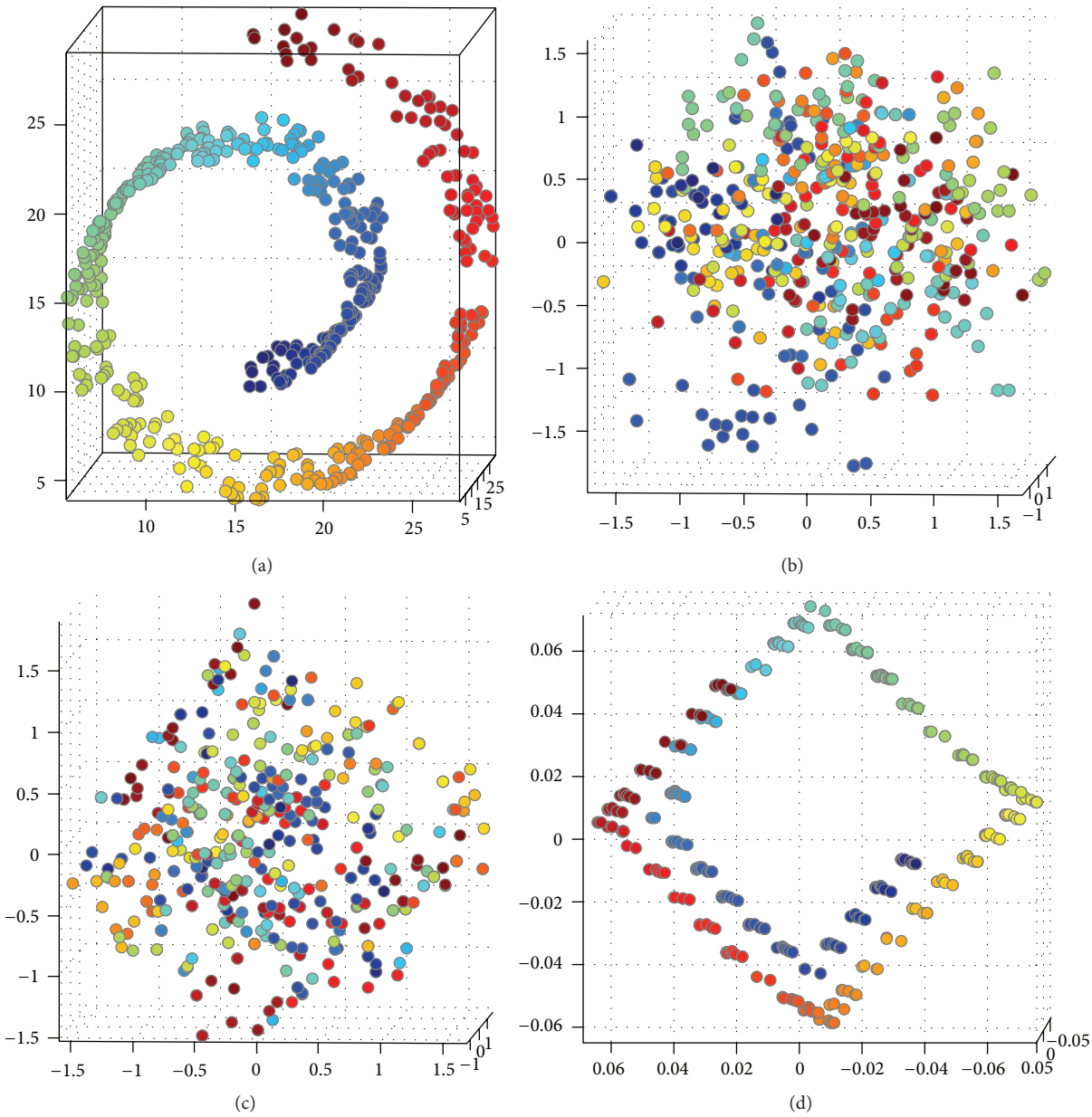


Figure 1: From (a) to (d), a Swiss roll and its hash codes (embedded to 3D by PCA) after applying SH, STH, and MVH-CG, respectively. MVH-CG can maintain the manifold of the Swiss roll in some sense. SH and STH fail to preserve the manifold.

addresses this problem in another way. By viewing the binary codes of the training data as pseudo-labels, it learns the hash functions via an extra pseudo-supervised learning stage. Nevertheless, learning errors in the self-taught stage may collapse the manifold structure of the learning data as illustrated in Figure 1.

Indeed, all these mentioned data-dependent methods aim at hashing the high-dimensional features of the training data with low-dimensional binary codes while preserving the underlying data structure. These methods normally suffer from loss of local geometric structure of the training data. However, by viewing this problem from a different angle and removing the constraint of Hamming space, it can be seen as a variation of the traditional dimensionality reduction problem. Among the large number of dimensionality reduction methods (see [13] for a survey), maximum variance unfolding (MVU) [14] can almost faithfully preserve the local geometric structure of the training data (e.g., the distances and angles between nearby samples in details).

Meanwhile, Liu et al. [15] recently proposed a scalable graph-based hashing method, named anchor graph hashing (AGH). They approximated the origin data by a small set of *anchors* and learned the hash functions by using the Nyström extension [16]. But the generalized eigenfunctions

are derived only for the Laplacian eigenmaps embedding and their performance may decline rapidly when their number increases.

In summary, the main contributions of this work can be described as follows. (i) Inspired by MVU, we propose maximum variance hashing (MVH), which directly embeds the high-dimensional data into a specified low-dimensional Hamming space and preserve the geometric properties of local neighborhoods. The idea is maximizing the total variance of the hash codes subject to the constraints imposed by the rigid rods between k nearest neighbors (kNN). (ii) To address the out-of-example extension difficulty, we propose a column generation-based solution of the derived optimization problem, named MVH-CG. As the size of training data increases, the construction of neighborhood graphs become infeasible. (iii) On the other hand, since the outputs of MVH-CG are a set of binary-valued functions, we can learn the hash functions on the anchor set and then apply them to any unseen data items directly. This motivates us to propose the anchor version of MVH (referred as MVH-A) to reduce the computational cost.

We put forward our algorithms and present the main results on several large-scale image datasets in the next sections.

2. Methodology

2.1. Notation. The following notations will be used throughout this paper:

(i) a bold lower-case letter ($\mathbf{x}$): a column vector;

(ii) a bold upper-case letter ($\mathbf{X}$): a matrix;

(iii) a calligraphic style upper-case letter ($\mathcal{X}$): a set;

(iv) $|\mathcal{X}|$: the number of elements in $\mathcal{X}$;

(v) $h(\cdot)$ or $d(\cdot,\cdot)$: a function with one or two inputs;

(vi) $\mathbb{R}^m$: the m-dimensional real number space;

(vii) (i, j): a pair of order numbers for two data samples.

2.2. Problem Definition. Given a set of samples $\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^m\}_{i=1}^n$, we would like to map each point to a low-dimensional binary code for fast nearest neighbor searching. Suppose that the desired number of dimensions of the embedded binary space is ℓ, the goal is to seek a transformation $f(\cdot) : \mathbb{R}^m \to \{0,1\}^\ell$, where the pairwise relationship of $\mathbf{x}_i, \mathbf{x}_j$ in $\mathbb{R}^m$ is kept, in some sense, with their counterpart $\mathbf{y}_i, \mathbf{y}_j$ in $\{0,1\}^\ell$. Here each code $\mathbf{y}_i = [h_1(\mathbf{x}_i); \ldots; h_\ell(\mathbf{x}_i)]$ is an ℓ-dimensional binary vector projected from $\mathbf{x}_i$ using a set of ℓ binary-valued functions $\mathcal{H} = \{h_k(\cdot)\}_{k=1}^\ell$.

Formally, we denote the relationship of $\mathbf{x}_i, \mathbf{x}_j$ as their Euclidean distance (Any other metric can be chosen based on the nature of $\mathcal{X}$, though here we use Euclidean distance as a normal setting.) $d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\|\mathbf{x}_i - \mathbf{x}_j\|^2}$. Meanwhile, the relationship of $\mathbf{y}_i$, $\mathbf{y}_j$ can be defined as their Hamming distance $d_{\text{Hamming}}(\mathbf{y}_i, \mathbf{y}_j)$ naturally. We minimize the following objective to keep the pairwise relationship:

$$\mathcal{O}(h_1, \ldots, h_\ell) = \sum_{\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}} \eta_{ij} \left(d(\mathbf{x}_i, \mathbf{x}_j) - C d_{\text{Hamming}}(\mathbf{y}_i, \mathbf{y}_j) \right)^2, \tag{1}$$

where η_{ij}, depending on the specific application, is the weight of how important the relationship of $\mathbf{x}_i$ and $\mathbf{x}_j$ should be kept during the transformation and C is a constant scale factor. Typically, it is reasonable to use the 0-1 adjacency matrix of the training data's kNN graph to define η_{ij} in order to preserve the local structure of $\mathcal{X}$. That is, the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ will be kept if and only if $\mathbf{x}_i$ is a kNN of $\mathbf{x}_j$ or the other way around. The kNN graph has been successfully used in STH to represent the local similarity structure, and the sparse nature of it greatly reduces the computational demand of the next optimization process. In addition, Weinberger and Saul [14] proved that if we add a small number of edges over the kNN graph, both the distances along the edges and the angles between edges in the original graph are preserved. Accordingly, we define η_{ij} as:

$$\eta_{ij} = \begin{cases} 1, & (i, j) \in \mathcal{N}, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $(i, j) \in \mathcal{N}$ if and only if $\mathbf{x}_i$ and $\mathbf{x}_j$ are k nearest neighbors themselves or common k nearest neighbors of another sample.

On the other hand, the Hamming distance used in $\{0,1\}^\ell$ may not be as descriptive as their counterpart in $\mathbb{R}^m$, especially when ℓ is small (which is desirable in practice). We therefore relax the discrete Hamming distance to the real weighted Hamming distance $\tilde{d}(\mathbf{y}_i, \mathbf{y}_j) = \sum_{k=1}^\ell w_k |h_k(\mathbf{x}_i) - h_k(\mathbf{x}_j)| = \sum_{k=1}^\ell w_k h_k^{ij}$, where w_k is a nonnegative weight factor associated with $h_k(\cdot)$ and h_k^{ij} is the shorthand symbol of $|h_k(\mathbf{x}_i) - h_k(\mathbf{x}_j)|$. Let $d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$, $\tilde{d}_{ij} = \tilde{d}(\mathbf{y}_i, \mathbf{y}_j)$, and $\xi_{ij} = \sum_{k=1}^\ell w_k h_k^{ij} - d_{ij}$ be the slack between them (we remove the constant scale factor C by merging it into the weight factors); the objective then is

$$\mathcal{O}(h_1, \ldots, h_\ell, w_1, \ldots, w_\ell) = \sum_{(i,j) \in \mathcal{N}} \xi_{ij}^2. \tag{3}$$

As discussed in [14], by preserving the pairwise distances in the extended kNN constraint set $\mathcal{N}$, we faithfully preserve the local manifold of $\mathcal{X}$. Direct optimization of (3), however, tends to crowd points together in one location due to the absence of pairs other than kNNs in the constraint, which is also a problem in the research of manifold learning. Various methods have been proposed in the literature to overcome this problem. t-SNE [17], for instance, uses the long-tailed Student's-t-distribution to make all pairwise information in use. But the constraint set is not sparse any more in t-SNE. Weinberger and Saul [14], instead, proposed to maximize the variance of the embedded codes. That is, the mapping

codes will be pulled apart as far as possible subject to the kNN distance constraints in the embedded space. The maximum variance formulation has two advantages—it is (i) a global formulation and (ii) economical in computation. The variance of $\{\mathbf{y}_i\}_{i=1}^n$ measured by weighted Hamming distance is $\sum_{1\le i,j\le n}\widetilde{d}_{ij} = \sum_{1\le i,j\le n}\sum_{k=1}^{\ell} w_k h_k^{ij} = \sum_{k=1}^{\ell} w_k a_k$. Here, $a_k = \sum_{1\le i,j\le n} h_k^{ij}$ is the variance of the kth bit. It is easy to see that $a_k = (n-s)s$, where $s = \sum_{i=1}^n h_k(\mathbf{x}_i)$ is the count of "1"s $h_k(\cdot)$ produced on $\mathcal{X}$. Combining them together, the optimization object can be written as:

$$\begin{aligned} &\min_{h_1(\cdot),\ldots,h_\ell(\cdot),w_1,\ldots,w_\ell} -\sum_{k=1}^{\ell} w_k a_k + \frac{\gamma}{2}\sum_{(i,j)\in\mathcal{N}} \xi_{ij}^2, \quad w_k \ge 0, \\ &\text{s.t.} \quad \xi_{ij} = \sum_{k=1}^{\ell} w_k h_k^{ij} - d_{ij}, \end{aligned} \tag{4}$$

where γ is the balancing parameter of the two terms. The introduced variable ξ_{ij} here is nontrivial, which will play a critical role in the derivation of (4)'s dual.

2.3. Column Generation Algorithm. There are two sets of variables to be optimized in (4)—the binary-valued hash functions $h_1(\cdot),\ldots,h_\ell(\cdot)$ and their weights $w_1,\ldots,w_\ell$. It is normally difficult to optimize them simultaneously, and the fact that the former one is a group of functions, even adds to the difficulty of solving the problem. We can use the column generation (CG) technique to find an approximate solution of it iteratively. It has been successfully applied in boosting algorithms [18, 19], which also have to generate a series of binary-valued functions and optimize their weights at the same time. Similar to the well-known expectation-maximization (EM) algorithm, column generation has a two-step iteration framework, where one set of variables are treated as constant in each step. The aim of column generation is to reduce the gap between the primal and the dual solutions iteratively.

We first consider $h_k(\cdot)$ in (4) as a known function and ξ_{ij} as a variable to be optimized. Then, the Lagrangian of it is

$$\begin{aligned} &\mathscr{L}\left(w_k, \xi_{ij}, v_k, u_{ij}\right) \\ &\quad = -\sum_{k=1}^{\ell} w_k a_k + \frac{\gamma}{2}\sum_{(i,j)\in\mathcal{X}} \xi_{ij}^2 - \sum_{k=1}^{\ell} v_k w_k \\ &\qquad + \sum_{(i,j)\in\mathcal{X}} u_{ij}\left(\sum_{k=1}^{\ell} w_k h_k^{ij} - d_{ij} - \xi_{ij}\right), \end{aligned} \tag{5}$$

where $v_k \ge 0$ and u_{ij} are Lagrange multipliers. At optimum, the first derivation of the Lagrangian w.r.t. the primal variables must vanish as follows:

$$\begin{aligned} \frac{\partial\mathscr{L}}{\partial w_k} &= -a_k - v_k + \sum_{(i,j)\in\mathcal{N}} u_{ij} h_k^{ij} \\ &= 0 \longrightarrow \sum_{(i,j)\in\mathcal{N}} u_{ij} h_k^{ij} - a_k = v_k \ge 0, \end{aligned} \tag{6a}$$

$$\frac{\partial\mathscr{L}}{\partial \xi_{ij}} = \gamma\xi_{ij} - u_{ij} = 0 \longrightarrow \xi_{ij} = \gamma^{-1} u_{ij}. \tag{6b}$$

The Lagrange dual function is

$$\begin{aligned} \inf_{w_k,\xi_{ij}} \mathscr{L} &= \inf_{w_k,\xi_{ij}} \sum_{k=1}^{\ell} w_k \left(-a_k - v_k + \sum_{(i,j)\in\mathcal{N}} u_{ij} h_k^{ij}\right) \\ &\qquad + \sum_{(i,j)\in\mathcal{N}} \left(\frac{\gamma}{2}\xi_{ij}^2 - u_{ij}\xi_{ij}\right) - \sum_{(i,j)\in\mathcal{N}} u_{ij} d_{ji} \\ &= \inf_{\xi_{ij}} \sum_{(i,j)\in\mathcal{N}} \left(\frac{\gamma}{2}\xi_{ij}^2 - u_{ij}\xi_{ij}\right) - \sum_{(i,j)\in\mathcal{N}} u_{ij} d_{ji} \\ &= -\frac{\gamma^{-1}}{2}\sum_{(i,j)\in\mathcal{N}} u_{ij}^2 - \sum_{(i,j)\in\mathcal{N}} u_{ij} d_{ij}. \end{aligned} \tag{7}$$

Then, it is easy to obtain the Lagrange dual as follows:

$$\begin{aligned} &\max_{u_{ij}} \quad -\frac{\gamma^{-1}}{2}\sum_{(i,j)\in\mathcal{N}} u_{ij}^2 - \sum_{(i,j)\in\mathcal{N}} u_{ij} d_{ij} \\ &\text{s.t.} \quad \sum_{(i,j)\in\mathcal{N}} u_{ij} h_k^{ij} \ge a_k. \end{aligned} \tag{8}$$

The idea of CG is to iteratively add a variable by selecting the most violated constraint of the dual, and then optimize the related variables by solving a restricted version of the origin optimization problem. It works on the basis that the sequence of restrict primal problems all have the same dual in which the most violated constraint indicates the steepest ascent direction of the dual. For (8), the subproblem for generating the most violated constraint is as follows:

$$\begin{aligned} h^*(\cdot) &= \arg\min_{h(\cdot)\in\mathscr{B}} \sum_{(i,j)\in\mathcal{N}} u_{ij} h_{h(\cdot)}^{ij} - a_{h(\cdot)} \\ &= \arg\min_{h(\cdot)\in\mathscr{B}} \sum_{(i,j)\in\mathcal{N}} u_{ij}\left|h(\mathbf{x}_i) - h(\mathbf{x}_j)\right| \\ &\qquad - \sum_{1\le i,j\le n}\left|h(\mathbf{x}_i) - h(\mathbf{x}_j)\right|, \end{aligned} \tag{9}$$

where $\mathscr{B}$ is the class of the base binary-valued hash function. Since there can be infinitely many functions in $\mathscr{B}$, we restrict it to be the decision stumps (Since decision stump is a deterministic model, the column generation process will converge when all the constraints in the dual are satisfied, which means that no new hash function can be generated. This convergence, however, usually happens after the required ℓ (typically less than 128) iterations in practice. Moreover, if it is a nondeterministic model here, column generation can produce new hash functions even after satisfying all constraints. We, therefore, do not mention the convergence in Algorithm 1.) [20], a machine learning model widely used in ensemble learning techniques, on the training set $\mathcal{X}$. By the restriction of decision stump, (9) can be solved by exhaustive search within reasonable time.

Input: Training data $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^n$; balancing parameter $\gamma > 0$; length of hash codes ℓ.
(1) **Initialize**: $\mathbf{w} = \mathbf{0}$; $\mathcal{H} = \emptyset$; assign a positive constant to each element of $\mathbf{u}$.
(2) **for** $t = 1, \ldots, \ell$ **do**
(3) Find a new binary hash function $h_t(\cdot)$ by solving (9); Add $h_t(\cdot)$ to $\mathcal{H}$ and the restricted primal problem (10); Solve (10) by Mosek to obtain the updated $\mathbf{w}$; update $\mathbf{u}$ by (11).
(4) Map $\mathcal{X}$ to $\mathcal{Y} = \{\mathbf{y}_i\}_{i=1}^n$ using $\mathcal{H}$.
Output: The learnt hash functions $\mathcal{H}$, their weights $\mathbf{w}$ and the binary codes $\mathcal{Y}$.

Algorithm 1: MVH-CG: Column generation for maximum variance hashing.

Input: Training data $\mathcal{X} = \{\mathrm{x}_i\}_{i=1}^n$; balancing parameter $\gamma > 0$; length of hash codes ℓ; number of anchors $c \ll n$.
(1) Generate the anchor set $\mathcal{A}$ by a clustering method (e.g. K-means) from $\mathcal{X}$;
(2) Find $\mathcal{H}$ via running Algorithm 1 on $\mathcal{A}$ with parameters γ and ℓ.
(3) Map $\mathcal{X}$ to $\mathcal{Y} = \{\mathbf{y}_i\}_{i=1}^n$ using $\mathcal{H}$.
Output: The learnt hash functions $\mathcal{H}$ and the binary codes $\mathcal{Y}$.

Algorithm 2: MVH-A: Maximum variance hashing with anchors.

We summarize our MVH-CG framework in Algorithm 1. In the tth iteration, we add a new function $h_t(\cdot) = h^*(\cdot)$ to the restricted primal problem. Let $r = |\mathcal{X}|$; we use $\mathbf{w} = [w_k]_{t\times 1}$, $\mathbf{a} = [a_k]_{t\times 1}$, $\mathbf{d} = [d_{ij}]_{r\times 1}$, and $\mathbf{u} = [u_{ij}]_{r\times 1}$ to, respectively, gather the corresponded scalars, and let $\mathbf{H} \in \mathbb{R}^{r\times t}$ to denote the learned hash functions' response on $\mathcal{N}$ such that the kth column of it is the gather of h_k^{ij}, $(i, j) \in \mathcal{N}$. Then, the restricted primal problem (without the introduced variables ξ_{ij}) can be written as:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\mathbf{w}^\top \mathbf{H}^\top \mathbf{H}\mathbf{w} - \left(\gamma^{-1}a^\top + d^\top \mathbf{H}\right)\mathbf{w} \\ \text{s.t.} \quad & \mathbf{w} \succeq 0. \end{aligned} \tag{10}$$

As a quadratic programming problem, (10) can be solved efficiently by the off-the-shelf solver Mosek [21]. The KKT condition (6b) establishes the connection between the primal variables $\mathbf{w}^*$ and the dual variables $\mathbf{u}^*$ at optimality as follows:

$$\mathbf{u}^* = \gamma\left(\mathbf{H}\mathbf{w}^* - \mathbf{d}\right). \tag{11}$$

The outputs of Algorithm 1 are the learnt binary-valued hash functions $\mathcal{H}$, their weights $\mathbf{w}$, and the binary codes $\mathcal{Y}$ of the training set $\mathcal{X}$. Given a new observation $\mathbf{x} \in \mathbb{R}^m$, $\mathcal{H}$ is used to obtain its ℓ-bit binary code as follows:

$$\mathbf{y} = f(\mathbf{x}) = \left[h_1(\mathbf{x}); \ldots; h_\ell(\mathbf{x})\right]. \tag{12}$$

The weight vector $\mathbf{w}$ is a result of the relaxation of the difficult discrete problem. We simply abandon it and use $\mathbf{y}$ only in hash applications.

2.4. Anchor Hashing Using MVH. In MVH-CG, we use the kNN set $\mathcal{N}$ to preserve the manifold structure in $\mathcal{X}$. The sparse nature of kNN matrix can reduce the number of variables in (8). Yet the size of training set $\mathcal{X}$ can be large, for example, the image dataset CIFAR-10 (http://www.cs.toronto.edu/~kriz/cifar.html) has 60,000 images, and the digits recognition dataset MNIST (http://yann.lecun.com/exdb/mnist/) has 70,000 samples. To solve the hashing problem more efficiently, Liu et al. [15] proposed to represent a data point by a set of *anchors*, which are the cluster centers obtained by running K-means on the whole (or a random selected small subsample of the) database. As the number of anchors are sufficiently small, the effective Laplacian eigenvector-based hashing method [12] can be processed on it in linear time. The main difficulty is how to generate hash codes for unseen points, which is known as *out-of-sample extension* problem. For this reason, [15] has to use the Nyström method [16] to learn eigenfunctions for a kernel matrix. Instead, our MVH-CG method learns a set of binary-valued hash functions, which can be directly applied to any data points. As a result, we only need to run the MVH-CG algorithm on the anchor set, and then apply the learnt binary-valued functions to hash the whole dataset. The anchor version of MVH (referred as MVH-A) is summarized in Algorithm 2.

3. Evaluation of the Algorithm

In this section, we evaluate the hashing behavior of MVH-CG and the influence of the parameters.

We first evaluate the hashing behavior of MVH-CG on a Swiss roll toy data. The Swiss roll is a 2D submanifold embedded in a 3D space and can be thought of as curling a piece of rectangular paper. We apply MVH-CG, spectral hashing [12] (SH) and self-taught hashing (STH) [22] on it with the same code length $\ell = 32$, respectively. To visualize the results, we embed the obtained hash codes into $\mathbb{R}^3$ by PCA. The results are illustrated in Figure 1. It is shown that all three methods tend to keep the neighborhood relationship during the mapping. MVH-CG maps the Swiss roll into

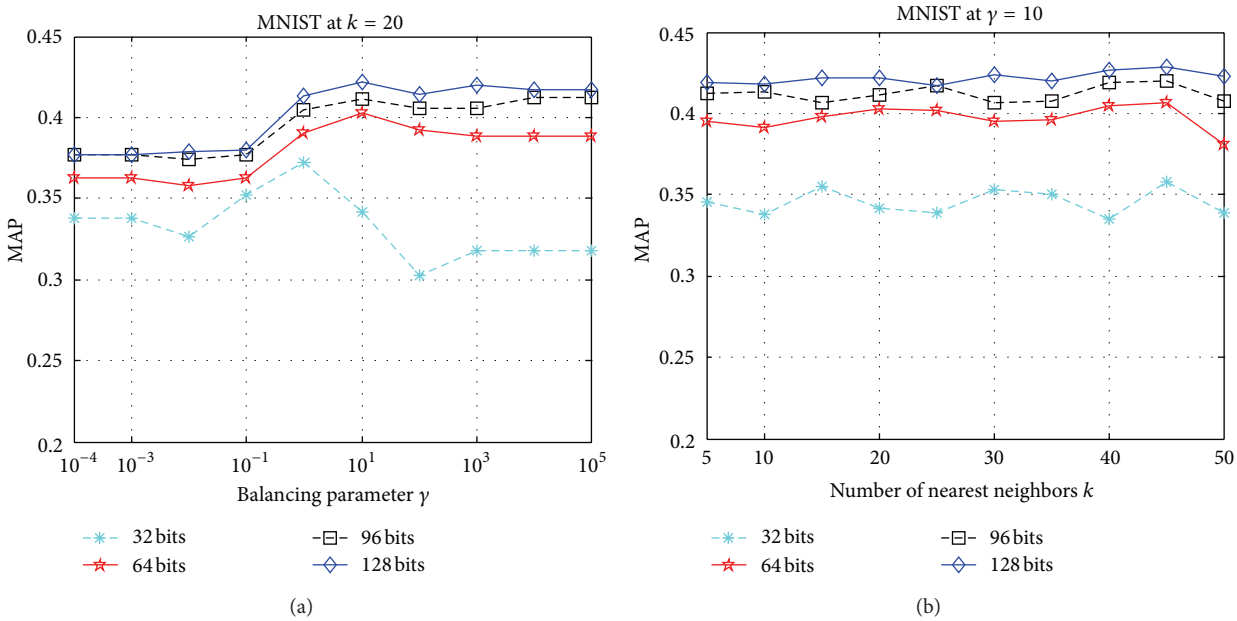


Figure 2: MAP results versus varying balancing parameter γ ((a), fixing $k = 20$) and number of nearest neighbors k ((b), fixing $\gamma = 10$) for MVH-CG. The comparison is conducted on a subset of the MNIST dataset.

a cube and can maintain the submanifold of it in some sense. SH and STH fail in preserving the manifold. For SH, one reason may be that it attempts to keep all pairwise relationships during the mapping. Studies in dimensionality reduction point out that kNN can achieve good approximation of the original manifold, and a method built on kNN kernel (used in MVH-CG) can analyze data that lies on a low-dimensional submanifold more faithfully than a predefined global kernel (used in SH) [23]. For STH, the failure may be due to the learning errors in its self-taught stage.

We then take the MNIST dataset as an example to evaluate the influence of the parameters. The MNIST dataset consists of 70,000 images of handwritten digits divided into 10 classes of 28×28 pixel image. We use the original 784-dimension pixel representation for the MNIST dataset. There are two parameters, k and γ, in MVH-CG. k is the kNN size, and γ is the balancing parameter between the two terms of object function (4). To eliminate the scale difference of these two terms, γ is multiplied with a constant n^2/r in experiments. We randomly generate 4,000 samples of the MNIST dataset, half for training and the rest for test, to evaluate the influences of k and γ. The results are summarized in Figure 2. From (a), we can see that the MAP curves rise after $\gamma = 1$, which indicates that the second term of (4) is somewhat more important; from (b), we see that the performance of MVH-CG does not change significantly with the number of nearest neighbors k. Based on the above observation, we set $\gamma = 10$ and $k = 20$ for the remainder of our experiments.

We also run an experiment to evaluate the influence of anchor set size c of MVH-A based on the MNIST dataset. We randomly select 1000 samples as test set and the others (which are 69,000 samples) for training. As described in Algorithm 2, in this experiment, we first reduce the 69,000 training samples into c anchors by K-means clustering and then run MVH-CG on the anchor set. The resulted MAP curves in Figure 3 basically remain stable from $c = 250$ to $c = 2000$. We, therefore, set $c = 1000$ for MVH-A.

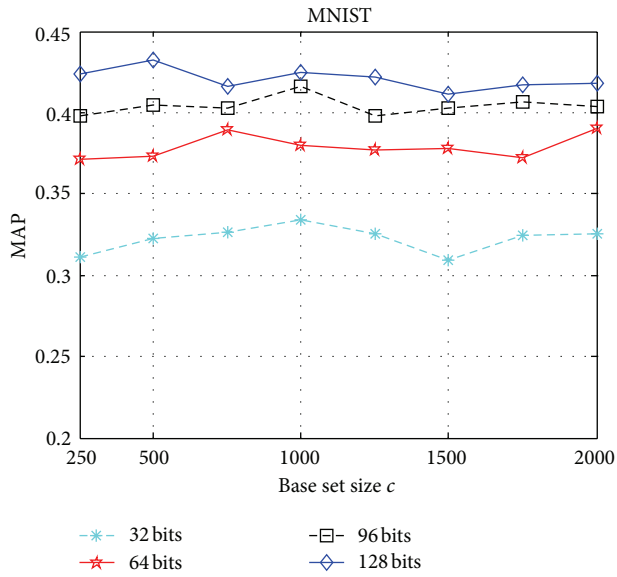


Figure 3: MAP results versus anchor set size c on the MNIST dataset.

4. Experiments

In this section, we evaluate the proposed hashing algorithm on the large-scale image datasets MNIST and CIFAR-10. The MNIST dataset consists of 70,000 images of handwritten

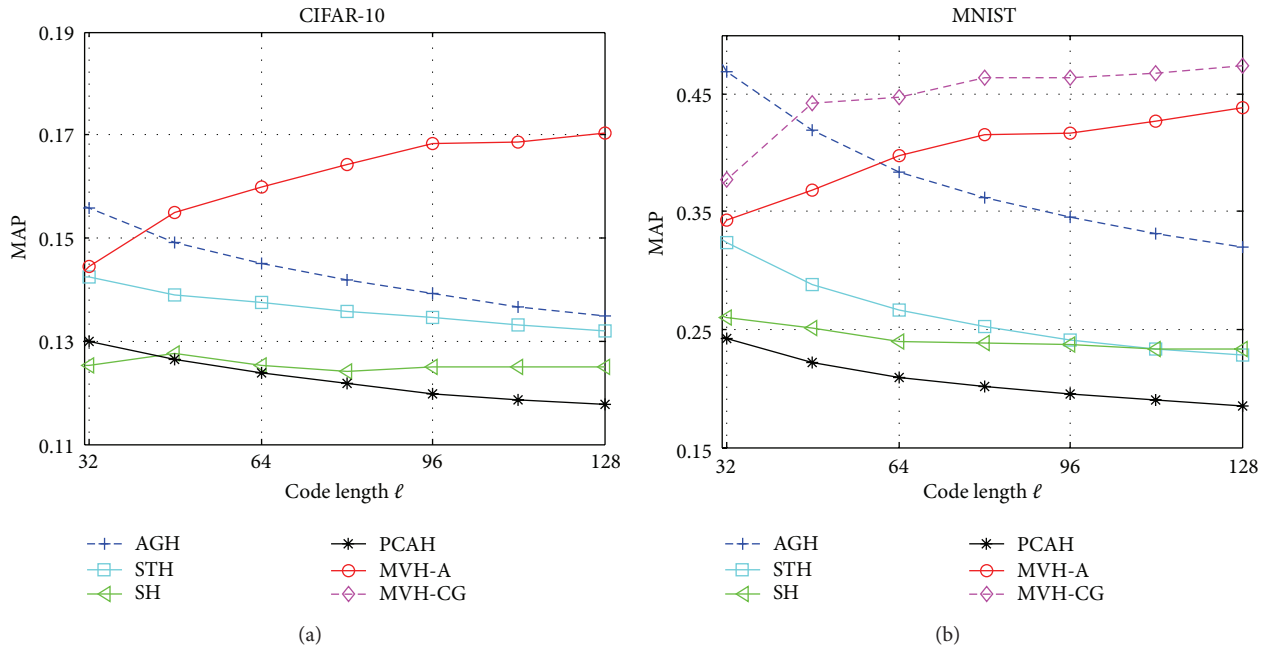


Figure 4: Comparison of different methods using MAP for varying code lengths on CIFAR-10 (a) and MNIST (b).

digits. The CIFAR-10 dataset consists of 60,000 images of 10 classes, which means that there are 6,000 samples for each class. In our experiments, we use the original 784-dimension pixel representation for the MNIST dataset and a 512-dimension GIST [24] feature for the CIFAR-10 dataset. Both of them are split into a test set with 1,000 images and a training set with all other samples. Since the proposed MVH method is fully unsupervised, we compare it with several unsupervised hashing algorithms including PCA-based hashing (PCAH) [11], spectral hashing (SH) [12], self-taught hashing (STH) [22], and anchor graph hashing (AGH) [15]. The performance of the comparison methods is measured by Mean Average Precision (MAP) or precision-recall curves for Hamming ranking.

4.1. Results on the MNIST Dataset. We report the experimental results based on MAP for Hamming ranking with code length from 32 to 128 bits in Figure 4(b). We can see that AGH obtains a high score at very short code length. Its performance, however, declines rapidly as ℓ increases and is inferior to MVH-CG after $\ell = 48$. The performance of PCAH and STH, similar to AGH, also drops down with longer bit lengths. By contrast, MVH-A and MVH-CG consistently improve their performance as code length grows. This property is important in very large-scale problems, when a short hash code, with a length of 32 bits, for example, is not enough to describe the whole dataset. MVH-CG is consistently superior to MVH-A as more data are used in the learning process. Yet, MVH-A also catches up with AGH at $\ell = 64$. We then plot the precision-recall curves for the compared methods in Figure 5. It can be seen that the curves of AGH are relatively high at the beginning, but they drop rapidly when more samples are returned. We also see that our MVH methods perform better at larger ℓ, which confirms the observation in Figure 4. PCAH performs worst in this case since it simply generates the hash hyperplanes by linear projects, which cannot capture the nonlinear similarity information behind the training data. SH is slightly better, but much worse than others, because it relies upon the strict uniform data assumption.

4.2. Results on the CIFAR-10 Dataset. The CIFAR-10 dataset is a manually labeled subset of the well-known 80 million tiny images dataset [4]. It consists of 60,000 images from 10 classes as in the examples shown in the top row of Figure 6. Each image is represented by a 512-dimension GIST [24] feature and then hashed by MVH-A. MVH-CG is not run since decision stump on the whole dataset is expensive. The bottom row of Figure 6 shows the returning list of the example query, where the first 8 results are correct and the last 2 are false positives. The MAP scores against code lengths are plotted in Figure 4(a). On this dataset, we can see that MVH-A yields rising performance as the number of bits increases. It outperforms all its competitors from $\ell = 48$ onward and achieves the highest MAP score at $\ell = 128$. PCAH and SH perform worst again. Figure 7 shows the precision-recall curves of hamming ranking for the compared methods with different code lengths. When $\ell = 32$, MVH-A is inferior to AGH. As the code length grows, the areas under the precision-recall curves of MVH-A are much broader than AGH and other methods. This trend is consistent with the MAP results.

5. Conclusion

This paper has proposed a novel unsupervised hashing method based on manifold learning theories, which can

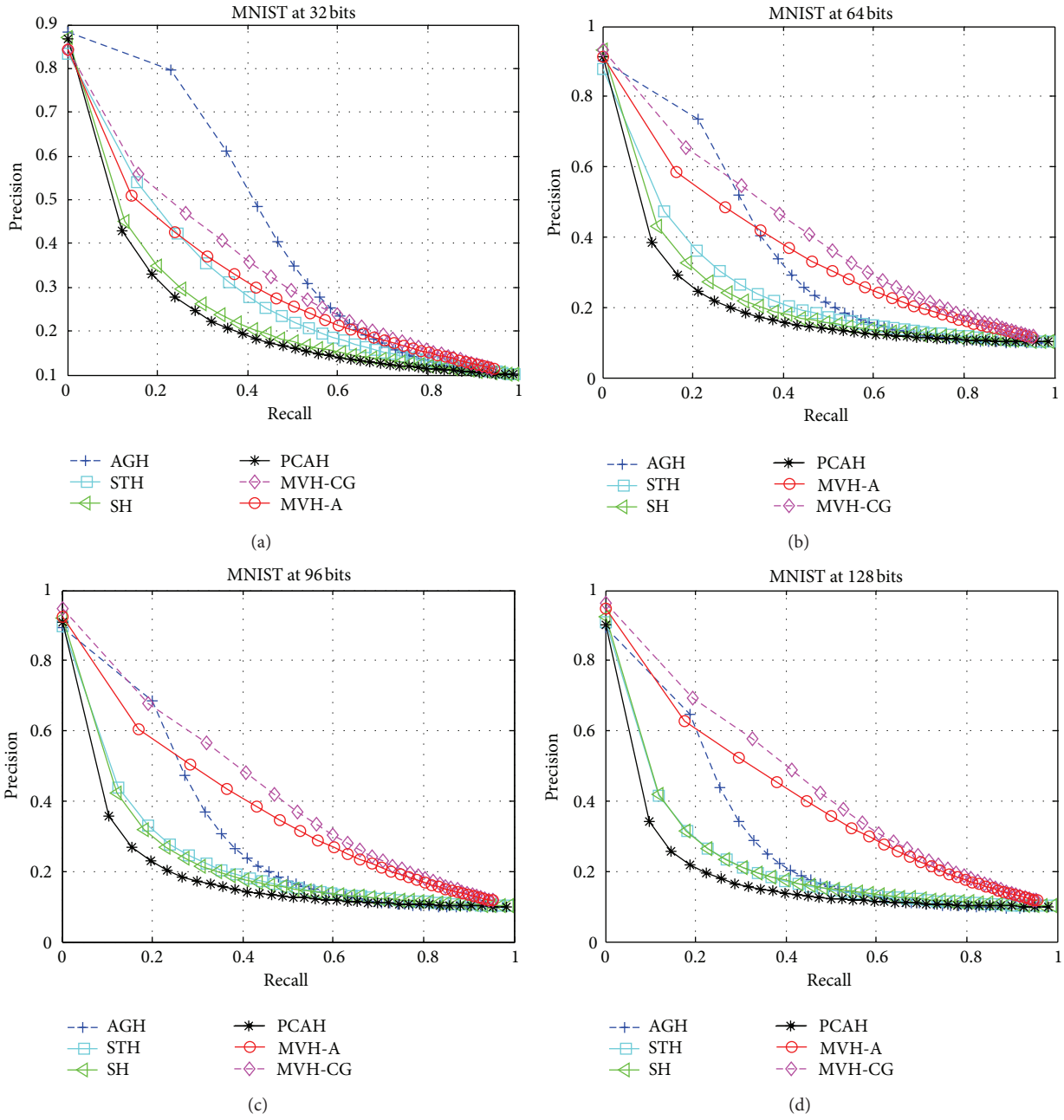


Figure 5: Precision-recall curves for competing methods on the MNIST dataset for different code lengths.

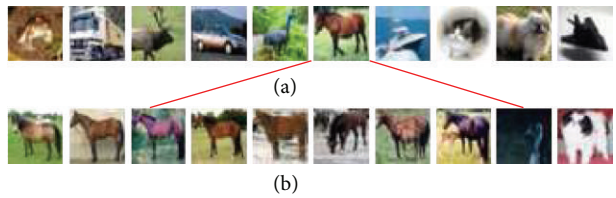


Figure 6: (a) Samples from the CIFAR-10 dataset, one for each category. (b) The results for a query of "horse" image returned by MVH-A with 128 bits. The last two returns are false positive.

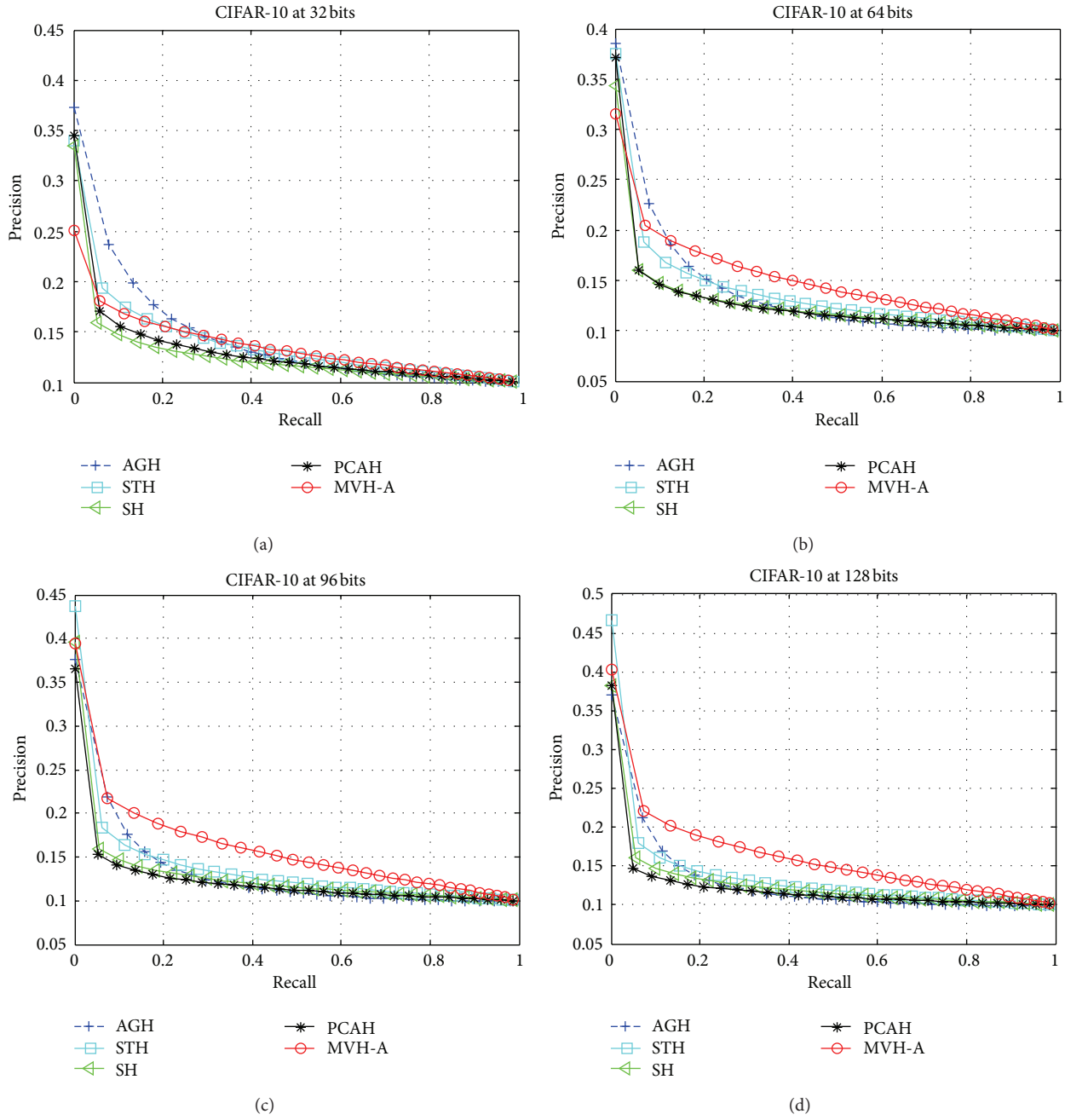


Figure 7: Precision-recall curves for competing methods on the CIFAR-10 dataset for different code lengths.

maximize the total variance of the hash codes while preserving the local structure of the training data. Two algorithms, MVH-CG and MVH-A, have been proposed to solve the derived optimization problem. Both of them can embed the input data into binary space while maintaining the submanifold with very short hash codes. The training process of MVH-A is faster than MVH-CG, but the anchor representation of MVH-A may degrade the retrieval performance of the resulted hash codes. Experimental results on large-scale image datasets show that, in the case of image retrieval, the proposed algorithms are consistently superior to the state-of-the-art unsupervised methods such as PCAH, SH, and STH and outperform AGH with relatively longer codes. The idea of manifold learning has a great potential for large-scale hashing. We are going to develop more efficient hashing method based on other manifold learning approaches.

Acknowledgments

The authors gratefully acknowledge the kind help from the Academic Editor Shengyong Chen. This work was supported by the National Nature Science Foundation of China under

NSFC nos. 61033008, 61272145, 60903041, and 61103080, Research Fund for the Doctoral Program of Higher Education of China under SRFDP no. 20104307110002, Hunan Provincial Innovation Foundation For Postgraduate under no. CX2010B028, Fund of Innovation in Graduate School of NUDT under nos. B100603 and B120605.

References

[1] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: state of the art and challenges," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 2, no. 1, pp. 1–19, 2006.

[2] S. Chen, J. Zhang, Y. Li, and J. Zhang, "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 118–127, 2012.

[3] L. Luo, C. Shen, C. Zhang, and A. van den Hengel, "Shape similarity analysis by self-tuning locally constrained mixeddiffusion," *IEEE Transactions on Multimedia*, 2013.

[4] A. Torralba, R. Fergus, and W. T. Freeman, "80 million tiny images: a large data set for nonparametric object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1958–1970, 2008.

[5] C. Strecha, A. Bronstein, M. Bronstein, and P. Fua, "LDAHash: improved matching with smaller descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 66–78, 2012.

[6] A. Gionis, P. Indyk, R. Motwani et al., "Similarity search in high dimensions via hashing," in *Proceedings of the International Conference on Very Large Data Bases*, pp. 518–529, 1999.

[7] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proceedings of the 20th Annual Symposium on Computational Geometry (SCG '04)*, pp. 253–262, June 2004.

[8] B. Kulis, P. Jain, and K. Grauman, "Fast similarity search for learned metrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2143–2157, 2009.

[9] B. Kulis and K. Grauman, "Kernelized locality-sensitive hashing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 6, pp. 1092–1104, 2012.

[10] M. Raginsky and S. Lazebnik, "Locality-sensitive binary codes from shift-invariant kernels," *Advances in Neural Information Processing Systems*, vol. 22, 2009.

[11] J. Wang, S. Kumar, and S. Chang, "Semi-supervised hashing for large scale search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2393–2406, 2012.

[12] Y. Weiss, A. Torralba, and R. Fergus, "Spectral hashing," *Advances in Neural Information Processing Systems*, 2008.

[13] L. Van der Maaten, E. Postma, and H. Van Den Herik, "Dimensionality reduction: a comparative review," *Journal of Machine Learning Research*, vol. 10, pp. 1–41, 2009.

[14] K. Weinberger and L. Saul, "Unsupervised learning of image manifolds by semidefinite programming," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 77–90, 2006.

[15] W. Liu, J. Wang, S. Kumar, and S. Chang, "Hashing with graphs," in *Proceedings of the International Conference on Machine Learning*, 2011.

[16] Y. Bengio, O. Delalleau, N. Le Roux, J. F. Paiement, P. Vincent, and M. Ouimet, "Learning eigenfunctions links spectral embedding and kernel PCA," *Neural Computation*, vol. 16, no. 10, pp. 2197–2219, 2004.

[17] L. Van Der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2625, 2008.

[18] A. Demiriz, K. P. Bennett, and J. Shawe-Taylor, "Linear programming boosting via column generation," *Machine Learning*, vol. 46, no. 1–3, pp. 225–254, 2002.

[19] C. Shen and H. Li, "On the dual formulation of boosting algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2216–2231, 2010.

[20] W. Iba and P. Langley, "Induction of one-level decision trees," in *Proceedings of the International Conference on Machine Learning*, pp. 233–240, 1992.

[21] Mosek, "The MOSEK interior point optimizer," http://www.mosek.com.

[22] D. Zhang, J. Wang, D. Cai, and J. Lu, "Self-taught hashing for fast similarity search," in *Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '10)*, pp. 18–25, 2010.

[23] L. Saul, K. Weinberger, J. Ham, F. Sha, and D. Lee, "Spectral methods for dimensionality reduction," *Semisupervised Learn*, pp. 293–308, 2006.

[24] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 928309, 14 pages
http://dx.doi.org/10.1155/2013/928309

Research Article

Detailed Simulation of Complex Hydraulic Problems with Macroscopic and Mesoscopic Mathematical Methods

Chiara Biscarini,[1] Silvia Di Francesco,[2] Fernando Nardi,[1] and Piergiorgio Manciola[3]

[1] *Warredoc, University for Foreigners of Perugia, Via XIV Settembre, 06100 Perugia, Italy*
[2] *CIPLA, Interuniversitary Centre for Environment, University of Perugia, Piazza Università 1, 06125 Perugia, Italy*
[3] *DICA, Department of Civil and Environmental Engineering, University of Perugia, Via G. Duranti 93, 06125 Perugia, Italy*

Correspondence should be addressed to Chiara Biscarini; chiara.biscarini@unistrapg.it

Received 30 January 2013; Accepted 15 March 2013

Academic Editor: Carlo Cattani

The numerical simulation of fast-moving fronts originating from dam or levee breaches is a challenging task for small scale engineering projects. In this work, the use of fully three-dimensional Navier-Stokes (NS) equations and lattice Boltzmann method (LBM) is proposed for testing the validity of, respectively, macroscopic and mesoscopic mathematical models. Macroscopic simulations are performed employing an open-source computational fluid dynamics (CFD) code that solves the NS combined with the volume of fluid (VOF) multiphase method to represent free-surface flows. The mesoscopic model is a front-tracking experimental variant of the LBM. In the proposed LBM the air-gas interface is represented as a surface with zero thickness that handles the passage of the density field from the light to the dense phase and vice versa. A single set of LBM equations represents the liquid phase, while the free surface is characterized by an additional variable, the liquid volume fraction. Case studies show advantages and disadvantages of the proposed LBM and NS with specific regard to the computational efficiency and accuracy in dealing with the simulation of flows through complex geometries. In particular, the validation of the model application is developed by simulating the flow propagating through a synthetic urban setting and comparing results with analytical and experimental laboratory measurements.

1. Introduction

The detailed three-dimensional (3D) simulation of free-surface flows is a challenging task for engineering projects [1] dealing with open flow hydraulics at both the large and small scale because of the computational burden and the large parameterization needed for representing the physics, geometry, and boundary conditions. Nevertheless, in recent times, the always larger availability of remotely sensed data, providing digital topographic and land cover information, and the decreasing impact of the computational burden make the application of small-scale high-resolution 3D numerical models feasible and attractive for engineering studies. In particular, it is foreseen that 3D hydraulic simulations shall support not only fluid mechanics analyses for engineering design at the small scale, but also large-scale applications in urban planning, river restoration, and flood mitigation and management projects.

There are several recent works in the literature that report the ability of the Navier-Stokes (NS) method using the two-dimensional (2D) shallow water (SW) equations [2] to simulate river surface flows, even if the SW hypothesis is not always valid for the inertial effects and significant gradients [3–5]. In fact, several fully three-dimensional NS simulations are available in the literature for representing not only dam-breaks [6, 7], but also unconfined surface flows along straight river segments, bends and confluences [8], as well as impulse waves and tsunamis [4, 9], and water falls [10] among others.

The lattice Boltzmann method (LBM) is an alternative numerical fluid dynamics scheme based on Boltzmann's kinetic equation [11] that represents the flow dynamics at the macroscopic level by incorporating a microscopic kinetic

approach that also preserves the conservation law [12, 13]. LBM has been adopted for the characterization of flows in porous media [14] and for multiphase flows [15–17] among others [18, 19]. The LBM demonstrated to work efficiently for describing the complex physics of non ideal flows and complex geometries [20]. There are also several works investigating the performances of the LBM in simulating shallow free-surface flows [21, 22], but with no emphasis on their use for practical engineering applications.

In this work a fully 3D front-tracking formulation of the LBM, that was originally presented [23] for foam dynamics and that was revisited for free-surface flow applications [24–26], is implemented, validated, and compared to the standard NS approach for transient hydraulic flows, including a synthetic case of flow propagation through an urban setting.

The storage effect of urban areas as well as the hydrograph separation effect resulting in different water transmission rates was investigated in several works [27]. Experiments of dam breaks were conducted [28] to investigate the effect of a flood flow propagating to inundate a single building [29]. Nevertheless, the application and potential validation of LBM for urban modeling of flash floods [30] has not been adequately developed yet.

The novel aspect of the proposed work is the investigation for the potential of the LBM to simulate the flooding in highly urbanized areas with specific regard to the detailed characterization of instantaneous discharge variations, like the one caused by dam or dyke breaks or by intense rainfall of flash floods, producing multidimensional street surface flow paths, which is one of the most complex and difficult hazardous phenomena to manage, prevent, and control. The interaction of water flow with anthropogenic features produces complex 3D flow structures that shall be simulated by taking into consideration all the resulting flow physics such as hydraulics jumps, flow constrictions, and back pressure effects. Disregarding those geometric and hydraulic features would result in a misinterpretation of the estimated water levels providing an inaccurate flood hazard and risk characterization.

In this study, given the impossibility of gathering measured data on such phenomena, a comparative analysis of the performance of LBM and NS in simulating flash flood propagation in urban areas is investigated by a comparison with experimental data of dam break hydraulic effects on a synthetic case study represented by an ideal city.

The present document is organized as follows: in the next section, the theoretical and numerical specifications of both LBM and NS are introduced and described. Then, in the case study section, the results of the application of the proposed methods are inserted with specific regard to (1) the test case of the LBM results compared to an analytical solution; (2) the analysis and comparison of LBM and NS codes applied on the simulation of a rapidly varying flow along a simple geometry of a straight river reach; (3) the evaluation and validation of the LBM and NS code application on the flow propagating through a synthetic urban domain by means of comparison with experimental data of a physical model. The last two sections represent, respectively, the discussion and conclusive remarks.

2. Methods

2.1. The Navier-Stokes Numerical Scheme. The most general macroscopic model for hydraulic flows may be represented by the incompressible Navier-Stokes equations as follows:

$$\rho\left(\frac{\partial \mathbf{u}}{\partial t} - \mathbf{F}\right) = -\nabla p + \mu\nabla^2\mathbf{u}, \tag{1}$$

where ρ is the fluid density, $\mathbf{F}$ is a specific prescribed body force (i.e., gravity force), and μ is the fluid dynamic viscosity.

As in hydraulic flows, at least two phases are always present, water and air; the above system of equations must be applied to a multiphase system. The simulation of multiphase flows is a challenging task due to the inherent complexity of the involved phenomena (i.e., moving interfaces with complex topology) and represents one of the leading edges of computational physics.

The following two main approaches are widely used to simulate multiphase flows: the Euler-Lagrange approach and the Euler-Euler approach. The latter approach, usually chosen in hydraulics applications, is based on the assumption that the volume of a phase cannot be occupied by other phases, thus introducing the concept of phase volume fractions as continuous functions of space and time.

For the present work, we employed the so-called volume of fluid (VOF) method, [31, 32] designed for two or more immiscible fluids where only one fluid (i.e., air) is compressible and the position of the interface between the fluids is of interest. The VOF method is a surface-tracking technique applied to a fixed Eulerian mesh, in which a specie transport equation is used to determine the relative volume fraction of the two phases, or phase fraction, in each computational cell. Practically, a single set of Reynolds-averaged Navier-Stokes equations is solved and shared by the fluids, and, for the additional phase, its volume fraction γ is tracked throughout the domain.

Therefore, the full set of governing equations for the fluid flow are as follows:

$$\begin{gathered}\nabla\cdot\mathbf{u} = 0,\\ \frac{\partial\rho\mathbf{u}}{\partial t} + \nabla\cdot(\rho\mathbf{u}\mathbf{u}) - \nabla\cdot\left((\mu+\mu_t)\,\mathbf{S}\right) = -\nabla p + \rho\mathbf{g} + \sigma K\frac{\nabla\gamma}{|\nabla\gamma|},\\ \frac{\partial\gamma}{\partial t} + \nabla\cdot(\mathbf{u}\gamma) = 0,\end{gathered} \tag{2}$$

where $\mathbf{u}$ is the velocity vector field, p is the pressure field, μ_t is the turbulent eddy viscosity, $\mathbf{S}$ is the strain rate tensor defined by $\mathbf{S} = (\nabla\mathbf{u} + \nabla\mathbf{u}^T)/2$, σ is the surface tension, and K is the surface curvature.

The nature of the VOF method means that an interface between the species is not explicitly computed, but rather emerges as a property of the phase fraction field. Since the phase fraction can have any value between 0 and 1, the interface is never sharply defined, but occupies a volume around the region where a sharp interface should exist.

Table 1: Velocity vectors for the D3Q19 models.

$\mathbf{c_i}$	Direction	Number of cells	f_i^{eq0}
0	0	1	f_1^{eq0}
c	1,. . .,6	6	f_2^{eq0}
$c \cdot \sqrt{2}$	7,. . .,18	12	f_3^{eq0}

The model used in this work is based on an open source computational fluid dynamics (CFD) platform named OpenFOAM [33], freely available on the Internet. OpenFOAM, primarily designed for problems in continuum mechanics, uses the tensorial approach and object-oriented techniques. It provides a fundamental platform to write C++ new solvers for different problems as long as the problem can be written in tensorial partial differential equation form.

2.2. The LBM Numerical Scheme. This section introduces the theoretical background of the proposed LBM with specific regard to the following:

(i) the single-phase LBM and the boundary conditions,

(ii) the front tracking concept,

that are, respectively, inserted in the two following sections.

2.2.1. Single-Phase Lattice Boltzmann Method and Boundary Conditions. Artificial and natural fluvial hydrodynamics may be simulated at the macroscopic scale using the LBM that is structured as a mesoscopic (i.e., scaled larger than micro and smaller than macro) method. LBM treats the water mass by analysing its density probability function, $f_i(\mathbf{x}, \mathbf{c_i}, t)$, where $\mathbf{x}$ is the particle spatial location within the lattice domain, t is the time variable, and $\mathbf{c_i}$ is the kinematic component. The dynamic of the particle distribution function is described in (3) that represents a discrete lattice Boltzmann equation, representing the evolution in time of the fluid flow by providing a numerical solution of the modification of the density probability function:

$$f_i\left(\mathbf{x} + \mathbf{c_i}\Delta t, t + \Delta t\right) - f_i\left(\mathbf{x}, t\right) + \Delta t \cdot \mathbf{F_i} = C_i\left(\mathbf{x}, t\right), \quad i = 1, \ldots, b, \tag{3}$$

where the left and right terms represent, respectively, the molecular free-streaming and the particle collisions. $\mathbf{F_i}$ is an additional term accounting for internal (e.g., collision) and external forcing processes (e.g., electric and magnetic fields or gravity).

For water-driven processes, it is reasonable to assume a discrete set of directions along which particles may route and, as a consequence, $\mathbf{c_i}$ is restricted to a maximum of 7 to 9 for 2D or 19 directions for 3D flows. This 19-direction discretized cubic lattice model implemented here (namely, D3Q19) is characterized by the set of predefined velocity vectors schematically represented in Figure 1 and reported in Table 1.

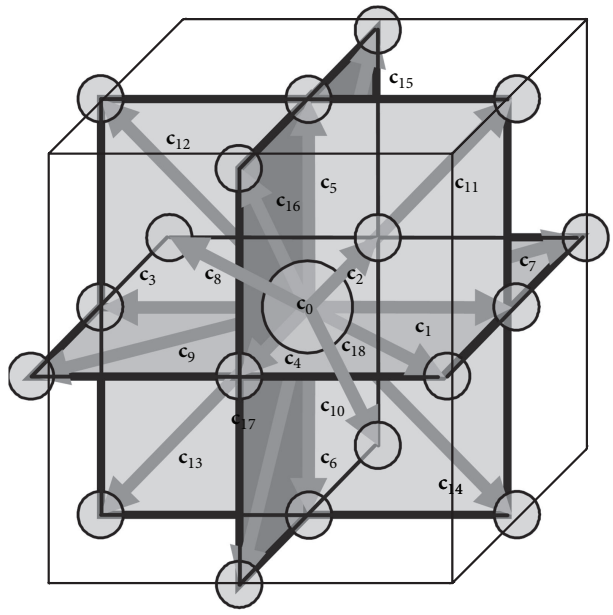


Figure 1: D3Q19 velocity model.

The collision component is integrated by implementing the Bhatnagar-Gross-Krook Model (BGK) [34]

$$C_i\left(\mathbf{x}, t\right) = -\frac{\Delta t}{\tau}\left[f_i\left(\mathbf{x}, t\right) - f_i^{\mathrm{eq}}\left(\mathbf{x}, t\right)\right] \tag{4}$$

that characterizes the density modifications caused by particle collisions as related to dynamic modifications from the local equilibrium f_i^{eq} [35, 36].

If the relative Knudsen and Mach numbers are small enough, this formulation is able to represent the dynamics of a fluid with pressure $P = c_s^2\rho$ and a kinematic viscosity $\upsilon = c_s^2(\tau - \Delta t/2)$, where c_s is the lattice sound velocity [12].

The proposed numerical approach is implemented on a structured Cartesian lattice (i.e., mesh) where the particles may only reside on the mesh nodes.

In the proposed model boundary nodes are solved using the bounce-back scheme meaning that particles colliding with a stationary wall simply reverse their momentum [12, 37]. In particular, the half-way bounce-back scheme is here applied by posing the reflecting wall at a medium distance between boundary nodes and the physical walls [38]. As a result, the proposed model can simulate both solid features (e.g., walls, levees, etc.) of known geometry or open boundaries of predefined velocity, discharge, or pressure function.

The turbulence flow field is implemented in the proposed LBM model by employing a Large Eddy Simulation (LES) approach [39] that characterizes the flow field simulating large-scale structures (resolved grid scales), which are assumed to be mostly predominating in defining the transport of mass and momentum, and not directly simulating the small-scale local effects. The definition of the boundary between small and large processes is based on the Smagorinsky model [40].

2.2.2. Front-Tracking Model. In the Front-Tracking (FT) formulation proposed here [23], the interface is treated as a zero thickness, mathematical surface, across which the density field jumps from the light to the dense phase and vice versa and is tracked throughout the computational domain through an additional scalar variable (i.e., liquid mass fraction). A single set of LB equations is solved for both the heavy fluids with respect to the light fluids (i.e., gas) in agreement with the front-tracking methodology that do not aim at providing any description of the physics of the phase-transition between the two phases. The main difference, which also represents the main advantage of the LBM front-tracking approach, with respect to the traditional continuum methods, is that the single-species transport equation, used in Navier-Stokes/VOF model to determine the relative volume fraction of the two phases in each computational cell, is not needed, since the free-surface tracking is automatically performed by advancing the fluid elements. In practical terms, only the liquid phase flow is numerically solved, while the gas characterization is neglected, and an additional variable ε (i.e., the liquid volume fraction) is structured to track the free surface throughout the computational domain. This variable is numerically solved using the following equation:

$$\varepsilon(\mathbf{x},t) = \frac{m(\mathbf{x},t)}{\rho(\mathbf{x},t)} : \begin{cases} \varepsilon(\mathbf{x}) = 0, & \forall \mathbf{x} \in G, \\ \varepsilon(\mathbf{x}) = 1, & \forall \mathbf{x} \in L, \\ 0 < \varepsilon(\mathbf{x}) < 1, & \forall \mathbf{x} \in I, \end{cases} \tag{5}$$

in which G, L, and I represent respectively empty (gas), liquid, and interface cells.

The computation scheme is integrated numerically in three steps as follows: (1) interface positioning and flow mass motion update defined by analysing cell fluxes; (2) definition of the boundary conditions at the interface to separate the gas and liquid state; (3) update of cell type. As a result, interface cells defined by the $\mathbf{x}$ position and time t are defined as follows:

$$\Delta m_i = \begin{cases} 0, & \forall \mathbf{x} \in G, \\ f_i^{\mathrm{pc}}(\mathbf{x} + \mathbf{c_i}\Delta t, t) - f_{-i}^{\mathrm{pc}}(\mathbf{x},t), & \forall \mathbf{x} \in L, \\ \left(\dfrac{\varepsilon(\mathbf{x},t) + \varepsilon(\mathbf{x} + \mathbf{c_i}\Delta t, t)}{2} \right) & \\ \quad \times \left[f_i^{\mathrm{pc}}(\mathbf{x} + \mathbf{c_i}\Delta t, t) - f_{-i}^{\mathrm{pc}}(\mathbf{x},t) \right], & \forall \mathbf{x} \in I, \end{cases} \tag{6}$$

in which pc characterizes the postcollision state and the sign is related to the mass direction where $-i$ has opposite direction with respect to i. Mass dynamics in space and time are defined as

$$m(\mathbf{x}, t + \Delta t) = m(\mathbf{x},t) + \sum_{i=1}^{18} \Delta m_i(\mathbf{x},t). \tag{7}$$

Combining (6) and (7), the continuity principium is in respect with the condition that direct modification of state, from liquid to gas or vice versa, is not permitted outside the interface. In fact, liquid and gas cells are only allowed to transform into interface cells, whereas interface cells can be transformed into both gas and liquid cells. In this way, mass and density are completely decoupled and, as a results, mass evolution does not affect the particle distribution functions.

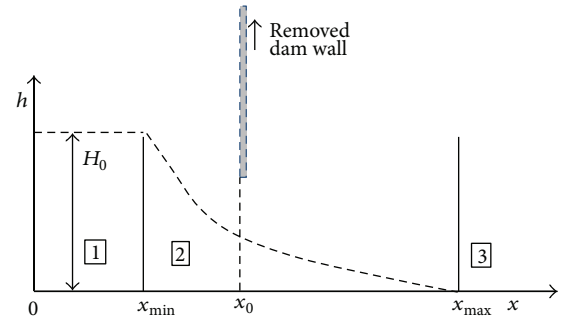


Figure 2: Schematic representation of the water surface profile for a 20 m long straight channel dam break.

3. Numerical Results for Three Case Studies

The proposed LBM and NS are applied for the following three test cases:

(a) the free-surface flow routed along a straight channel, with LBM modeling results that are validated by comparison with two different types of analytical solutions;

(b) the free-surface flow routed along a straight channel deriving from the partial asymmetrical collapse of a vertical dam embankment with results that are compared in terms of hydraulics and computational efficiency;

(c) the simulation of a severe transient flow through a synthetic urban district.

The three tests are hereafter presented in detail.

3.1. Test (a): Fast Flow Propagating along a Straight Channel. The routing of a fast moving flow resulting from a dam break, propagated through a 20 m long straight channel (see Figure 2), is simulated, and the numerical results are compared to the analytical solution provided by Ritter [41].

The channel bed is horizontal. Bottom and wall friction is set to null. The numerical mesh is made up of 2,000 cubic elements with a grid-spacing of 0.01 m. At time zero, a 0.5 m by 7 m volume of water ideally positioned at the most, upstream section of the channel is released instantaneously and two resulting waves propagate, respectively, downstream and upstream back to the reservoir. At a given time t from the start of the simulation (i.e., dam break/breach or ideal vertical wall removal), three different regions are defined (Figure 2) as follows:

(1) for $x < x_{\min}$, we have $h = H_0 = 0.5$ m,

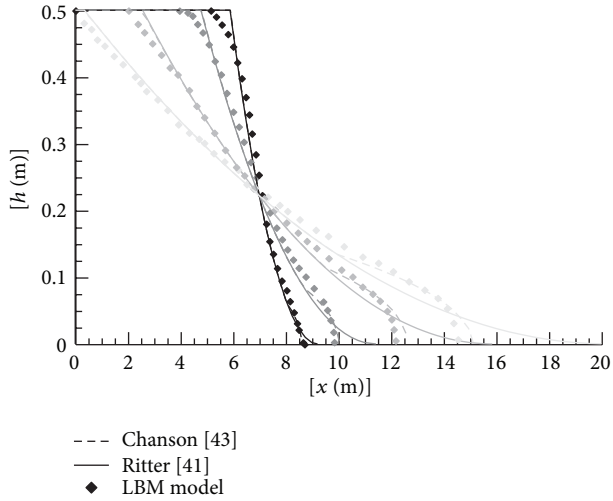


Figure 3: LBM simulation results, analytical solution, and experimental data for the 20 m long straight channel dam break at times 0.5 s, 1 s, 2 s, and 3 s.

(2) for $x_{\min} < x < x_{\max}$, the water surface profile is parabolic

$$h = \frac{H_0}{(x_{\min} - x_{\max})^2}\left(x^2 - 2x_{\max}x + {x_{\max}}^2\right),$$
$$x_{\min} = -t\sqrt{gH_0} + x_0, \quad (8)$$
$$x_{\max} = 2t\sqrt{gH_0} + x_0,$$

(3) for $x > x_{\max}$, we have $h = 0$,

where h is the water elevation. Figure 3 shows the agreement between numerical and analytical data apart some minor expected differences at the boundaries [42]. In fact, where Ritter's analytical scheme is not suited for this comparison, other methods were selected like the Chanson's solution [43]:

$$\frac{8}{3}\frac{1}{f}\frac{(1-U/2)^3}{U^2} = \frac{t}{\sqrt{g/H_0}},$$
$$\frac{x_s}{H_0} = \frac{x_0}{H_0} + \left(\frac{3}{2}U - 1\right)\frac{t}{\sqrt{g/H_0}} + \frac{4}{fU^2}\left(1 - \frac{U}{2}\right)^4,$$
$$\frac{h}{H_0} = \frac{1}{9}\left(2 - \frac{(x-x_0)}{H_0\left(t/\sqrt{g/H_0}\right)}\right)^2$$
$$\text{if } -\frac{t}{\sqrt{g/H_0}} \le \frac{(x-x_0)}{H_0} \le \left(\frac{3}{2}U - 1\right)\frac{t}{\sqrt{g/H_0}},$$
$$\frac{h}{H_0} = \sqrt{\frac{f}{4}U^2\frac{(x_s - x)}{H_0}}$$
$$\text{if } \left(\frac{3}{2}U - 1\right)\frac{t}{\sqrt{g/H_0}} \le \frac{(x-x_0)}{H_0} \le \frac{(x_s - x_0)}{H_0}, \quad (9)$$

that is also plotted in Figure 3.

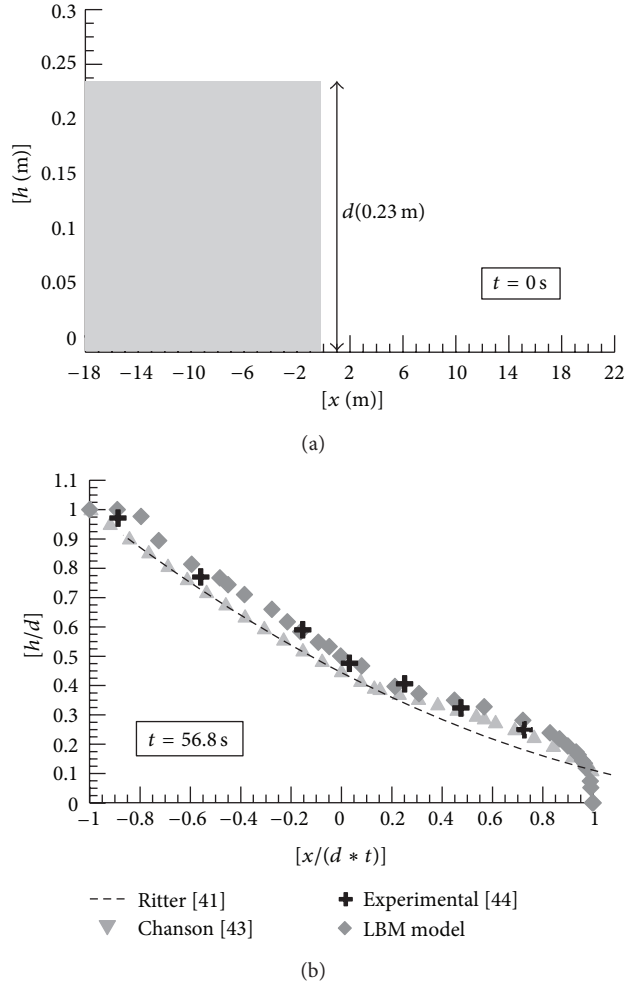


Figure 4: Comparison between numerical LBM and analytical and experimental results at time 56.8 s for the 40 m long straight channel dam break.

To further prove the validity of the code, a dam break flow propagating along a 40 m long straight channel is simulated. At time zero, a 0.23 m by 18 m water volume is released instantaneously. In Figure 4, LBM numerical results are compared to the Ritter [41] and Chanson [43] as well as to the experimental measurements provided by [44]. The comparison is performed in terms of dimensionless instantaneous free-surface profiles (h/d) between numerical and analytical data (assuming $f = 0.02$).

3.2. Test (b): Asymmetrical Dam Break—LBM and NS Numerical Models. The second test is characterized by the simulation of a submersion wave caused by the partial collapse of a nonsymmetrical sluice gate (or partial asymmetric dam breach). The spatial 3D computational domain is 200 m × 200 m × 20 m, where the boundaries of the domain are impermeable walls. The nonuniform initial condition for the liquid phase fraction is specified: the water surface level is initially 10 m for the upstream section and 5 m for the downstream section. An unsteady flow is generated by

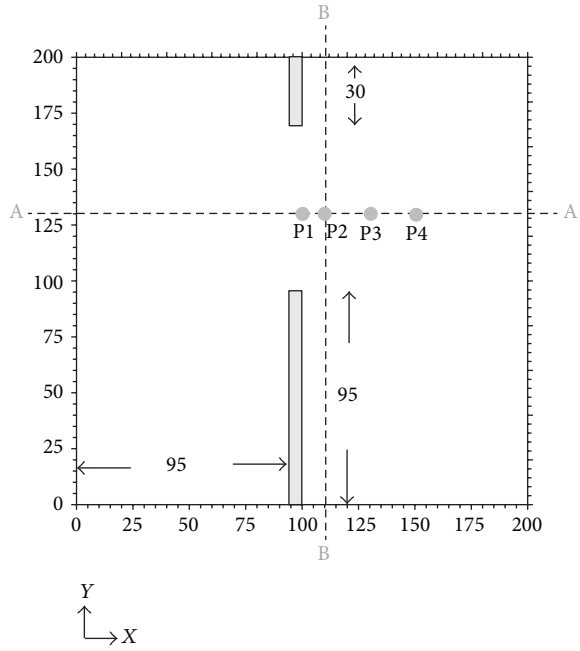


Figure 5: Geometric schematization in plan view for test (b). Sections A-A and B-B and Points P1 (100, 130), P2 (110, 130), P3 (130, 130), and P4 (150, 130) are shown. All dimensions are in m.

the instantaneous collapse of an asymmetrical 75 m long portion of the dam (see Figure 5). The bottom is flat, and ground resistance to the motion is neglected.

Numerical results of several authors are available in the literature for this case [6, 45, 46], that is, a standard benchmark for hydraulic models in dam break applications. We chose to compare the LBM and NS, described in the previous section, in terms of computational efficiency. Calculations have been performed with a time step of 0.01 s on a computational mesh of 800,000 nodes (i.e., lattice sites for the LBM). Figure 6 shows the comparison of the computed water level 7.2 seconds after the breach, when the flow reaches the left side of the tank, while Figure 7 shows the simulated water surface levels along sections A-A (longitudinal $Y = 130$ m) and B-B (transverse $X = 110$ m). Water level hydrographs at selected points of the domain have been also evaluated, and Figure 8 shows the results obtained for points P1 and P3 with both LBM and NS approaches.

Results show that NS and LBM, using the same grid resolution, are in good agreement in terms of water heights and hydrograph, along the section AA, while the results for the section BB (Figure 7(b)) are remarkably different especially in terms of water depths. This is expected with specific regard to the central part of the domain where the dam break originates and the flow is significantly 3D, while in the other portions of the domain, through which the emptying of the system applies (from the noncollapsing area towards where the breach is) in the direction of the flow that is parallel to the gate, the two models give different results with greater water depths evaluated by NS.

From the computational point of view, using the same hardware, the LBM model has completed the tests in one-fifth of the time needed by the NS. Results are summarized in Table 2 that also shows that the LBM code is able to provide a higher level of details of the flow pattern using the same computational time or to provide the same accuracy and details in less time. Figure 9, synthesizing the LBM versus NS computational efficiency comparison test, shows that the LBM model can use a mesh 65% denser than NS one (ΔX(LBM) = 0.35 cm, ΔX(NS) = 1 m) for the same computational time.

Table 2: Computational efficiency of LBM and NS.

Mathematical Model	Δx	Computational time (s)	Simulation time (s)
NS	1.0	6941	7.2
LBM	1.0	1512	7.2
LBM	0.8	3011.04	7.2
LBM	0.6	4676.4	7.2
LBM	0.4	6350	7.2
LBM	0.3	7300.08	7.2

3.3. Test (c): Severe Transient Flow through a Synthetic Urban District. This test presents the simulation of the effects of a severe transient flow on a synthetic urban area. This case study replicates the experiment developed at the Hydraulic Laboratory of the Civil and Environmental Engineering Department of the Université Catholique de Louvain in Belgium [47]. The geometric configuration is characterized by a regular distribution of rectangular obstacles (buildings) intercepting the flow in geometric domain. The realistic average ratio between building and street widths of 1 to 3 was selected by interpreting a real urban configuration (the aerial view of Brussels in Belgium). Buildings' heights are simulated as high enough in order not to be overtopped by the flow. The experimental physical setup consists of a 36 m long by 3.6 m wide channel, with a flat profile (Figure 10). Buildings, wooden parallelepipeds of 30 cm by 30 cm, and streets that are 10 cm wide are positioned along the flow direction. The initial reservoir water level is 0.40 m upstream. Water surface measurements were performed [47] by means of several resistive level gauges for estimating the water surface geometric and dynamic properties (depths and surface velocity fields).

The aim of this test is to investigate the potential of LBM and NS to correctly represent the complex dynamics of transient flows. In particular, the physics of interest is related to those flood phenomena propagated through urban setting for which an initial severe transient phase is often followed by a slow transient that could approximate under certain steady state conditions. Nevertheless, since the first instants correspond to a severe transient flow involving specific flow features characterized by high velocities and transcritical flows, wave impacts on structures, or higher water rise following wave reflection against a structure, the LBM and NS are challenged to provide such unsteady irregular behavior followed by the regular steady conditions of the following sequences.

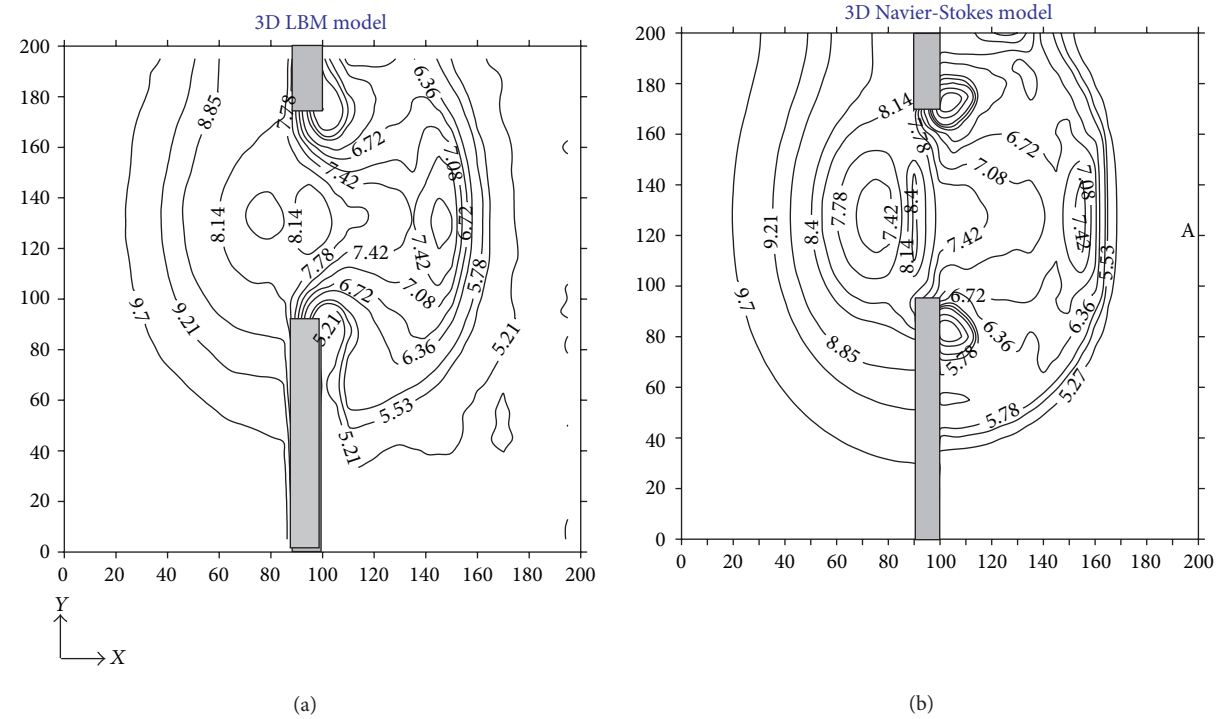


Figure 6: Comparison of the computed water level 7.2 seconds after the breach—LBM versus and NS are, respectively, plotted in the left and right.

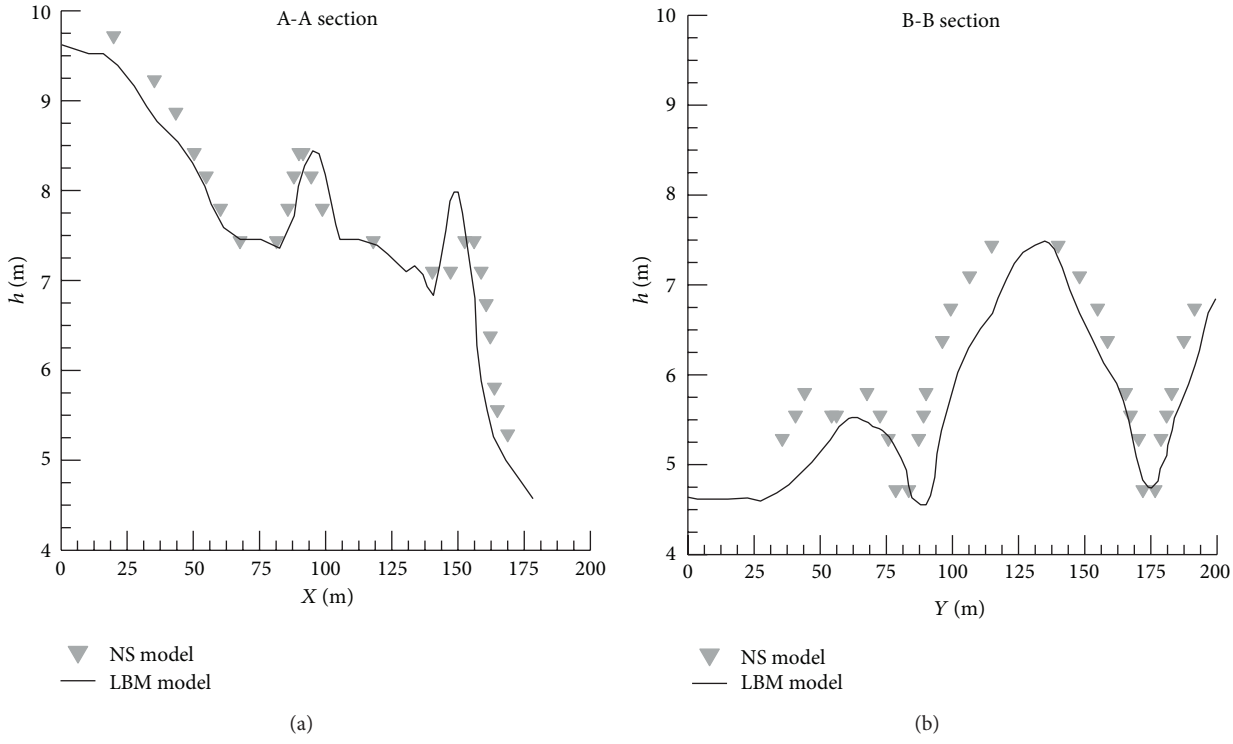


Figure 7: Comparison between water surface elevations obtained with LBM and NS in the A-A and B-B sections at 7.2 seconds after the failure.

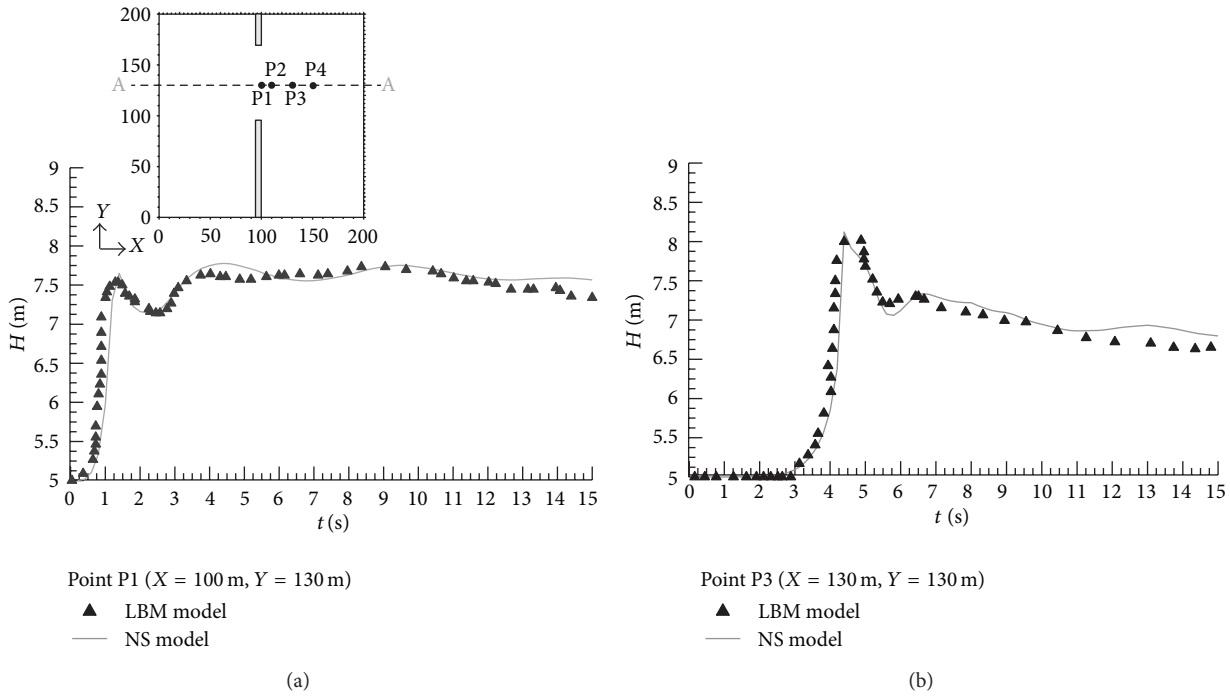


Figure 8: Water level hydrographs at points P1 (100, 130) and P3 (130, 130) for test (b). Continuous line LBM, scatter points for NS.

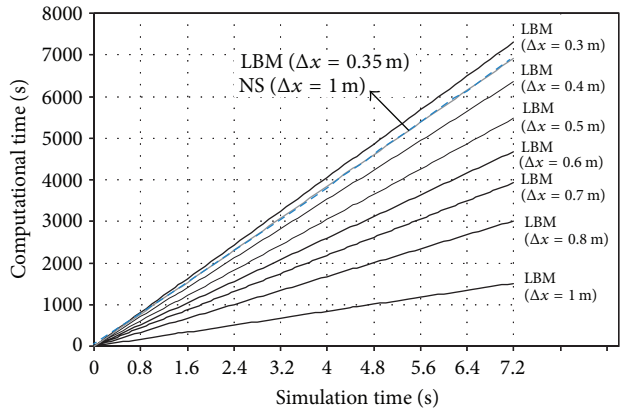


Figure 9: Computational comparison test for and LBM models for test (b).

3.3.1. Simulation Setup—Mesoscopic Approach Using LBM. Due to the large-scale of the domain in horizontal XY plane, the test case has been replicated using a computational mesh of cubes of 0.025 m resolution producing a total number of 4,976,640 nodes. The downstream region has been represented assigning a liquid volume fraction of 0.5 to the first vertical row of nodes. Considering the dimension of the domain, the mesh is quite rough and only few computational nodes describe the streets that are only 0.1 m wide.

3.3.2. Simulation Setup—Continuous Approach Using NS. For all simulations, the computational domain is the same with the following dimensions: 36 m long, 3.6 m wide, and 1 m high. Five scenarios have been developed for the NS simulation, varying the resolution of the mesh in the three directions. The computational mesh for each scenario derives from the intersection and refinement of the mesh composed of hexahedra (configuration A, see Table 3) with the 3D surface of the synthetic urban setting (Figure 11). In configuration A a grid variation along the vertical with a ratio of 1 to 10 is used. The result of such a refinement process is represented by configuration B (see Table 3). For the five scenarios, M1 to M5, the diverse grid resolutions lead to discrepancies between the numerical results and physical measurements both inside and outside the city. The comparison between experimental data and numerical results provided by the LBM and the NS, scenarios M1, M2, M5, for the section $y = 2$ m at $t = 5$, 6 and 10 s after the break are presented in Figure 12.

The free-surface profile includes a hydraulic jump at 5 s after the break. In the graphs for $t = 5$ s and $t = 6$ s, when the flow depth in the city is still low, the flow reflection against the buildings is clearly visible for both LBM and NS simulations,

Table 3: Geometric characteristics of computational meshes adopted for scenarios M1, M2, M3, M4, and M5.

	M1	M2	M3	M4	M5
Number of TOT cells	$3.24E+04$	$2.59E+05$	$5.18E+05$	$1.04E+06$	$2.07E+06$
Min volume cell (m^3)	$9.79E-04$	$1.25E-04$	$6.33E-05$	$3.16E-05$	$1.58E-05$
Max volume cell (m^3)	$9.79E-03$	$1.25E-03$	$6.33E-04$	$3.16E-04$	$1.58E-04$
Number of points	$3.78E+04$ ($180 \times 18 \times 10$)	$2.80E+05$ ($360 \times 36 \times 20$)	$5.48E+05$ ($360 \times 36 \times 40$)	$1.08E+06$ ($360 \times 72 \times 40$)	$2.16E+06$ ($720 \times 72 \times 40$)
Number of faces	$1.02E+05$	$7.98E+05$	$1.58E+06$	$3.15E+06$	$6.30E+06$
Number of TOT cells	$6.16E+04$	$4.01E+05$	$7.96E+05$	$1.12E+06$	$2.02E+06$
Min volume cell	$6.10E-06$	$1.50E-06$	$7.70E-07$	$3.10E-06$	$1.25E-05$
Max volume cell	$9.79E-03$	$1.25E-03$	$6.33E-04$	$3.16E-04$	$1.58E-04$
Number of points	$8.35E+04$	$4.84E+05$	$9.37E+05$	$1.21E+06$	$2.12E+06$
Number of faces	$2.07E+05$	$1.29E+06$	$2.53E+06$	$3.44E+06$	$6.17E+06$

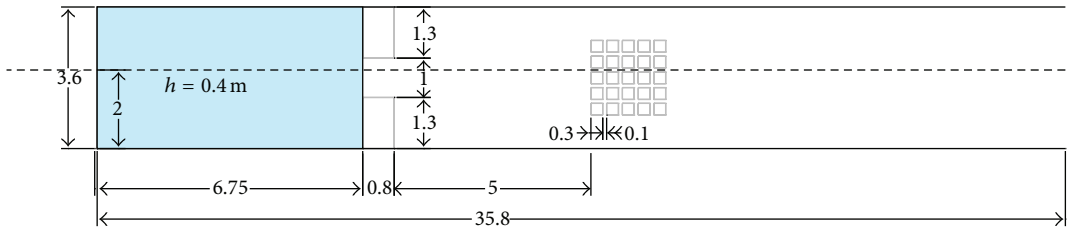


Figure 10: Schematic representation of experimental and computational domain.

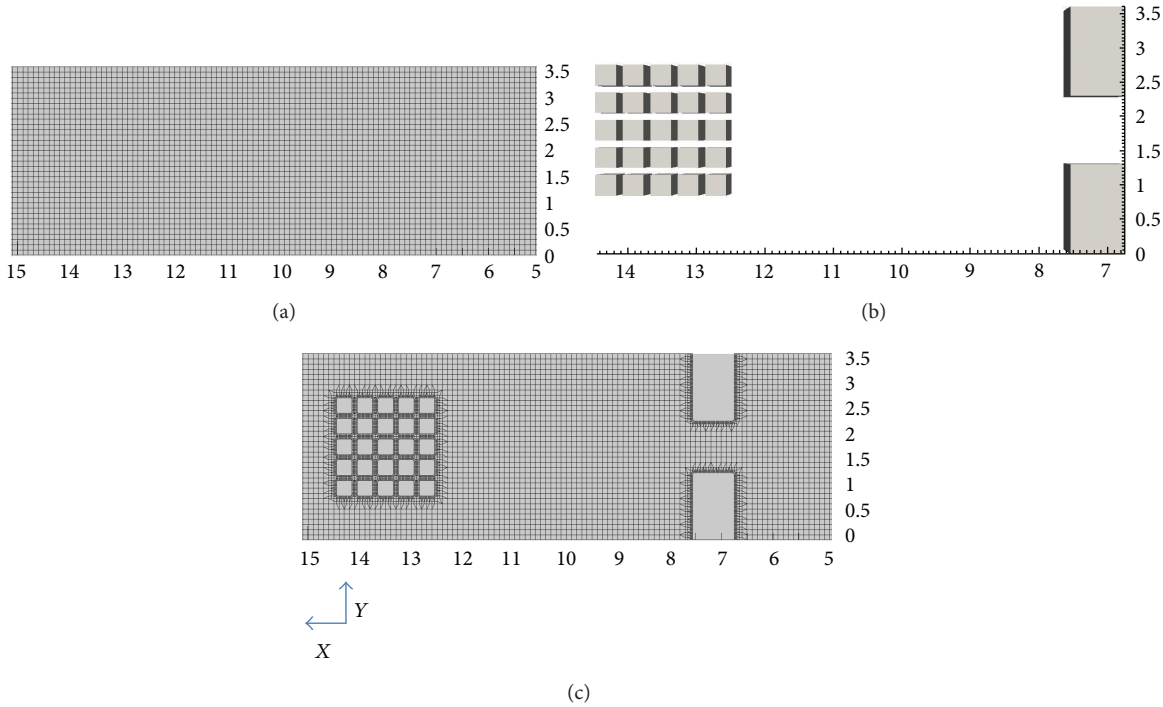


Figure 11: Computational mesh definition procedure. Example of intersection between 3D obstacle surface and original mesh.

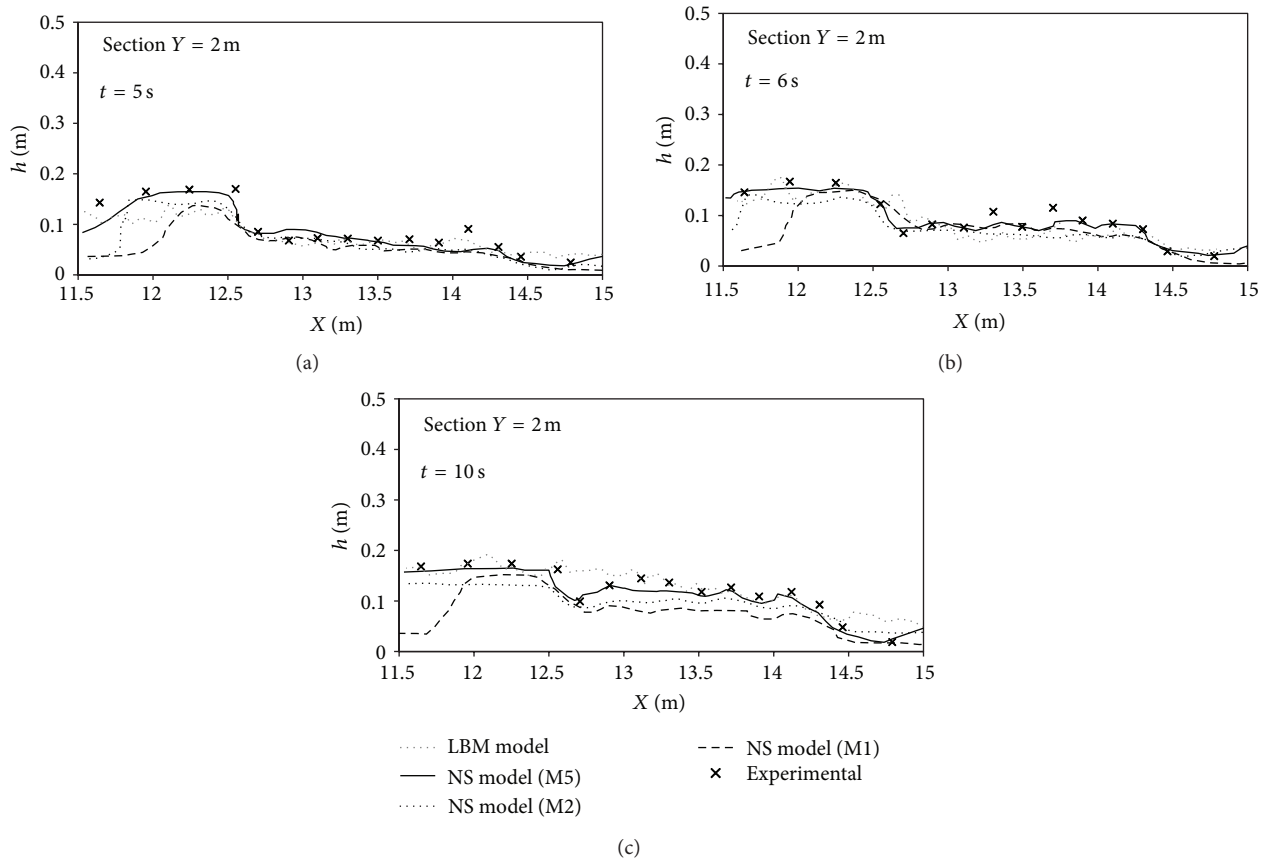


Figure 12: Comparison between experimental and simulated free-surface value for the section $Y = 2$ m for scenarios M1, M5 of continuous approach, and LBM for $t = 5$, 6, and 10 s after the break.

and the LBM model correctly predicts the free surface at low x values as the NS model with the finest mesh. At $t = 10$ s, the upstream hydraulic jump is still present, but the water level has significantly increased in the streets. The flow in the streets evolves from supercritical Fr > 1 (with a control section near the street entrance) to subcritical (with a control section at the street exit).

The overall agreement between the numerical and experimental results is promising for the use of the mesoscopic numerical model for replicating the complex physics of rapidly varying flows in such a dense urban area, even if, due to the mesh resolution, results of lattice-based simulations are less accurate than the NS because of the mesh refinement (especially for M5). Analyzing the effect of mesh refinement, the obstacle mesh refinement of M3–M5 is able to better reproduce the unsteady hydraulic features and flow processes like hydraulics jumps, while the coarser meshes of M1-M2 are characterized by some errors where the flow behavior is irregular.

For LBM at the time steps $t = 5$ and 6 s, significant differences appear between the two numerical models. For example, the depression and the hydraulic jump at the entrance of the urban district ($x = 5.20$ m) is absent in the LBM coarse mesh result. One limitation of the used LBM code is that the computational mesh is not refined locally for example in a single area of interest. This disadvantage with respect to the used NS model clearly indicates the utility of employing mesh refinement techniques, as those developed for standard LBM [48].

Figure 13 shows the results at $t = 4$ seconds for 5 different meshes, while in Figure 14 snapshots of free-surface evolution at different time steps (0, 2, 4, 6, 8, and 10 s) are inserted. Figure 15 reports the value of velocity magnitude for the same time steps.

Observations indicated that the flow rises at the city front before entering the streets, after the wave impact, a phenomenon that is similar to the impact against a single obstacle rather than a series of obstacles. A hydraulic jump is represented at the impact section (Figure 14), with the water level that is locally higher with respect to the same case without buildings, while a wake zone is developed immediately downstream of the city.

4. Conclusions

In this study an experimental fully 3D flow modeling framework is implemented to numerically simulate the unsteady irregular flows originating from dam breaks with specific regard to the complex interaction between the geometrical urban features and the associated physical hydraulic processes. The proposed novel technique is based on a front-tracking variant of the lattice Boltzmann method that is here

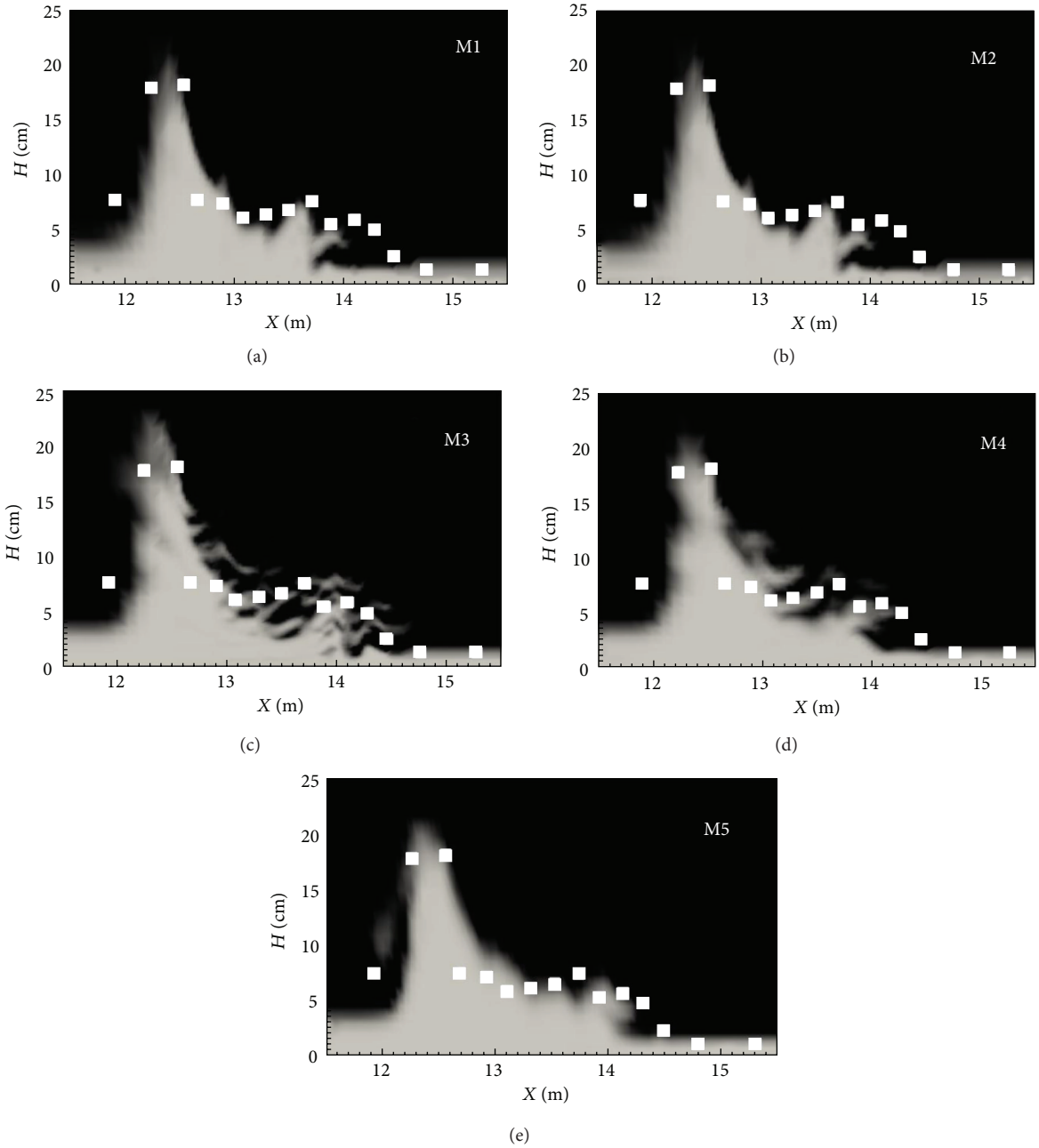


Figure 13: Comparison of experimental free-surface value (points) and simulated fraction fill results for 6 different computational meshes at time = 4 s from the release at section $Y = 2$ m.

investigated as a valid alternative to the standard numerical Navier-Stokes model in the treatment of instantaneously varying flows interacting with artificial structures. In particular, test cases evaluate the accuracy and the computational efficiency of numerical simulations with respect to analytical solutions and experimental data. The critical analysis of model performances provides the following conclusive remarks.

The proposed LBM is able to predict with a high level of accuracy typical hydraulic phenomena characterizing engineering applications related to the flooding of simple straight channels (test (a)) and asymmetrical front waves in rectangular basins (test (b)). More specifically, as compared to the fully 3D NS model, it is reported the validity and robustness of LBM results that, while implementing the same resolution of the geometric domain, are demonstrated to be more computationally efficient (test (b)).

The last test investigates the diverse hydraulic unsteady processes, like the hydraulic jump, return, and diverged flow, that originate from the passage of the flood wave through a synthetic urban setting (test (c)). The application of the presented 3D numerical schemes shows the major advantages

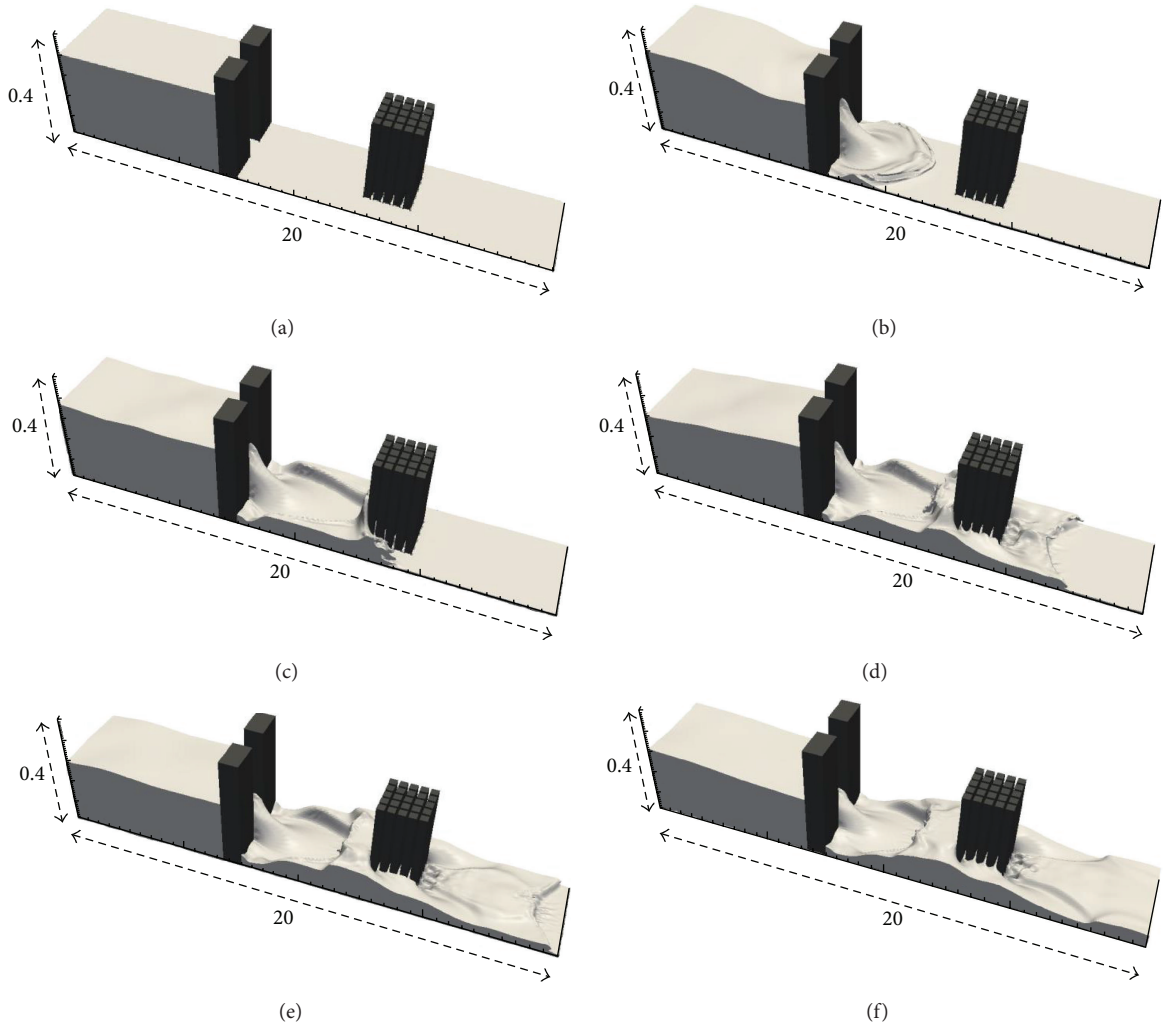


Figure 14: Free-surface evolution for scenario M5 for different time steps ((a) = 0 s, (b) =2 s, (c) = 4 s, (d) = 6 s, (e) = 8 s, and (f) =10 s).

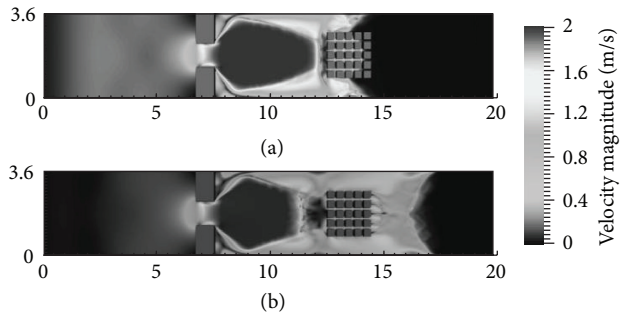


Figure 15: Evolution of velocity magnitude for scenario M5 at different time steps—t = 4 s (a), and 6 s (b).

and disadvantages of both LBM and NS schemes confirming the ability of LBM in representing such abruptly changing and irregular flow dynamics. Nevertheless, the mesoscopic LBM approach is significantly bound by the regular structured mesh, while the continuous NS model performs well, also in large domains, especially when using a more accurate grid (M5) that is adaptively refined to more accurately represent obstacles.

References

[1] P. Manciola, S. Di Francesco, and C. Biscarini, "Flood protection and risk management: the case of Tescio River basin," in *Proceedings of the Role of Hydrology in Water Resources Management Symposium*, vol. 327 of *IAHS Red Book Series*, pp. 174–183, Capri, Italy, October 2008.

[2] Y. Jia and S. S. Y. Wang, "Numerical model for channel flow and morphological change studies," *Journal of Hydraulic Engineering*, vol. 125, no. 9, pp. 924–933, 1999.

[3] P. G. Manciola, A. Mazzoni, and F. Savi, "Formation and propagation of steep waves: an investigative experimental interpretation," in *Proceedings of the Specialty Conference on Modelling of Flood Propagation Over Initially Dry Areas*, pp. 283–297, Milano, Italy, July 1994.

[4] C. Biscarini, "Computational fluid dynamics modeling of landslide generated water waves," *Landslide Springer*, vol. 7, no. 2, pp. 117–124, 2010.

[5] A. Ferrari, L. Fraccarollo, M. Dumbser, E. F. Toro, and A. Armanini, "Three-dimensional flow evolution after a dam break," *Journal of Fluid Mechanics*, vol. 663, pp. 456–477, 2010.

[6] C. Biscarini, S. Di Francesco, and P. Manciola, "CFD modelling approach for dam break flow studies," *Hydrology and Earth System Sciences*, vol. 14, no. 4, pp. 705–718, 2010.

[7] D. Dutykh and D. Mitsotakis, "On the relevance of the dam break problem in the context of nonlinear shallow water equations," *Discrete and Continuous Dynamical Systems B*, vol. 13, no. 4, pp. 799–818, 2010.

[8] H. Liu, J. G. Zhou, and R. Burrows, "Lattice Boltzmann simulations of the transient shallow water flows," *Advances in Water Resources*, vol. 33, no. 4, pp. 387–396, 2010.

[9] C. L. Mader, "Modeling the 1958 Lituya Bay Mega-Tsunami, II," *Science of Tsunami Hazards*, vol. 20, pp. 241–250, 2002.

[10] C. Biscarini and M. Testa, "Three-dimensional numerical modelling of the marmore water falls," *Progress in Computational Fluid Dynamics*, vol. 11, no. 2/2011, pp. 105–115, 2011.

[11] R. Benzi, S. Succi, and M. Vergassola, "The lattice Boltzmann equation: theory and applications," *Physics Report*, vol. 222, no. 3, pp. 145–197, 1992.

[12] S. Succi, *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond*, Numerical Mathematics and Scientific Computation, The Clarendon Press, Oxford, UK, 2001.

[13] S. Ubertini, P. Asinari, and S. Succi, "Three ways to lattice Boltzmann: a unified time-marching picture," *Physical Review E*, vol. 81, no. 1, Article ID 016311, 2010.

[14] J. F. Chau, D. Or, and M. C. Sukop, "Simulation of gaseous diffusion in partially saturated porous media under variable gravity with lattice Boltzmann methods," *Water Resources Research*, vol. 41, no. 8, Article ID W08410, 2005.

[15] X. Shan and H. Chen, "Lattice Boltzmann model for simulating flows with multiple phases and components," *Physical Review E*, vol. 47, pp. 1815–1819, 1993.

[16] D. Chiappini, G. Bella, S. Succi, F. Toschi, and S. Ubertini, "Improved lattice Boltzmann without parasitic currents for Rayleigh-Taylor instability," *Communications in Computational Physics*, vol. 7, no. 3, pp. 423–444, 2010.

[17] G. Falcucci, S. Ubertini, and S. Succi, "Lattice Boltzmann simulations of phase-separating flows at large density ratios: the case of doubly-attractive pseudo-potentials," *Soft Matter*, vol. 6, no. 18, pp. 4357–4365, 2010.

[18] Y. Peng, J. G. Zhou, and R. Burrows, "Modeling free-surface flow in rectangular shallow basins by using lattice boltzmann method," *Journal of Hydraulic Engineering*, vol. 137, no. 12, pp. 1680–1685, 2012.

[19] G. Falcucci, G. Bella, G. Chiatti, S. Chibbaro, M. Sbragaglia, and S. Succi, "Lattice Boltzmann models with mid-range interactions," *Communications in Computational Physics*, vol. 2, no. 6, pp. 1071–1084, 2007.

[20] C. Biscarini, S. Di Francesco, F. Nardi, and P. Manciola, "Three-dimensional hydraulic modelling of unsteady free-surface flows using the Lattice Boltzmann Method," *Journal of Hydraulic Engineering*. In review.

[21] J. G. Zhou, *Lattice Boltzmann Methods For Shallow Water flows*, Springer, 2004.

[22] Y. Li and P. Huang, "A coupled lattice Boltzmann model for the shallow water-contamination system," *International Journal for Numerical Methods in Fluids*, vol. 59, pp. 195–213, 2009.

[23] C. Körner, M. Thies, T. Hofmann, N. Thürey, and U. Rüde, "Lattice Boltzmann model for free surface flow for modeling foaming," *Journal of Statistical Physics*, vol. 121, no. 1-2, pp. 179–196, 2005.

[24] N. Thürey, T. Pohl, U. Rüde, M. Öchsner, and C. Körner, "Optimization and stabilization of LBM free surface flow simulations using adaptive parameterization," *Computers and Fluids*, vol. 35, no. 8-9, pp. 934–939, 2006.

[25] C. Biscarini, S. D. Francesco, and M. Mencattini, "Application of the lattice Boltzmann method for large-scale hydraulic problems," *International Journal of Numerical Methods for Heat and Fluid Flow*, vol. 21, no. 5, pp. 584–601, 2011.

[26] G. Falcucci, S. Ubertini, C. Biscarini et al., "Lattice boltzmann methods for multiphase flow simulations across scales," *Communications in Computational Physics*, vol. 9, no. 2, pp. 269–296, 2011.

[27] H. M. Solo-Gabriele and F. E. Perkins, "Streamflow and suspended sediment transport in an urban environment," *Journal of Hydraulic Engineering*, vol. 123, no. 9, pp. 807–811, 1997.

[28] S. Soares-Frazão and Y. Zech, "Experimental study of dam-break flow against an isolated obstacle," *Journal of Hydraulic Research*, vol. 45, pp. 27–36, 2007.

[29] S. Di Francesco, C. Biscarini, and P. Manciola, "Mesoscopic modeling of dam break flow," in *Proceedings of the 3rd International Junior Researcher and Engineer Workshop on Hydraulic Structures (IJREWHS '10)*, R. Jannsenn and H. Chanson, Eds., Hydraulic Model Report CH80/10, School of Civil Eng., The University of Queensland, Brisbane, Australia, 2010.

[30] D. Yu and S. N. Lane, "Urban fluvial flood modelling using a two dimensional diffusion-wave treatment: 1. Mesh resolution effects," *Hydrological Processes*, vol. 20, pp. 1541–1565, 2006.

[31] S. Zaleski, J. Li, and S. Succi, "Two-dimensional Navier-Stokes simulation of deformation and breakup of liquid patches," *Physical Review Letters*, vol. 75, pp. 244–247, 1995.

[32] O. Ubbink and R. I. Issa, "A method for capturing sharp fluid interfaces on arbitrary meshes," *Journal of Computational Physics*, vol. 153, no. 1, pp. 26–50, 1999.

[33] *Open FOAM, User Guide, Version 1.5*, The Open Source CFD Toolbox, 2008.

[34] P. L. Bhatnagar, E. P. Gross, and M. Krook, "A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems," *Physical Review*, vol. 94, no. 3, pp. 511–525, 1954.

[35] S. Chen and G. D. Doolen, "Lattice Boltzmann method for fluid flows," *Annual Review of Fluid Mechanics*, vol. 30, pp. 329–364, 1998.

[36] F. Tosi, S. Ubertini, S. Succi, H. Chen, and I. V. Karlin, "Numerical stability of entropic versus positivity-enforcing lattice Boltzmann schemes," *Mathematics and Computers in Simulation*, vol. 72, no. 2-6, pp. 227–231, 2006.

[37] P. A. Skordos, "Initial and boundary conditions for the lattice Boltzmann method," *Physical Review E*, vol. 48, no. 6, pp. 4823–4842, 1993.

[38] T. Inamuro, M. Yoshino, and F. Ogino, "A non-slip boundary condition for lattice Boltzmann method," *Journal of Computational Physics* , vol. 155, p. 307, 1999.

[39] B. Galperin and S. A. Orszag, *Large Eddy Simulation of Complex Engineering and Geophysical Flows*, Cambridge University Press, New York, NY, USA, 1993.

[40] J. Smagorinsky, "General circulation experiments with the primitive equations. I. The basic experiment," *Monthly Weather Review*, vol. 91, pp. 99–164, 1963.

[41] A. Ritter, "The propagation of water waves," *Ver Deutsch ingenieur zeitschr*, vol. 36, no. 33, part 3, pp. 947–954, 1892.

[42] R. Dressler, "Comparison of theories and experiments for the hydraulic dam-break wave," *Proceedings of the International Association of Scientific Hydrology Assemblée Générale, Rome, Italy*, vol. 3, no. 38, pp. 319–328, 1954.

[43] H. Chanson, "Application of the method of characteristics to the dam break wave problem," *Journal of Hydraulic Research*, vol. 47, no. 1, pp. 41–49, 2009.

[44] Y. Cavaillé, *Contribution à l'étude de l'écoulement variable accompagnant la vidange brusque d'une retenue*, vol. 410 of *Publications Scientifiques et Techniques du Ministère de l'air*, Paris, France, 1965.

[45] F. Alcrudo and P. Garcia-Navarro, "High-resolution Godunov-type scheme in finite volumes for the 2D shallow-water equations," *International Journal for Numerical Methods in Fluids*, vol. 16, no. 6, pp. 489–505, 1993.

[46] R. J. Fennema and M. H. Chaudhry, "Explicit methods for 2-D transient free-surface flows," *Journal of Hydraulic Engineering*, vol. 116, no. 8, pp. 1013–1034, 1990.

[47] S. Soares-Frazão and Y. Zech, "Dam-break flow through an idealised city," *Journal of Hydraulic Research*, vol. 46, no. 5, pp. 648–658, 2008.

[48] O. Filippova and D. Hänel, "Grid Refinement for Lattice-BGK Models," *Journal of Computational Physics*, vol. 147, no. 1, pp. 219–228, 1998.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 316501, 6 pages
http://dx.doi.org/10.1155/2013/316501

Research Article

Strong List Edge Coloring of Subcubic Graphs

Hongping Ma,[1] Zhengke Miao,[1] Hong Zhu,[1] Jianhua Zhang,[2] and Rong Luo[3]

[1] *School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou 221116, China*
[2] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*
[3] *Department of Mathematics, West Virginia University, Morgantown, WV 26506–6310, USA*

Correspondence should be addressed to Hongping Ma; hpma@163.com

Received 9 February 2013; Accepted 31 March 2013

Academic Editor: Carlo Cattani

We study strong list edge coloring of subcubic graphs, and we prove that every subcubic graph with maximum average degree less than 15/7, 27/11, 13/5, and 36/13 can be strongly list edge colored with six, seven, eight, and nine colors, respectively.

1. Introduction

All graphs in this paper are finite and simple. For a graph G with vertex set $V(G)$ and edge set $E(G)$, a proper edge coloring of G is an assignment of colors to the edges of G so that no two adjacent edges receive the same color. A strong edge coloring is a proper edge coloring so that two edges adjacent to a common edge receive different colors. The strong chromatic index of G, denoted by $\chi_s'(G)$, is the minimum number of colors needed for a strong edge coloring of G.

Strong edge coloring was introduced by Fouquet and Jolivet [1, 2]. This type of coloring can be used to represent the conflict-free channel assignment in radio networks.

Denote by Δ the maximum degree of the graph. In 1985, Erdös and Nešetřil conjectured that the strong chromatic index of a graph is at most $(5/4)\Delta^2$ when Δ is even and $(1/4)(5\Delta^2 - 2\Delta + 1)$ when Δ is odd. Andersen proved the conjecture for $\Delta = 3$ [3]. Strong edge coloring of cubic Halin graphs has been studied in [4, 5].

Let $\text{mad}(G) = \max_{H\subseteq G, |V(H)|\geq 1}(2|E(H)|/|V(H)|)$ be the maximum average degree of the graph G. Hocquard and Valicov [6] considered the subcubic graphs with bounded maximum average degree, and they proved the following results.

Theorem 1. *Let G be a subcubic graph.*

(i) *If* $\text{mad}(G) < 15/7$, *then* $\chi_s'(G) \leq 6$.

(ii) *If* $\text{mad}(G) < 27/11$, *then* $\chi_s'(G) \leq 7$.

(iii) *If* $\text{mad}(G) < 13/5$, *then* $\chi_s'(G) \leq 8$.

(iv) *If* $\text{mad}(G) < 36/13$, *then* $\chi_s'(G) \leq 9$.

The main purpose of this paper is to generalize the study of list version so that the admissible colors on edges are constrained. An edge list L of a graph G is a mapping that assigns a finite set to each edge of G. Denote $L = \{L(e) : e \in E(G)\}$. We say that L is a k-edge list if $|L(e)| \geq k$ for each edge e in G. The graph G is strongly L-edge colorable if there exists a strong edge coloring c of G such that $c(e) \in L(e)$ for every edge e of G. For a positive integer k, a graph G is strongly k-edge choosable if for every k-edge list L, G is strongly L-edge colorable. The strong list chromatic index $\chi_{ls}'(G)$ is the minimum positive integer k for which G is strongly k-edge choosable.

In this paper, we consider strong list edge coloring of subcubic graphs and extend Theorem 1 to the list version. We prove the following theorem.

Theorem 2. *Let G be a subcubic graph.*

(i) *If* $\text{mad}(G) < 15/7$, *then* $\chi_{ls}'(G) \leq 6$.

(ii) *If* $\text{mad}(G) < 27/11$, *then* $\chi_{ls}'(G) \leq 7$.

(iii) *If* $\text{mad}(G) < 13/5$, *then* $\chi_{ls}'(G) \leq 8$.

(iv) *If* $\text{mad}(G) < 36/13$, *then* $\chi_{ls}'(G) \leq 9$.

The paper is organized as follows. In Section 2, we will prove two lemmas which will be applied a lot in the proof of

Theorem 2. Theorem 2 will be proved in Sections 3, 4, 5, and 6.

Before proceeding we introduce some notations and definitions. The degree of a vertex v in a graph is denoted by $d(v)$. A vertex of degree k is called a k-vertex. A k-neighbor of v is a k-vertex adjacent to v. $N(v)$ is the set of the neighbors of v. A t-thread of G is a path $P_t = x_1x_2\cdots x_t$ with $d(x_i) = 2$ for $i = 1, 2, \ldots, t$. Two edges are at distance at most 2 if either they are adjacent or they are adjacent to a common edge. Denote by $N_2(e)$ the set of edges at distance at most 2 from the edge e. Define $\mathrm{LSC}(N_2(e))$ as the set of colors used by edges in $N_2(e)$. Denote $L'(e) = L(e) \setminus \mathrm{LSC}(N_2(e))$.

2. Lemmas

Lemma 3. *Let G be the graph obtained from a path $x_1x_2x_3x_4x_5$ by adding two vertices u, v so that u is adjacent to x_3 and v is adjacent to x_4. Let L be an edge list of G. If $|L(x_ix_{i+1})| \geq 3$ for $i = 1, 2, 3$, $|L(x_4x_5)| \geq 2$, $|L(x_3u)| \geq 5$, and $|L(x_4v)| \geq 4$, then G has a strong L-edge coloring.*

Proof. If there is a color $a \in L(x_1x_2) \cap L(x_4x_5)$, then we first color the edges x_1x_2 and x_4x_5 with the color a. So we can further color the rest of edges with the available colors in the order of x_2x_3, x_3x_4, x_4v, and x_3u.

Now we assume that $L(x_1x_2) \cap L(x_4x_5) = \emptyset$. Denote $L(x_1x_2) = \{a, b, c\}$ and $L(x_4x_5) = \{d, e\}$. If there is a color $t \in L(x_2x_3) \setminus \{a, b, c\}$, we first color the edge x_2x_3 with the color t. Then after coloring all the other edges, the edge x_1x_2 still has one color available. Therefore, we can color the edges with an available color in the order of x_4x_5, x_3x_4, x_4v, and x_3u. Thus we can further assume $L(x_2x_3) = \{a, b, c\}$. Similarly, we can also assume that $L(x_3x_4) = \{a, b, c\}$. If there is a color $s \in \{a, b, c\} \cap L(x_4v)$, then we first color the edges x_1x_2 and x_4v with the color s. So we can further color the rest of edges with the available colors in the order of x_2x_3, x_3x_4, x_3u, and x_4x_5. Therefore we can assume that $\{a, b, c\} \cap L(x_4v) = \emptyset$. We first color the edges x_1x_2, x_2x_3, x_3x_4, and x_4x_5 with the colors a, b, c, and d, respectively. Then $|L'(x_3u)| \geq 1$, $|L'(x_4v)| \geq 3$, and thus we can further color the edges x_3u and x_4v. This completes the proof of the lemma. □

Lemma 4. *Let $P = xyzuv$ be a path and L an edge list so that $|L(e)| \geq 2$ if $e \in \{xy, zu, uv\}$ and $|L(yz)| \geq 3$. Then P has a strong L-edge coloring.*

Proof. We only need to prove the lemma when each $|L(e)|$ is equal to its lower bound. If there is a color $a \in L(xy) \cap L(uv)$, we color both xy and uv with a. Then $|L'(zu)| \geq 1$ and $|L'(yz)| \geq 2$, so we can further color the edges zu and yz.

Now assume that $L(xy) \cap L(uv) = \emptyset$. Denote $L(xy) = \{a, b\}$ and $L(uv) = \{c, d\}$. If $a \in L(zu)$, we first color zu with a and the edge xy with b and then color the edge yz with an available color. Since $a, b \notin L(uv)$, there is still one color available for the edge uv. Thus P has a strong L-edge coloring. Similarly we can obtain a strong L-edge coloring of P if $L(zu) \cap L(uv) \neq \emptyset$. Now we further assume $L(zu) \cap [L(xy) \cup L(uv)] = \emptyset$. That is, $L(xy)$, $L(zu)$, and $L(uv)$ are mutually disjoint, and it is easy to see that P has a strong L-edge coloring. □

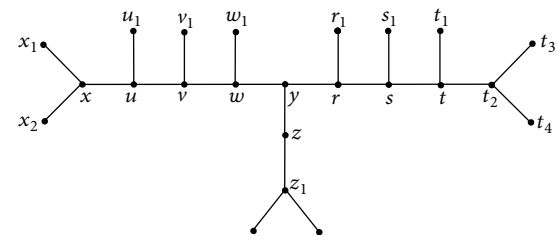


Figure 1: The configuration of Claim 4(1).

3. Proof of (i) of Theorem 2

Let H be a counterexample with $|E(H)|$ as small as possible. Then there exists a 6-edge list L such that H is not strongly L-edge colorable. We can assume that H is connected; otherwise, we can color independently each connected component. A 3-vertex is bad if it is adjacent to a 1-vertex; otherwise it is good.

Claim 1. A 1-vertex is adjacent to a 3-vertex in H and each bad 3-vertex is adjacent to two 3-vertices.

Proof. Let u be a 1-vertex and v its neighbor. Since H is a minimum counterexample, $H \setminus \{uv\}$ has a strong L-edge coloring. If $d(v) = 2$ or v is adjacent to only one 3-vertex, we have $|L'(uv)| \geq 1$, and thus we can easily extend the coloring to H, a contradiction. □

Claim 2. H does not contain a t-thread with $t \geq 3$.

Proof. Suppose that H contains a t-thread $x_1x_2\cdots x_t$ with $t \geq 3$. Then $H' = H \setminus \{x_1x_2, x_2x_3\}$ has a strong L-edge coloring by the minimality of H. It is easy to see that $|L'(x_1x_2)| \geq 2$ and $|L'(x_2x_3)| \geq 2$. Hence we can extend the coloring of H' to H, a contradiction. □

Claim 3. H does not contain a path $x_2x_3x_4x_5$ such that x_2, x_3, x_4, x_5 are all bad 3-vertices.

Proof. Suppose that H contains such a path $x_2x_3x_4x_5$. Let x_1, u, v be the 1-neighbors of x_2, x_3, x_4, respectively. By the minimality of H, $H \setminus [\{x_ix_{i+1} \mid i = 1, \ldots, 4\} \cup \{x_3u, x_4v\}]$ has a strong L-edge coloring f. Since $|L(e)| \geq 6$, we have $|L'(x_ix_{i+1})| \geq 3$ for each $i = 1, 2, 3$, $|L'(x_4x_5)| \geq 2$, $|L'(x_3u)| \geq 5$, and $|L'(x_4v)| \geq 4$. By Lemma 3, we can further extend the coloring to the rest of the edges of H to obtain a strong L-edge coloring of H, a contradiction. □

Claim 4. H does not contain the following three configurations:

(1) a path $xuvwyrstt_2$ such that u, v, w, r, s, and t are bad 3-vertices, y is a good vertex, and another neighbor z of y is a 2-vertex (see Figure 1),

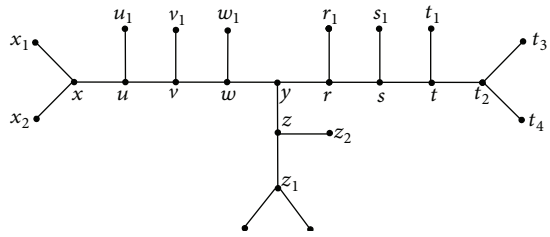


Figure 2: The configuration of Claim 4(2).

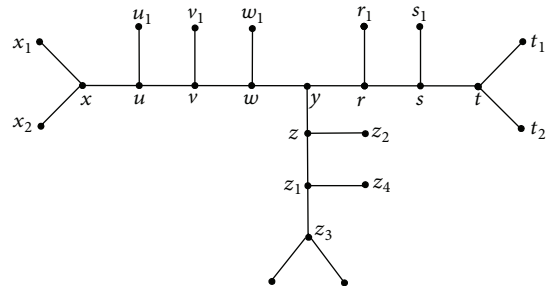


Figure 3: The configuration of Claim 4(3).

(2) a path $xuvwyrstt_2$ such that u, v, w, r, s, and t are bad 3-vertices, y is a good vertex, and another neighbor z of y is a bad 3-vertex (see Figure 2),

(3) a path $xuvwyrs$ such that u, v, w, r, and s are bad 3-vertices, y is a good vertex, and another neighbor z of y is a bad 3-vertex which is also adjacent to a bad 3-vertex (see Figure 3).

Proof. (1) Suppose that there exists a path $xuvwyrstt_2$ such that u, v, w, r, s, and t are bad 3-vertices, y is a good vertex, and y is adjacent to a 2-vertex (see Figure 1). Denote $N(u) = \{x, u_1, v\}$, $N(v) = \{u, v_1, w\}$, $N(w) = \{v, w_1, y\}$, $N(r) = \{y, r_1, s\}$, $N(s) = \{r, s_1, t\}$, and $N(t) = \{s, t_1, t_2\}$, where u_1, v_1, w_1, r_1, s_1, t_1 are 1-vertices, and $N(x) = \{u, x_1, x_2\}$, $N(y) = \{w, r, z\}$, $N(z) = \{y, z_1\}$, and $N(t_2) = \{t, t_3, t_4\}$. Then $H \setminus \{uu_1, uv, vv_1, vw, ww_1, wy, yz, yr, rr_1, rs, ss_1, st, tt_1\}$ has a strong L-edge coloring C by the minimality of H. We are going to extend the coloring C to H. For each uncolored edge e in H, we use $L'(e)$ to denote the set of colors available for e. Then $L'(uu_1) = L(uu_1) \setminus \{C(xu), C(xx_1), C(xx_2)\}$ has at least three colors because $|L(e)| \geq 6$ for each edge e in H. Similarly we have the following:

(1) $|L'(e)| \geq 3$ if $e \in \{uu_1, uv, st, tt_1, yz\}$,

(2) $|L'(e)| \geq 5$ if $e \in \{vv_1, vw, wy, yr, rs, ss_1\}$,

(3) $|L'(e) \geq 6$ if $e \in \{ww_1, rr_1\}$.

By Lemma 3, we only need to

(i) either extend the coloring C to $yz, yr, rs, ss_1, st, tt_1$ so that there are still two colors available for rr_1,

(ii) or extend the coloring C to $yz, yr, rr_1, rs, ss_1, st, tt_1$ so that $|\{C(yz), C(yr), C(rr_1), C(rs)\} \cap [L(wy) \setminus \{C(zz_1)\}]| \leq 3$.

If there is a color $a \in L'(yz) \cap L'(st)$, we color the edges yz and st with the color a. Then we may further color the edges tt_1, ss_1, rs, and yr. This gives a coloring in (i). Thus $L'(yz) \cap L'(st) = \emptyset$. Similarly we can show $L'(yz) \cap L'(ss_1) = \emptyset$. Denote $L'(yz) = \{a, b, c\}$.

If $a \notin L'(yw)$, we first color the edge yz with the color a, and then by Lemma 3, we can further extend C to the edges yz, yr, rr_1, rs, ss_1, st, and tt_1. Since the edge yz is colored with a not in $L'(wy)$, such extension satisfies (ii). Therefore $L'(yz) \subseteq L'(wy)$. Similarly we can show that $L'(wy) = L'(yr) = L'(rs)$. Denote $L'(wy) = \{a, b, c, d, e\}$. We first color the edges yz, yr, and rs with a, b, and c, respectively. Then $L'(wy) \setminus \{a, b, c\}$ has two colors. By Lemma 3, we can first extend C to the edges uu_1, uv, vv_1, vw, ww_1, and wy. Since $\{a, b, c\} \cap [L'(st) \cup L'(ss_1)] = \emptyset$, we can further color the edges tt_1, st, rr_1, and ss_1 in order.

In each case, we can extend the coloring C to a strong L-edge coloring of H, a contradiction. Therefore, the configuration in Figure 1 does not exist.

Similarly we can also show that the configurations in Figures 2 and 3 do not exist either. □

Let $M(x) = d(x) - (15/7)$ be the initial charge of x for each vertex x. Then $\sum_{x \in V(H)} M(x) < 0$. We assign a new charge to each vertex according to the following rules.

R1. For each good 3-vertex x, if $xx_1x_2 \cdots x_tx_{t+1}$ is a path in which $x_1x_2 \cdots x_t$ is a maximal t-thread, then x sends $1/14$ to each x_i for $i = 1, \ldots, t$, and if $xx_1x_2 \cdots x_tx_{t+1}$ is a path in which x_i is a bad 3-vertex for each $i = 1, \ldots, t$, x sends $1/7$ to each x_i for $i = 1, \ldots, t$.

R2. Each bad 3-vertex sends $8/7$ to its 1-neighbor.

Now we consider the new charge $M'(x)$ for each vertex x.

(1) If $d(x) = 1$, then by Claim 1, x is adjacent to a bad 3-vertex. Thus $M'(x) = 1 - (15/7) + (8/7) = 0$.

(2) If $d(x) = 2$, then by R1, x receives $2 \times (1/14) = (1/7)$ in total from some 3-vertices. Thus $M'(x) = 2 - (15/7) + (1/7) = 0$.

(3) If x is a bad 3-vertex, then $M'(x) = (8/7) - (8/7) = 0$.

(4) Assume that x is a good 3-vertex. Denote by t and s the numbers of bad 3-vertices and 2-vertices receiving charges from x, respectively. By Claim 2 and Claim 4, we have $0 \leq t \leq 6$ and $0 \leq s \leq 6$. Hence $M'(x) = 3 - (15/7) - (t/7) - (s/14) = (6/7) - (t/7) - (s/14)$.

If $t = 0$, then $s \leq 6$. Hence $M'(x) \geq (6/7) - (6/14) > 0$.

If $0 < t \leq 3$, then $s \leq 4$. Hence $M'(x) \geq (6/7) - (3/7) - (4/14) > 0$.

If $t = 4$, then $s \leq 2$. Hence $M'(x) \geq (6/7) - (4/7) - (2/14) > 0$.

If $t = 5$, then $s \leq 2$. Hence $M'(x) \geq (6/7) - (5/7) - (2/14) \geq 0$.

If $t = 6$, then $s = 0$. Hence $M'(x) \geq (6/7) - (6/7) \geq 0$.

Therefore we have $0 \leq \sum_{x \in V(H)} M'(x) = \sum_{x \in V(H)} M(x) < 0$.

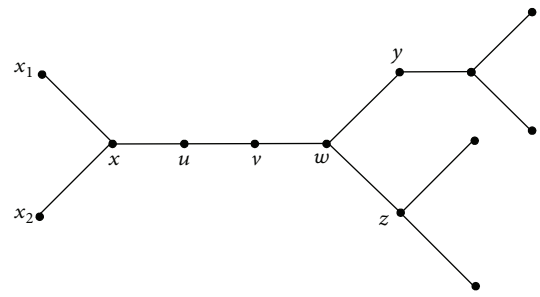


Figure 4: The configuration of Claim 7(1).

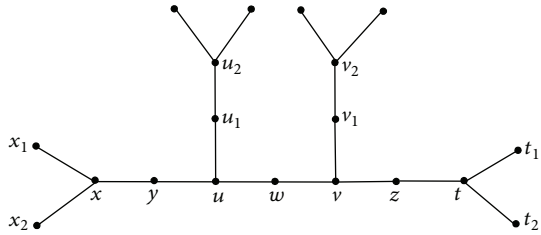


Figure 5: The configuration of Claim 7(2).

4. Proof of (ii) of Theorem 2

Let H be a counterexample with $|E(H)|$ as small as possible. Then there exists a 7-edge list L such that H is not strongly L-edge colorable.

Claim 5. There is no 1-vertex in H.

Proof. Suppose to the contrary that H contains a 1-vertex u, such that v is its neighbor. Since H is a minimum counterexample, $H' = H \setminus \{uv\}$ has a strong L-edge coloring. Hence $|L'(uv)| \geq 1$, we can easily extend this coloring to H, a contradiction. □

Claim 6. H does not contain a t-thread with $t \geq 3$.

Proof. The proof is similar to that of Claim 2 and thus omitted. □

Claim 7. H does not contain the following two configurations:

(1) A path $xuvw$ such that u and v are 2-vertices, x is a 3-vertex, and w is a 3-vertex which has two 2-neighbors and one 3-neighbor (see Figure 4).

(2) A path uwv where $d(u) = d(v) = 3$ and $d(w) = 2$ and both u and v have three 2-neighbors (see Figure 5).

Proof. (1) Suppose that H contains a path $xuvw$ in Figure 4. Let $H' = H \setminus \{uv, vw, wy\}$. Since H is a minimal counterexample, H' has a strong L-edge coloring. Then $|L'(wy)| \geq 1$, $|L'(vw)| \geq 2$, and $|L'(uv)| \geq 3$, and we can extend the coloring to H, a contradiction.

(2) Suppose that H contains the configuration in Figure 5, where y, u_1, w, v_1, and z are 2-vertices and u and v are 3-vertices. Since H is a minimum counterexample, $H' = H \setminus \{yu, uu_1, uw, wv, vv_1, vz\}$ has a strong L-edge coloring. Now $|L'(uu_1)| \geq 3$ and $|L'(vv_1)| \geq 3$. We first color the two edges uu_1 and vv_1 to obtain a strong L-edge coloring of $H \setminus \{yu, uw, wv, vz\}$. It is easy to check that $|L'(yu)| \geq 2$, $|L'(uw)| \geq 3$, $|L'(wv)| \geq 3$, and $|L'(vz)| \geq 2$. By Lemma 4, we can further extend the coloring to H, a contradiction. □

Let $M(x) = d(x) - (27/11)$ be the initial charge of x for each vertex x. Then $\sum_{x \in V(H)} M(x) < 0$. A 2_k-vertex is a 2-vertex with k 3-neighbors. We assign a new charge to each vertex according to the following rules.

R1. Let x be a 2_2-vertex and u a 3-neighbor of x. If u is adjacent to three 2-vertices, then u sends $2/11$ to x; otherwise u sends $3/11$ to x.

R2. Let x be a 2_1-vertex and u be its 3-neighbor. u sends $5/11$ to x.

(1) If x is a 2_1-vertex, then by R2, x receives $5/11$ from its 3-neighbor. Thus $M'(x) = 2 - (27/11) + (5/11) = 0$.

(2) If x is a 2_2-vertex with two neighbors u and v and if one of u, v has three 2-neighbors, then by Claim 7 the other one has at most two 2-neighbors. Hence x receives $(2/11) + (3/11) = 5/11$ from its neighbors. If neither u nor v has three 2-neighbors, then x receives $(3/11) + (3/11) = 6/11$ from its neighbors. Therefore $M'(x) \geq 2 - (27/11) + (5/11) \geq 0$.

(3) Assume $d(x) = 3$. By Claim 7, we only consider the following three cases: (a) if x is adjacent to three 2_2-vertices, then x sends out $3 \times (2/11) = 6/11$ to its neighbors; (b) if x is adjacent to at most two 2_2-vertices, then it sends out at most $2 \times (3/11) = 6/11$ to its neighbors; (c) if x is adjacent to one 2_1-vertex, then it sends out $5/11$ to its neighbors. In each case, we have $M'(x) \geq 3 - (27/11) - (6/11) = 0$.

Therefore we have $0 \leq \sum_{x \in V(H)} M'(x) = \sum_{x \in V(H)} M(x) < 0$.

5. Proof of (iii) of Theorem 2

Let H be a counterexample with $|E(H)|$ as small as possible. Then there exists an 8-edge list L such that H is not strongly L-edge colorable.

Claim 8. There is no 1-vertex in H.

Proof. The proof is similar to that of Claim 5 and thus omitted. □

Claim 9. There are no two adjacent 2-vertices in H.

Proof. Suppose that there are two adjacent 2-vertices u and v. Let w be the other neighbor of v. Since H is a minimum counterexample, $H \setminus \{uv, vw\}$ has a strong L-edge coloring. Then $|L'(uv)| \geq 3$ and $|L'(vw)| \geq 1$. Hence, we can extend the coloring to H easily, a contradiction. □

Claim 10. A 3-vertex is not adjacent to three 2-vertices in H.

Proof. Suppose to the contrary that a 3-vertex u is adjacent to three 2-vertices x, v, and w. Since H is a minimum counterexample, $H \setminus \{xu, uv, uw\}$ has a strong L-edge coloring. Thus $|L'(xu)| \geq 3$, $|L'(uv)| \geq 3$, and $|L'(uw)| \geq 3$. Therefore, we can extend the coloring to H, a contradiction. □

Claim 11. H does not contain a path $P = uvwxy$ where u, w, and y are 2-vertices and v and x are 3-vertices.

Proof. Suppose to the contrary that H contains a path $uvwxy$, where u, w, and y are 2-vertices and v and x are 3-vertices. By the minimality of H, $H \setminus \{vw, wx\}$ has a strong L-edge coloring. Uncolor uv and xy. It is easy to check that $|L'(uv)| \geq 2$, $|L'(xy)| \geq 2$, $|L'(vw)| \geq 3$, and $|L'(wx)| \geq 3$. By Lemma 4, we can extend the coloring to the path $uvwxy$ to obtain a strong L-edge coloring of H, a contradiction. □

Let $M(x) = d(x) - (13/5)$ be the initial charge of x for each vertex x. Then $\sum_{x \in V(H)} M(x) < 0$. We assign a new charge to each vertex according to the following rule.

R. Let x be a 3-vertex and t the number of 2-neighbors of x. x sends $2/5t$ to each adjacent 2-vertices if $t \neq 0$.

Obviously $M'(x) \geq 0$ if $d(x) = 3$.

Let x be a 2-vertex and u, v its neighbors. By Claims 9 and 10, both u and v are 3-vertices and have at most two 2-neighbors. If one 3-neighbor of x is adjacent to two 2-vertices, then by Claim 11, the other neighbor of x is adjacent to only one 2-vertex. Hence x receives $(1/5) + (2/5) = 3/5$ from its neighbors. If each neighbor of x is adjacent to only one 2-vertex, then x receives $(2/5) + (2/5) = 4/5$ from its neighbors. Hence $M'(x) \geq 2 - (13/5) + (3/5) = 0$.

Therefore we have $0 \leq \sum_{x \in V(H)} M'(x) = \sum_{x \in V(H)} M(x) < 0$. This contradiction completes the proof.

6. Proof of (iv) of Theorem 2

Let H be a counterexample with $|E(H)|$ as small as possible. Then there exists a 9-edge list L such that H is not strongly L-edge colorable.

Claim 12. There is no 1-vertex in H.

Proof. The proof is similar to that of Claim 5 and thus omitted. □

Claim 13. There are no two adjacent 2-vertices in H.

Proof. The proof is similar to that of Claim 9 and thus omitted. □

Claim 14. No 3-vertex is adjacent to two 2-vertices in H.

Proof. Suppose to the contrary that a 3-vertex u is adjacent to two 2-vertices v, w. Let x be the third neighbor of u. Since H is a minimum counterexample, $H \setminus \{uv, uw, ux\}$ has a strong L-edge coloring. Then $|L'(ux)| \geq 1$, $|L'(uv)| \geq 3$, and $|L'(uw)| \geq 3$. Hence we can extend the coloring to H, a contradiction. □

Claim 15. If a 3-vertex x has a 2-neighbor, then each 3-neighbor of x is not adjacent to a 2-vertex.

Proof. Suppose to the contrary that x has a 3-neighbor y such that y is also adjacent to a 2-vertex u. Let z be the 2-neighbor of x. We first assume that $u = z$. By the choice of H, $H \setminus \{xz, yz\}$ has a strong L-edge coloring with the edges xz and yz uncolored. Then it is easy to see that $|L'(xz)| \geq 3$ and $|L'(yz)| \geq 3$, and thus we may further color the edges xz and yz to get a strong L-edge coloring of H, a contradiction.

Now we assume that $u \neq z$. Let v be the other neighbor of z. Let C be a strong L-edge coloring of $H \setminus \{vz, zx, xy, yu\}$ obtained from a strong L-edge coloring of $H - \{z\}$ by uncoloring the edges xy and yu. It is easy to see that $|L'(vz)| \geq 2$, $|L'(zx)| \geq 3$, $|L'(xy)| \geq 2$, and $|L'(yu)| \geq 2$. By Lemma 4, C can be extended to those four uncolored edges, and thus H has a strong L-edge coloring, a contradiction. □

Let $M(x) = d(x) - (36/13)$ be the initial charge of x for each vertex x. Then $\sum_{x \in V(H)} M(x) < 0$. We assign a new charge to each vertex according to the following rules.

R1. Each 2-vertex receives $5/13$ from each adjacent vertex.

R2. If a 3-vertex x is adjacent to a 2-vertex then x receives $1/13$ from each 3-neighbor.

Obviously if $d(x) = 2$, then $M'(x) = 2 - (36/13) + (10/13) = 0$.

If $d(x) = 3$ and x is not adjacent to a 2-vertex, then $M'(x) \geq 3 - (36/13) - 3 \times (1/13) = 0$.

If $d(x) = 3$ and x is adjacent to a 2-vertex, then by Claims 14 and 15, $M'(x) = 3 - (36/13) - (5/13) + 2 \times (1/13) = 0$.

Therefore we have $0 \leq \sum_{x \in V(H)} M'(x) = \sum_{x \in V(H)} M(x) < 0$. This contradiction completes the proof.

7. Conclusion

This paper studies strong list edge coloring of subcubic graphs. The result can be used to deal with the conflict-free channel assignment problem in wireless radio networks when the admissible channels on the links between transceivers are constrained. We believe that the upper bounds on the maximum average degree in Theorem 2 are not sharp. It would be interesting to find sharp upper bounds for the maximum average degree.

Acknowledgments

The authors acknowledge the support of the National Natural Science Foundation of China (nos. 11101351 and 11171288) and NSF of University in Jiangsu province (no. 11KJB110014).

References

[1] J. L. Fouquet and J. L. Jolivet, "Strong edge-coloring of graphs and applications to multi-k- gons," *Ars Combinatoria A*, vol. 16, pp. 141–150, 1983.

[2] J. L. Fouquet and J. L. Jolivet, "Strong edge-coloring of cubic planar graphs," *Progress in Graph Theory*, pp. 247–264, 1984.

[3] L. D. Andersen, "The strong chromatic index of a cubic graph is at most 10," *Discrete Mathematics*, vol. 108, no. 1–3, pp. 231–252, 1992.

[4] K. W. Lih and D. D. F. Liu, "On the strong chromatic index of cubic Halin graphs," *Applied Mathematics Letters*, vol. 25, no. 5, pp. 898–901, 2012.

[5] W. C. Shiu and W. K. Tam, "The strong chromatic index of complete cubic Halin graphs," *Applied Mathematics Letters*, vol. 22, no. 5, pp. 754–758, 2009.

[6] H. Hocquard and P. Valicov, "Strong edge colouring of subcubic graphs," *Discrete Applied Mathematics*, vol. 159, no. 15, pp. 1650–1657, 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 817496, 11 pages
http://dx.doi.org/10.1155/2013/817496

Research Article

Approximated Slack Scaling for Structural Support Vector Machines in Scene Depth Analysis

Sheng Liu,[1,2] Binbin Zhai,[1] Sixian Chan,[1] Feng Li,[1] and Ye Zhan[3]

[1] *College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China*
[2] *Key Laboratory of Visual Media Intelligent Processing Technology of Zhejiang Province, China*
[3] *School of Accounting, Zhejiang University of Finance and Economics, Hangzhou, China*

Correspondence should be addressed to Sheng Liu; edliu@zjut.edu.cn

Received 31 January 2013; Accepted 31 March 2013

Academic Editor: Carlo Cattani

Based upon the framework of the structural support vector machines, this paper proposes two approaches to the depth restoration towards different scenes, that is, margin rescaling and the slack rescaling. The results show that both approaches achieve high convergence, while the slack approach yields better performance in prediction accuracy. However, due to its nondecomposability nature, the application of the slack approach is limited. This paper therefore introduces a novel approximation slack method to solve this problem, in which we propose a modified way of defining the loss functions to ensure the decomposability of the object function. During the training process, a bundle method is used to improve the computing efficiency. The results on Middlebury datasets show that proposed depth inference method solves the nondecomposability of slack scaling method and achieves relative acceptable accuracy. Our approximation approach can be an alternative for the slack scaling method to ensure efficient computation.

1. Introduction

Learning for stereo vision has been a challenging subject for a long time. Owing to the increment of ground truth datasets, considerable progress has been achieved, that is, using the scene structure of input images to learn a probability distribution model for matching [1–4] and adopting an expectation maximization algorithm to estimate disparity and then relearn the model parameters based on the estimation [5]. Although these methods have shown exciting results, the shortcoming is obvious, that is, the parameters must be preset or initialized manually on the basis of their prior knowledge. In [6], a new supervised machine learning method was proposed to handle such problem based on conditional random fields (CRFs), and the results had shown a promising future.

As mentioned above, supervised image labeling has been a long-lasting problem in computer vision. In recent years, CRFs have become a popular alternative to address this problem [7, 8], where the spatial correlations among neighboring pixels are incorporated by defining proper unary and pairwise potential functions on the related pixels. In addition, support vector machines have been widely used in image labeling [9], but they are less successful as noisy label results occurred for the absence of consideration of the spatial correlations.

Recently, structured prediction has caused widespread attention, and many new approaches have been proposed. Structured learning approaches solve the above-mentioned problems. In its computation process, both inputs and outputs are well structured, and strong internal correlations are revealed. It is formulated as the learning of complex functional dependencies between multivariate input and output representations. Structured learning has significant impact on addressing important computer vision tasks including image denoising [10], stereo [11], segmentation [12, 13], object localization [14, 15], and human pose estimation [16, 17]. A common way is to generalize the max-margin binary/multiclass classification to incorporate with structured information [14, 18–20]. It has been utilized in many respects, such as sequence labeling, image segmentation, grammar parsing, dependency parsing, bipartite matching, and text segmentation [21]. Furthermore, with the development of SVMs, structured information is introduced which

generated two new support vector machines named max-margin-based and slack-based SSVMs, respectively.

Max-margin method, with its decomposability of the error function, is possible to find the most violating constraint using the maximum a posteriori (MAP) inference algorithm for prediction [21]. But the shortcomings of the max-margin method are also obvious: it requires the error function being linearly comparable with the features, and it is sensitive to the most violating label. A label with large error would greatly decrease the separability of any other labels. An alternative choice is the slack scaling method. It has a fixed margin of 1 and reduces the violations in proportion to their errors which provide excellent accuracy. However, due to the nondecomposability of its error function, the slack method is not used widely. Therefore, we proposed an approximation method which modifies the slack method while reserving its normal properties. Depending on different given tasks, the proposed approximation method is effective to design most suitable loss functions and generate the corresponding solver.

This paper is organized as follows. In Section 2, we briefly discuss the principles of the SSVM. Our approach is proposed in Section 3 including steps to conduct the structural support vector machine, the typical max-margin method, and the expression of the improved slack method. Section 4 elaborates an approximation of the slack method. Section 5 provides the feature vectors which are utilized in our algorithm. As for Section 6, relative conditions and strategies for the training will be discussed and improved to make the training more efficient. Finally, we apply both methods for the depth restoration and make a detailed comparison between them.

2. Structural Support Vector Machine

Derived from statistical machine learning, the discriminative models focus on the posterior probability $p(y \mid x, \omega)$ and have been viewed as the most successful techniques for structural prediction. Here x is the input sample in the input space $\boldsymbol{\chi}$ and y is the associated output in the output space $\boldsymbol{\gamma}$. Given a feasible training set, for the training sample x_i and their associated truth output y_t, firstly a model for $p(y \mid x, \omega)$ will be learnt that the correct labels y_t have a higher probability than the wrong labels y, that is, $p(y_t \mid x, \omega) \geq p(y \mid x, \omega)$,and secondly, it can perform prediction by MAP estimation for a new sample x:

$$y^* = \arg\max_y p(y \mid x, \omega). \tag{1}$$

Under the framework of CRFs, $p(y \mid x)$ is modeled by a log linear model, which is often assumed as follows:

$$\log p(y \mid x, \omega) = \langle \omega, \Phi(x, y) \rangle - A_\omega(x), \tag{2}$$

where $\Phi(x, y)$ is a certain relationship between the input and its output; the second term, $A_\omega(x)$, is the normalization factor to make $p(y \mid x, \omega)$ a valid probability distribution.

By adopting the framework of max-margin method, the structural support vector machine tries to learn the weight vector, denoting the ω-parameterized model, to predict the correct output labels. And then, the optimization problem that results from the learning can be written as

$$\min_{\omega,\xi} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} \xi_i \tag{3}$$

subject to

$$\langle \omega, \Phi(x_i, y_t) \rangle - \langle \omega, \Phi(x_i, y) \rangle \geq \Delta(y, y_t) - \xi_i, \tag{4}$$

here, i from 1 to n denotes different samples, y is the label that is not equal to the true label y_t and $\Delta(y, y_t)$ denotes the loss between the two labels, ξ_i is the slack variables. Thus, the most violated constrains can be found by solving

$$y^* = \arg\max_y \left(\Delta(y, y_t) + F(x, y)\right), \tag{5}$$

where $F(x, y) = \langle \omega, \Phi(x, y) \rangle$ is the discriminative function. Therefore, y^* is reformulated as the minimization problem of energy, that is, $\operatorname{argmax}_y F(x, y) = \operatorname{argmin}_y E(x, y)$.

3. Our Approach

3.1. Problem Formulation. In stereo matching tasks, stereo images are two (or more) images of the same object taken from different views, named the left image (reference image) and right image, respectively. Assume that the right view image is just a horizontal shift of the left view, and the two images are the same size $R \times C$. Denoting $I(r, c)$ is the pixel on the cross of rth row and cth column in reference image, and $I'(r, c)$ the pixel on the same position in right image. The matching is aimed at finding the pixel-wise disparity which minimizes the energy

$$\begin{aligned} Y^* &= \arg\min_{\Upsilon} E\left(I, I', Y\right) \\ &= \arg\min_{\Upsilon} \left(\sum_{r,c} \left\| I(r, c) - I\left(r, c - y_{r,c}\right) \right\|^2 + E_{\text{smooth}}(Y) \right), \end{aligned} \tag{6}$$

where $y_{r,c}$ denotes the local disparity and $E_{\text{smooth}}(Y)$ is the smooth term which usually takes the form of *Pot's Model*

$$E_{\text{smooth}}(Y) = \begin{cases} 0 & y_i = y_j \\ p & \text{otherwise}, \end{cases} \tag{7}$$

where i and j are the index of neighboring pixels, y_i and y_j represent the neighboring disparity label, and p is a constant for penalty.

Normally the features of I and I' represent certain categories of visual information, for example, color, texture, or gradient. However, each category suits different situations. Texture features work well in boundary regions which usually are rich-textured but not applicative in weakly textured regions. Gradient-based features have opposite characters in comparison with texture features. In addition, different categories of features are not easy to be combined for learning. Simply expanding the dimension of feature vectors to involve

more features from different categories is dangerous due to sampling effect and scale. The highly weighted features will greatly influence the final results, also suppressing other features. Therefore, the data term should be constructed in the form of $\langle \omega_n, \emptyset(I, I', Y)\rangle$, where ω_n is the unary weight parameter which can balance the components in the combination feature vector against the sampling effect and different scales. These parameters W can be learnt from training examples.

By expanding the squared difference in data term, we will get three terms; that is, $I^2(r,c)$, $I'^2(r,c-y_{r,c})$, and $-2I(r,c)I'(r,c-y_{r,c})$. We use $\emptyset(I,I',Y) - \emptyset(I,I',Y_t)$ as the constraints in training phase, where Y_t is the ground truth, the term $I^2(r,c)$ would be canceled out by the subtraction because of its independency of label Y. We use $\|I(r,c) - I'(r,c-y_{r,c})\|^2$ to take the place of $I^2(r,c)$. Parameters working on these terms can balance the difference between $I(r,c)$ and $I'(r,c-y_{r,c})$, which is caused by sampling effect and camera settings. Overall, the data term is built as $\emptyset(I,I',Y) = [\|I(r,c) - I'(r,c-y_{r,c})\|^2, I'^2(r,c-y_{r,c}), -2I(r,c)I'(r,c-y_{r,c})]^T$.

3.2. Max-Margin Formulation for Stereo Learning. Assuming a learnt pairwise weight $\omega_e = p$, then the parameter W can be denoted as $W = (\omega_n, \omega_e)^T$, and the energy is written as $E(I,I',Y) = \langle W, \Phi(I,I',Y)\rangle$. Here $\Phi(I,I',Y)$ is the vector including data term $\emptyset(I,I',Y)$ and also the smooth term. The energy on ground truth Y_t should be minimized, that is, for all possible Y we have $E(I,I',Y) \geq E(I,I',Y_t)$. By adopting the margin scaling and adding the slack variables ξ_t to account for violations, the optimization problem reads, for $\eta > 0$,

$$\min_{W,\xi_t} \frac{\|W\|^2}{2} + \frac{\eta}{T}\sum_{t=1}^{T}\xi_t,$$
$$\text{s.t.} \quad \left\langle W, \Phi\left(I,I',Y\right)\right\rangle - \left\langle W, \Phi\left(I,I',Y_t\right)\right\rangle \geq \Delta\left(Y_t,Y\right) - \xi_t \qquad \forall t, Y, \xi_t \geq 0. \tag{8}$$

3.3. Slack Scaling Formulation. The margin rescaling method requires the label loss $\Delta(Y_t,Y)$ to be linearly comparable with the feature values $\Phi(I,I',Y)$. However, this is normally hard to be satisfied in structured learning, since $\Delta(Y_t,Y)$ counts the loss over each pixel in the image, and thus the aggregate value is much larger than feature values. Especially in stereo matching tasks, the pixel-wise loss may reach up to hundreds, which makes the overall loss even larger. Thus, we would like to adopt slack scaling, as it is invariant to the label loss scale. Nevertheless, the slack rescaling formulation is difficult to be solved, because no efficient approximation algorithm for Y^* exists. We follow the method introduced in [21] to solve this problem.

The slack rescaling optimization formulation is as follow:

$$\min_{W,\xi_t} \frac{\|W\|^2}{2} + \frac{\eta}{T}\sum_{t=1}^{T}\xi_t,$$
$$\text{s.t.} \quad \left\langle W, \Phi\left(I,I',Y\right)\right\rangle - \left\langle W, \Phi\left(I,I',Y_t\right)\right\rangle \geq 1 - \frac{\xi_t}{\Delta\left(Y_t,Y\right)} \qquad \forall t, Y, \xi_t \geq 0. \tag{9}$$

4. The Approximation for Slack Scaling

For the slack scaling optimization formulation, the inference engine problem is to find

$$y^S = \arg\min_y \left(s(y) + \frac{\xi}{L(y)}\right), \tag{10}$$

where $y \in \{s(y) - s(y_i) < 1 - \xi/L(y)\}$ is the set of the most violating label, ξ is the slack variable, and $s(y) = \langle W, \Phi(I,I',Y)\rangle$, $L(y) = \Delta(Y_t,Y)$.

As it is seen in the formulation, because $L(y)$ must be considered entirely, the second part of the formula cannot be decomposed easily. Thus, an approximation y^A is used to take the place of y^S and make it possible to be decomposed into the local parts.

It should be noted that $s(y) + \xi/L(y)$ is concave, and it has been proved approximated in the form of a linear function with respect to $L(y)$ [22]. The linearization and to be approximation procedure will be shown in the following parts.

4.1. Linearization and Approximation. According to [22], a concave function can be expressed in a linear form. Therefore, (10) is expressed as

$$s(y) + \frac{\xi_i}{L(y)} \geq \max_{\lambda\geq 0}\left(s(y) - \lambda L(y) + 2\sqrt{\xi\lambda}\right). \tag{11}$$

The aim of the inference problem is to find the optimal label y which minimizes the left side of (11). Therefore, we have

$$\begin{aligned}
&\min_y \left(s(y) + \frac{\xi}{L(y)}\right) \\
&\quad = \min_y \left(s(y) - \min_{\lambda\geq 0}\left(\lambda L(y) - 2\sqrt{\xi\lambda}\right)\right) \\
&\quad = \min_y \max_{\lambda\geq 0}\left(s(y) - \lambda L(y) + 2\sqrt{\xi\lambda}\right) \\
&\quad = \max_{\lambda\geq 0}\min_y \left(s(y) - \lambda L(y) + 2\sqrt{\xi\lambda}\right).
\end{aligned} \tag{12}$$

Here, let $F'(y_\lambda;\lambda) = s(y) - \lambda L(y) + 2\sqrt{\xi\lambda}$, thus

$$F(\lambda) = \min_y F'\left(y_\lambda;\lambda\right), \tag{13}$$

which leads to the simplified formulation as

$$\min_{y}\left(s(y)+\frac{\xi}{L(y)}\right)=\max_{\lambda\geq 0}\min_{y}F'(y_\lambda;\lambda)=\max_{\lambda\geq 0}F(\lambda). \tag{14}$$

For a fixed λ, firstly the most optimal label y_λ can be computed through minimization

$$y_\lambda=\arg\min_{y}\left(s(y)-\lambda L(y)\right). \tag{15}$$

Then, y_λ can be substituted into the formula $F(\lambda)$. We can find a λ that enables $F(\lambda)$ to catch its maximum, because $F(\lambda)$ is a function which is convex with respect to λ. $F(\lambda)$ can be seen as the max of a set of convex functions; therefore, $F(\lambda)$ is convex as well.

With the help of linear search algorithm such as Golden Search, the maximum of $F(\lambda)$ can be acquired in an efficient way. During the search procedure, it will encounter many different λs. By evaluating the $F(\lambda)$ for each λ, we can get different labels. The goal is to find the optimal label to get a minimum of $s(y)+\xi/L(y)$, which is denoted as y^A.

4.2. The Determination of Interval for λ. Since a simple constrain has been given out, $\lambda \geq 0$, it is obvious that $\lambda = 0$ can be the lower bound of λ as λ_l. However, if $\lambda = 0$, it will be hard to distinguish the $F'(y_\lambda;\lambda)$ between different labels in the early iterations, due to the neglect for the different loss $L(y)$. Let $\lambda_l = \varepsilon/L_{\max}$, where $L_{\max}$ is the possible maximal label loss and ε is the tolerance of the difference between two continuous iterations for this algorithm. In this way, a proper correct λ_l is obtained.

Then we come to determine the upper bound of λ. It is sufficient to find an upper bound λ as λ_u such that it returns $F(\lambda) = F(\lambda_u)$ for any $\lambda \geq \lambda_u$. And it also satisfies

$$y'=\arg\min_{y}s(y)-\lambda_u L(y), \tag{16}$$

which leads to the following formula

$$s(y')-\lambda_u L(y')\leq\min_{y,L(y)<L(y')}s(y)-\lambda_u L(y). \tag{17}$$

Here, let $y_1 = \operatorname{argmin}_y s(y)$ and L_ε be the minimal difference between $L(y)$ and $L(y')$, such as $L_\varepsilon = 1$ for Hamming loss. Then the right side of the function becomes $s(y_1)-\lambda_u(L(y')-L_\varepsilon)$. That requires $\lambda_u \geq s(y')-s(y_1)$ ($L_\varepsilon = 1$).

Since $s(y') < s(y_i)+1-\xi/L(y') < s(y_i)+1-\xi/L_{\max}$, so λ_u can be set as $\lambda_u = s(y_i)+1-\xi/L_{\max}-s(y_1)$.

5. Construction of Feature Vector

Image features are the terms used to describe images, as well as the clues for distinguishing the differences of images. Some image features may be the basic visual features, while others are defined for specific applications. Three types of features are used in this paper, that is, color, texture, and edge features.

5.1. Color Features. Color features are the basic visual description of images. Generally, color features are based on the characteristics of pixels, and each pixel in the image or the image region makes its own contribution to the color features. However, as a global feature, it is not sensitive to the changes of the size of the image or image region and also the directions in image. In other words, color features cannot capture the local characteristics of the image. And due to its nonuniqueness, pixels in different objects may share the same color features. Two basic color descriptions are RGB color space and YCbCr color space. While RGB concentrates on the gray levels of the pixels, the YCbCr pays close attention to the intensity, chromaticity, and the color difference. In YCbCr color space, the channel Y represents the intensity of the color, while channels Cb and Cr denote the chromaticity for blue and red, respectively. YCbCr color space can be easily obtained just by a linear transformation from RGB color space. Both the RGB and YCbCr color features are shown in Figure 1. In this paper, we use both RGB and YCbCr as the color features in the training process.

5.2. Texture Features. Similar to color features, texture features are also global features. The major difference is that texture features describe the statistical characteristics of the pixels in the image region. And the texture features have the properties of rotational invariance and noise immunity, but they are sensitive to the revolution of images, if the revolution changes, different features may be generated. On top of that, the light and the reflection on the surface of the objects may make it hard for computing the texture features.

In [23], Laws developed a method for computing texture features. According to this method, different convolution kernels, which were named Laws' masks, will be applied to our images. And the results will give some characteristics of the images. Here, the 2D Laws' masks can be generated from the following small kernels both with the length 3 and 5: $L_3 = [1\ 2\ 1]$, $E_3 = [1\ 0\ -1]$, $S_3 = [1\ -2\ 1]$, $L_5 = [1\ 4\ 6\ 4\ 1]$, $E_5 = [-1\ -2\ 0\ 2\ 1]$, $S_5 = [-1\ 0\ 2\ 0\ -1]$, $W_5 = [-1\ 2\ 0\ -2\ 1]$, $R_5 = [1\ -4\ 6\ -4\ 1]$.

Here, L denotes the average gray levels, E denotes the edge features, S stands for extracting the spots in the image, W stands for extracting the wave feature, and R stands for extracting the ripples in the image.

In order to generate the 2-D Laws' masks, we adopted matrix multiplication by a vertical 1D kernel and a horizontal 1-D kernel, such as $L_5E_5 = L_5^T \times E_5$. Take the masks scaled 3×3, for example, all the possible masks were listed in Table 1. After the convolve operation with these masks on an image sized M × N, the gray-scale texture feature image sized (N-masks_size + 1) × (M-masks_size + 1) will be generated. Figure 2 demonstrates the texture feature results generated by the 3×3 Laws' masks.

5.3. Edge Features. The object edge is the visual features of the discontinuity in the local image region which has a significant change in intensity. Generally, in images, the pixels along the edge have a smooth change in gray levels; however, on the

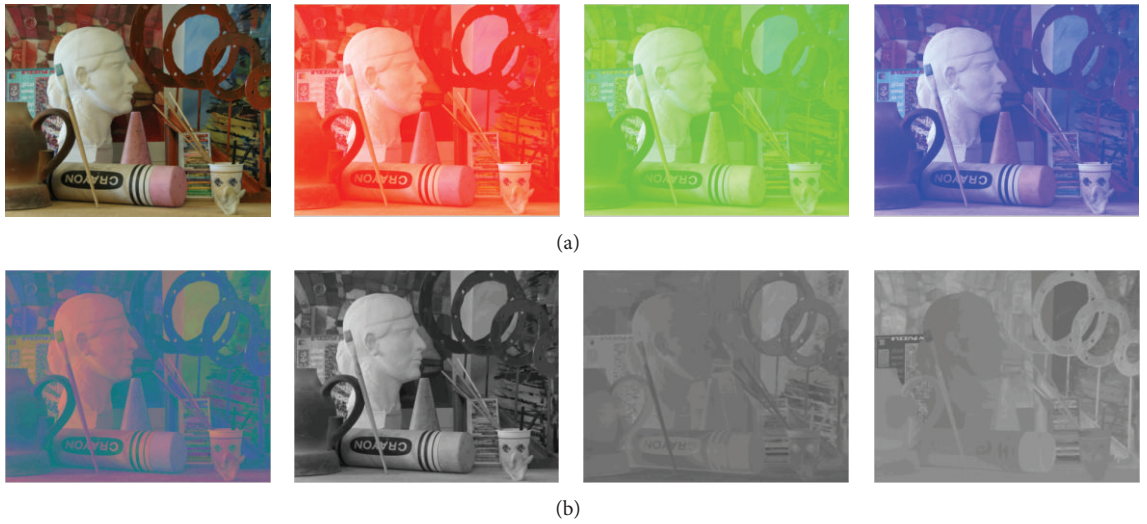


Figure 1: The color features of the image: RGB color features (first row) and YCbCr color features (second row). From left to right, first row: the original image in RGB color space, R channel, G channel, and B channel; second row: the original image in YCbCr color space, Y channel, Cb channel, and Cr channel.

Table 1: The possible Laws' masks scaled 3×3.

Masks	Method	Description
L_3L_3	$L_3^TL_3$	The gray level intensity within 3 neighboring pixels in both vertical and horizontal directions
L_3E_3	$L_3^TE_3$	In horizontal direction edge diction and in vertical direction gray level intensity
L_3S_3	$L_3^TS_3$	In horizontal direction spots detection and in vertical direction gray level intensity
E_3L_3	$E_3^TL_3$	In horizontal direction gray level intensity and in vertical direction edge diction
E_3E_3	$E_3^TE_3$	Edge detection in both vertical and horizontal directions
E_3S_3	$E_3^TS_3$	In horizontal direction spots detection and in vertical direction edge diction
S_3L_3	$S_3^TL_3$	In horizontal direction gray level intensity and in vertical direction spots detection
S_3E_3	$S_3^TE_3$	In horizontal direction edge diction and in vertical direction spots detection
S_3S_3	$S_3^TS_3$	Spots detection in both vertical and horizontal directions

direction which is vertical to the edge, the intensity of pixels change sharply.

The former denoted features are the local visual features. From the description, they are the surface features of the objects. On the other hand, the edge features are the measurement of the local compatibility. In this paper, 4 different Prewitt edge detectors which were directed in 0°, 45°, 90° and 135° were adopted in order to extract the edge features. The detectors in different directions and corresponding results are shown in Figure 3. By applying the 4 detectors, almost all the edges in the images can be captured.

6. Parameter Learning and Inference Problem

6.1. Bundle Method for Parameter Learning. For parameter learning, this paper utilizes the bundle method. Due to the formulation, such as

$$\min_{W,\xi_t} \frac{\|W\|^2}{2} + \frac{\eta}{T}\sum_{t=1}^{T}\xi_t,$$

$$\text{s.t.} \quad \left\langle W, \Phi\left(I, I', Y\right)\right\rangle - \left\langle W, \Phi\left(I, I', Y_t\right)\right\rangle \geq \Delta\left(Y_t, Y\right) - \xi_t \quad \forall t, Y, \xi_t \geq 0. \tag{18}$$

In order to obtain the optimal parameter, the constraints can be rearranged in the following form:

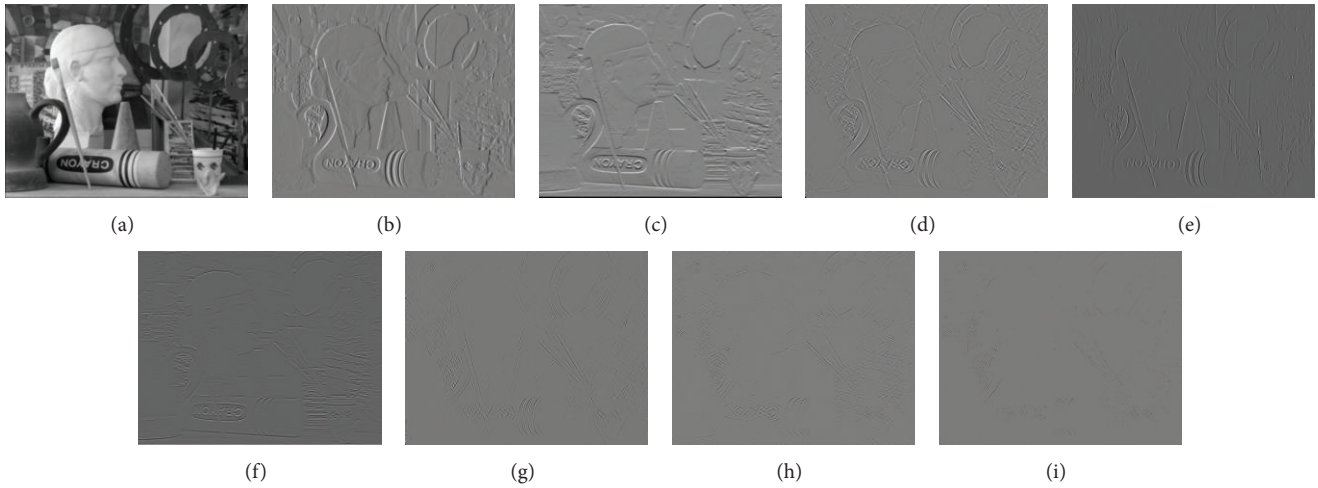
$$\left\langle W, \Phi\left(I, I', Y_t\right)\right\rangle \geq \left\langle W, \Phi\left(I, I', Y\right)\right\rangle - \Delta\left(Y_t, Y\right) + \xi_t. \tag{19}$$

This formula means that it is lower bounded by $\langle W, \Phi(I, I', Y_t)\rangle$. Then it generates the objective function to find the most violated constraints

$$Y^* = \arg\min_{Y}\left(\left\langle W, \Phi\left(I, I', Y\right)\right\rangle - \Delta\left(Y_t, Y\right)\right). \tag{20}$$

Thus, this forms an inference problem. And the bundle method can guarantee the optimal solution in a small number of iterations, so the problem can be solved efficiently. Algorithms 1 and 2 provide the parameter learning algorithm for both margin and slack method.

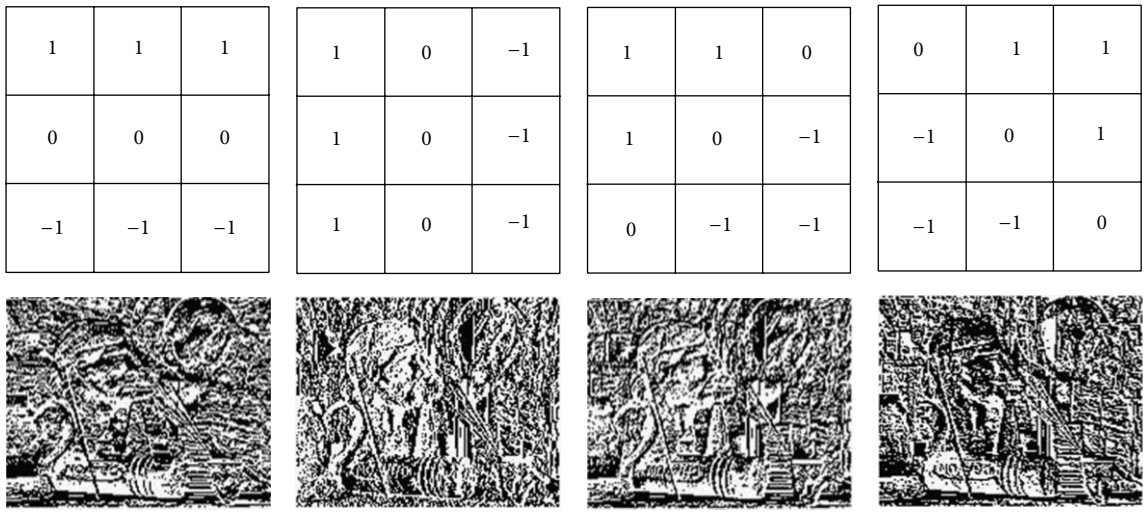
Both the margin and slack method refer to the optimal inference problems, so the best solution for them can be obtained via a standard graph-cuts algorithm (see reference [8] for detail). The frameworks seem to be the same, but in Algorithm 2, the inference engine is not similar to that in Algorithm 1. In this case, it needs to be approximated into a linear form, so that it searches for the best λ in the interval by the golden search algorithm.



(a) (b) (c) (d) (e)

(f) (g) (h) (i)

Figure 2: The outputs after the convolution of all the Laws' masks scaled 3×3.



(a) (b) (c) (d)

Figure 3: The results achieved by different edge detectors in 4 directions.

6.2. Golden Searching. In this paper, we adopted the golden searching algorithm during searching for the best approximation of the optimal label.

Firstly, suppose that there exists a continuous concave function **f** over the interval $[a, b]$, meanwhile it has only one minimum or maximum in the interval. Taking the minimum case for example, the binary searching algorithm is not the optimal algorithm for minimum searching, shown as follows:

Take the middle point as

$$m = \frac{a+b}{2}, \tag{21}$$

then two different points x_1 and x_2 are determined by

$$\begin{aligned} x_1 &= m - \frac{\delta}{2} \\ x_2 &= m + \frac{\delta}{2}, \end{aligned} \tag{22}$$

such that $f(x_1) \neq f(x_2)$. If $f(x_1) < f(x_2)$, the interval will be updated by $[a, x_1]$, otherwise $[x_2, b]$ will be the new interval. Obviously, each iteration step should call the binary searching for two times, which is not optimal.

In order to optimize the iteration process, there should be a factor which is capable of reducing the interval, named

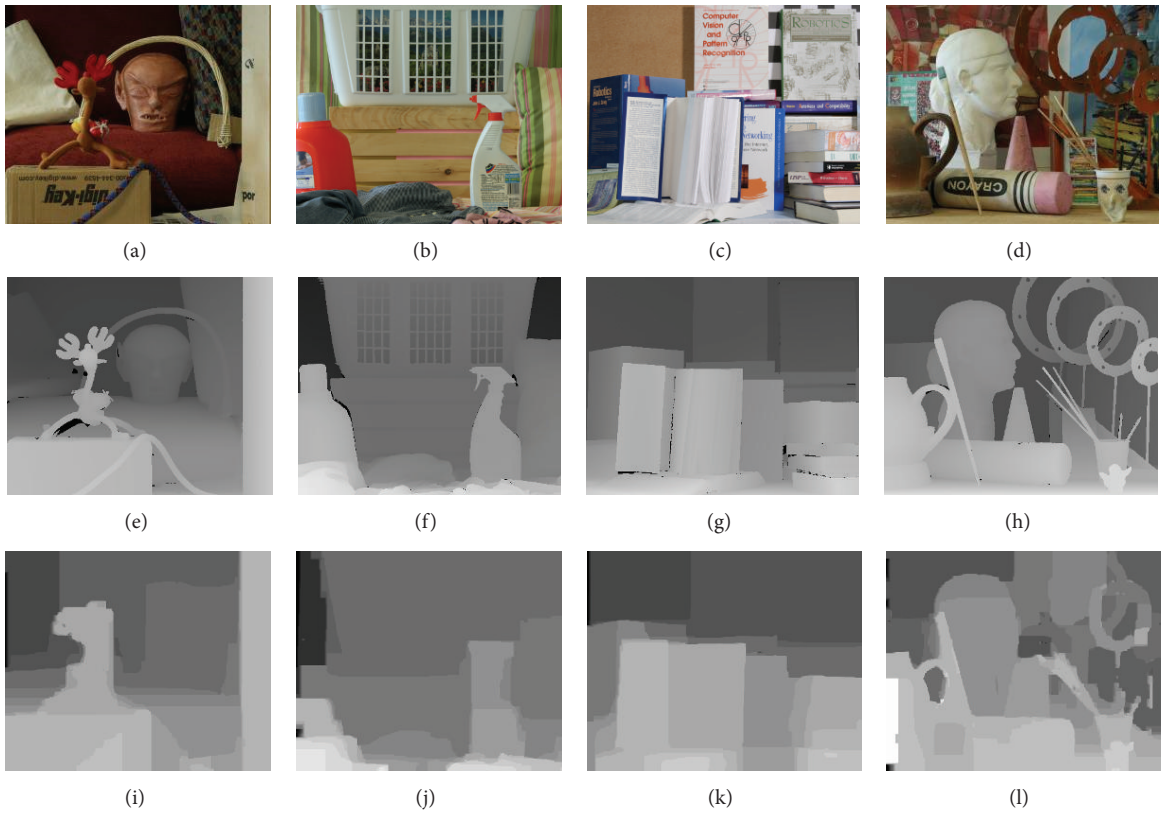


FIGURE 4: Inference depth maps by Max-margin method for different scene. From row 1 to 3: images, ground truth, and the obtained depth map. From column 1 to 4 are four scenes: 1st art, 2nd book, 3rd laundry, and 4th reindeer.

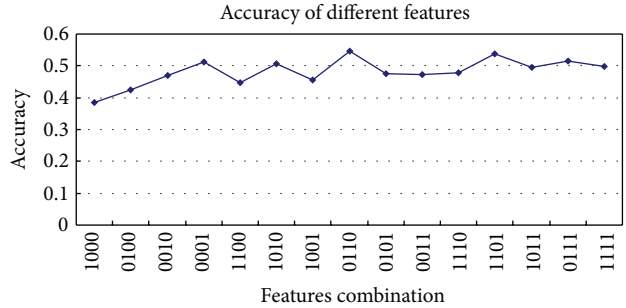


FIGURE 5: The inference accuracy of different features combination.

c. For x_1 and x_2 in the interval $[a, b]$, there are two different cases.

(1) If $f(x_1) < f(x_2)$, then the interval becomes $[a, x_2]$, and the interval size is compressed by c as follows:

$$\frac{x_2 - a}{b - a} = c, \tag{23}$$

as a result,

$$x_2 = (1 - c)\, a + cb. \tag{24}$$

(2) If $f(x_1) > f(x_2)$, similarly the interval is compressed by c and the new interval is $[x_1, b]$, then

$$\frac{b - x_1}{b - a} = c, \tag{25}$$

x_1 is obtained by

$$x_1 = (1 - c)\, b + ca. \tag{26}$$

Obviously, if the factor c is determined, it is easy to locate the points x_1 and x_2 in the interval. There are two rules for Cases (1) and (2), respectively, while Algorithm 3 shows the algorithm for golden searching.

Rule 1. If $f(x_1) < f(x_2)$, set $x_2 = x_1$, then compute another new x_1.

Rule 2. If $f(x_1) > f(x_2)$, set $x_1 = x_2$, then compute another new x_2.

7. Experiments and Results

We test the proposed methods on the Middlebury stereo datasets. The dataset contains many different scenes, that is, *art, books, dolls, laundry, moebius,* and *reindeer*, and each scene is consisted of 2 ground-truth images, related to view 1 and view 5 in each scene, and several different images which

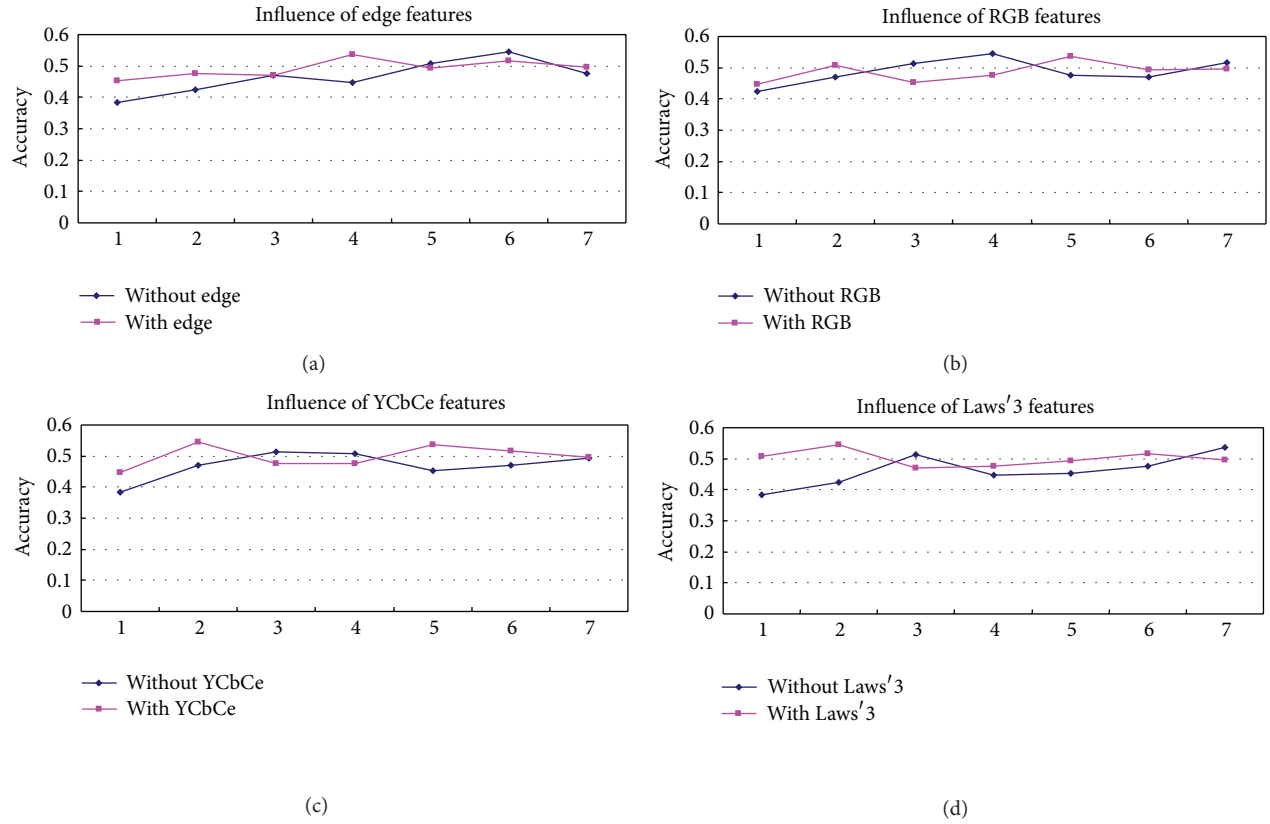


FIGURE 6: (a) the effect of the edge features. 1 to 7 means 1000, 0100, 0010, 1100, 1010, 0110, 1110. It shows that the edge features can boost the accuracy. (b) the effect of the RGB features. 1 to 7 means 0100, 0010, 0001, 0110, 0101, 0011, 0111. It shows that the RGB features can reduce the accuracy and the color feature with the texture features that can boost the inference accuracy. (c) the effect of the YCbCr features. 1 to 7 means 1000, 0010, 0001, 1010, 1001, 0011, 1011. It is easy to find that the YCbCr features have a similar effect on the accuracy with the RGB features. (d) the effect of the Laws' masks scaled 3×3. 1 to 7 means 1000, 0100, 0001, 1100, 1001, 0101, 1101. It is easy to find that the texture features can boost the accuracy in most of the situations.

were caught from different views. The ground-truth images are used as the label images of each scene, and its labels were compressed from 0–255 to 0–22 for the computing efficiency, and two neighbor view images are adopted to extract the different features.

Two groups of features are introduced in our experiments. The first group is local visual features, such as colors and textures, including the 3 dimensions of RGB color channels, the 3 dimensions YCbCr color channels, the 9 dimensions texture features, the outputs of Laws' masks scaled 3×3, and the 4 dimensions edge features, the outputs of the different Prewitt edge detectors. The second group is the graph edge features, which are the absolute difference between labels of neighboring pixels and one-dimensional bias constant. Practically, the method for conducting features may construct a large amount of dimensions, which can supply a rich set for choosing the suitable features to learn the parameters of the wanted model. By adopting the features and the Max-margin method, it may be easy for us to get the reasonable depth for different scenes, as shown in Figure 4.

7.1. Comparison on Inference Accuracy with Different Feature Combination. Suppose that the ground truth is denoted as y_t and the output results as y_o. Defining C_y as the number of the matched pixels in y_t and y_o and C_n as the number of different pixels in y_o from y_t, the inference accuracy can be denoted as

$$\text{acc} = \frac{C_y}{C_y + C_n}, \tag{27}$$

which stands for the ratio of the correct output.

In order to study the effects of different features, we have tested different combinations of image features. For the convenience of the expression, 1 denotes the state of the feature which was chosen, and 0 otherwise. Figure 5 shows the inference accuracy of different feature combinations for the 2nd scene *book*. Note, the order of the features arranged from left to right is RGB, YCbCr, laws' masks scaled 3×3, and the edge features. For example, 1000 denoted that only the RGB feature was chosen.

The combination of features does not always boost the accuracy of the results. In a word, some features have a negative effect on the results while others have a positive effect. In order to test it, a comparison between the set with a certain feature and another without it has been carried out. The results show that an offset effect does exist between

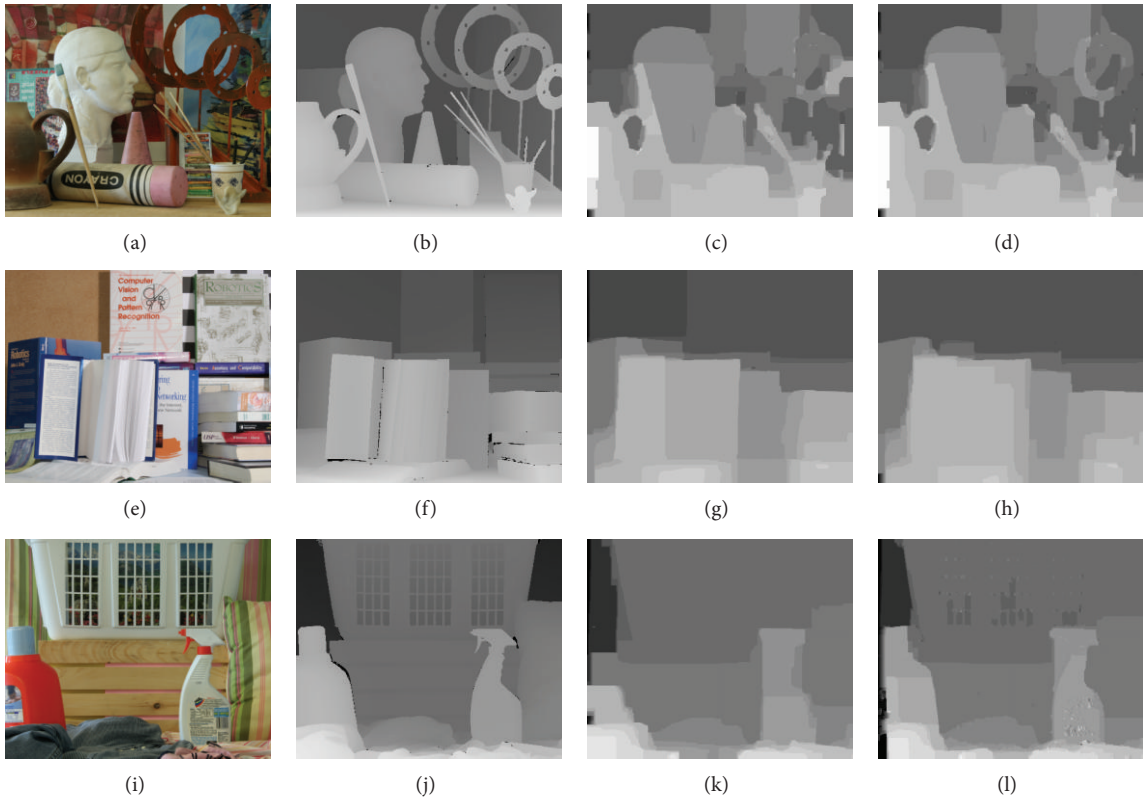


Figure 7: Depth inference results of different images by Max-margin and proposed slack method. From column 1 to 4: images, ground truth, the result of margin method, and the result of proposed slack method. And from row 1 to 3 are three scenes in Middlebury datasets: *art, book*, and *laundry*.

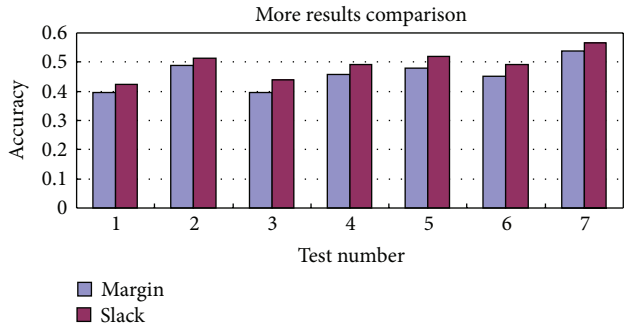


Figure 8: Different inference accuracy shows that the proposed slack method performs better than the margin method through the comparison of the inference accuracy.

Input: data X_t, label Y_t, size T, tolerance ε
Initialize parameter $W \rightarrow 0$, constraint set $R \rightarrow \emptyset$
Repeat
  for $t = 1$ to T
    $Y^* = \arg\min_Y [s(y) - \Delta(y_t, y)]$
  end for
  increase constraint set $R \leftarrow R \cup \{Y^*\}$
  $(W, \xi) \leftarrow$ solve the QP using all the existing Y^*
Until $\sum_t [\Delta(y_t, y) + s(y_t) - s(y)] \leq \xi + \varepsilon$

Algorithm 1: The parameter learning for margin method.

features, such as between color and edge feature, and also some features do boost the result, such as the textures in most situations (see Figures 6(a), 6(b), 6(c), and 6(d)).

7.2. Comparison between Margin and Slack Methods. To overcome the above-mentioned shortcomings of the Max-margin method, this paper adopts the slack scaling method to improve the results. In order to solve the nondecomposability problem, we introduce an approximated algorithm as described in Section 4 to make the slack method feasible. Both methods are tested on the Middlebury database, see Figure 7. As in Figure 8, the comparison results of inference accuracy for scene *art* show that the slack method performs better than the margin method.

7.3. Comparison on the Convergent Properties. To take a step further, the convergent property between margin and slack methods is compared. In the training procedure, the convergence of both margin and slack methods requires the use of the bundle method and *one-slack* trick. Take the

Input: data X_t, label Y_t, size T, tolerance ε
Initialize parameter $W \rightarrow 0$, constraint set $R \rightarrow \emptyset$
Repeat
  for $t = 1$ to T
    $Y^* = \arg\min_Y [s(y) + \xi/\Delta(y_t, y)]$
  end for
  increase constraint set $R \leftarrow R \cup \{Y^*\}$
  $(W, \xi) \leftarrow$ solve the QP using all the existing Y^*
Until $\sum_t [\xi/\Delta(y_t, y) + s(y_t) - s(y)] \leq \xi + \varepsilon$

Algorithm 2: The parameter learning for slack method.

Input: interval $[a, b]$, reduction factor c, tolerance ε
Initialize $x_1 \rightarrow x_1 = (1 - c)\, b + ca$,
  $x_2 \rightarrow x_2 = (1 - c)\, a + cb$,
Repeat
  If $(f(x_1) < f(x_2))$
    $b = x_2$, $x_2 = x_1$
    $f(x_2) = f(x_1)$
    $x_1 \rightarrow x_1 = (1 - c)\, b + ca$
  Else
    $a = x_1$, $x_1 = x_2$
    $f(x_1) = f(x_2)$
  $x_2 \rightarrow x_2 = (1 - c)\, a + cb$
  end If
Until $abs\,(x_1 - x_2) \leq \varepsilon$

Algorithm 3: The algorithm for golden searching.

margin method for example, the bundle method is used by rearranging the terms, then the constraints will be

$$\xi \geq \Delta(y, y_t) + s(y_t) - s(y). \tag{28}$$

This means that the constraints are up bounded by ξ. Given the current parameter, the objective function can be optimized using the bundle method, where the most violation constraint is

$$y^* = \arg\min_y \left(s(y) - \Delta(y_t, y)\right). \tag{29}$$

While the bundle method has the ability to achieve the optimal solution, the *one-slack* trick makes the procedure convergent in a small number of iterations. The computing process of the margin and the slack methods is examined to observe the convergence speed of the iteration. The error between two continuous iterations in the objective function is denoted as *itaeps*. Figure 9 shows the convergent property, indicating that both methods could converge in several iterations, while the slack method produces better accuracy without too much loss in convergence.

8. Conclusion

This paper presented two methods for the depth restoration of different scenes using structural vector machine. The proposed methods, including both margin and slack, have their own advantages and disadvantages, respectively. While the form of margin rescaling method can be decomposed into local parts easily, it is hard for the slack rescaling method to perform such operation. In contrast, the slack one outperforms the margin rescaling method in accuracy outstandingly. Besides the advantageous promotion in accuracy, there is no need for the slack rescaling method to abandon too many convergences while computing the parameters. The proposed approximation aiming at the slack rescaling approach manages to solve the decomposability problem successfully and make it computable in an efficient way. The pity is that the approximation method requires the formulation being concave which may be an over strong constraint. Our future works focus on these optimization algorithms, including improving the computing speed and enhancing the accuracy of the results.

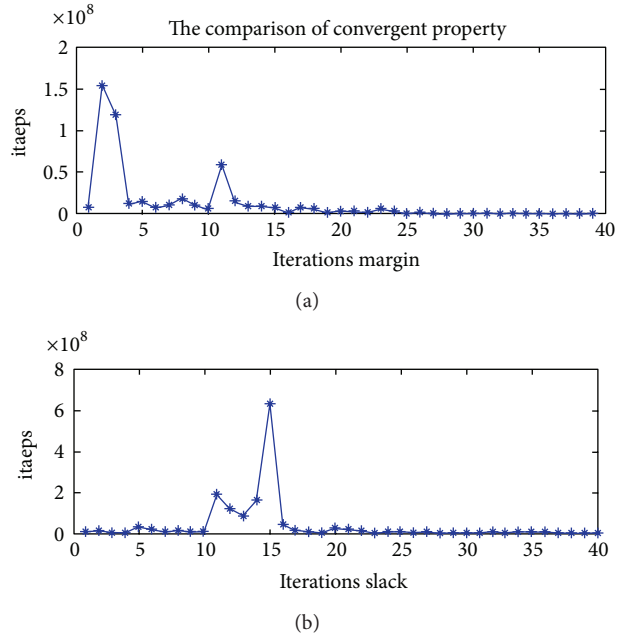


Figure 9: The comparison of the convergent property between margin and slack. Both the two methods converge in a small number of iterations. With the increasing accuracy, the slack method has a pronounced advantage in convergence compared to the margin method.

References

[1] D. Kong and H. Tao, "A method for learning matching errors in stereo computation," in *Proceedings of the British Machine Vision Conference (BMCV '04)*, 2004.

[2] S. Y. Chen and Z. J. Wang, "Acceleration strategies in generalized belief propagation," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 41–48, 2012.

[3] W. L. Wang, B. B. Xia, Q. Guan, and C. Shengyong, "Neural network based 3D model reconstruction with highly distorted stereoscopic sensors," in *Proceedings of the 8th International Workshop on Artificial Neural Networks (IWANN '05)*, vol. 3512 of *Lecture Notes on Computer Science*, pp. 661–668, June 2005.

[4] T. Fridman, J. Razumovskaya, N. Verberkmoes, G. Hurst, V. Protopopescu, and Y. Xu, "The probability distribution for a random match between an experimental-theoretical spectral pair in tandem mass spectrometry," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 455–476, 2005.

[5] L. Zhang and S. M. Seitz, "Parameter estimation for MRF stereo," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 288–295, June 2005.

[6] Y. Li and D. P. Huttenlocher, "Learning for stereo vision using the structured support vector machine," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, June 2008.

[7] D. Batra, R. Sukthankar, and T. Chen, "Learning class-specific affinities for image labelling," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.

[8] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[9] N. Cristianini and J. S. Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.

[10] J. J. McAuley, T. S. Caetano, A. J. Smola, and M. O. Franz, "Learning high-order MRF priors of color images," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 617–624, June 2006.

[11] P. Carr and R. Hartley, "Minimizing energy functions on 4-connected lattices using elimination," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 2042–2049, October 2009.

[12] M. Szummer, P. Kohli, and D. Hoiem, "Learning CRFs using graph cuts," in *Proceedings of the European Conference on Computer Vision (ECCV '08)*, pp. 582–595, 2008.

[13] D. Anguelov, B. Taskar, V. Chatalbashev et al., "Discriminative learning of Markov Random fields for segmentation of 3D scan data," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 169–176, June 2005.

[14] B. Taskar, V. Chatalbashev, and D. Koller, "Learning associative Markov networks," in *Proceedings of the 21th International Conference on Machine Learning (ICML '04)*, pp. 807–814, July 2004.

[15] C. Ionescu, L. Bo, and C. Sminchisescu, "Structural SVM for visual localization and continuous state estimation," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 1157–1164, September 2009.

[16] M. Blaschko and C. Lampert, "Learning to localize objects with structured output regression," in *Proceedings of the European Conference on Computer Vision (ECCV '08)*, pp. 2–15, 2008.

[17] M. Kim and V. Pavlovic, "Dimensionality reduction using covariance operator inverse regression," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.

[18] L. Bo and C. Sminchisescu, "Structured output-associative regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1–8, 2009.

[19] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun, "Large margin methods for structured and interdependent output variables," *Journal of Machine Learning Research*, vol. 6, pp. 1453–1484, 2005.

[20] B. Taskar, C. Guestrin, and D. Koller, "Max-margin Markov networks," *Advances in Neural Information Processing Systems*, vol. 16, 2003.

[21] S. Sarawagi and R. Gupta, "Accurate max-margin training for structured output spaces," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 888–895, July 2008.

[22] B. Taskar, C. Guestrin, and D. Koller, "Max-margin Markov networks," *Advances in neural information processing systems*, vol. 16, 2003.

[23] K. Laws, *Textured image segmentation [Ph.D. thesis]*, University of Southern California, 1980.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 986218, 10 pages
http://dx.doi.org/10.1155/2013/986218

Research Article

Hybrid Discrete Differential Evolution Algorithm for Lot Splitting with Capacity Constraints in Flexible Job Scheduling

Xinli Xu,[1] Li Li,[1] Lixia Fan,[2] Jing Zhang,[2] Xuhua Yang,[1] and Wanliang Wang[1]

[1] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*
[2] *College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Xinli Xu; 10165956@qq.com

Received 6 December 2012; Revised 31 January 2013; Accepted 20 February 2013

Academic Editor: Shengyong Chen

A two-level batch chromosome coding scheme is proposed to solve the lot splitting problem with equipment capacity constraints in flexible job shop scheduling, which includes a lot splitting chromosome and a lot scheduling chromosome. To balance global search and local exploration of the differential evolution algorithm, a hybrid discrete differential evolution algorithm (HDDE) is presented, in which the local strategy with dynamic random searching based on the critical path and a random mutation operator is developed. The performance of HDDE was experimented with 14 benchmark problems and the practical dye vat scheduling problem. The simulation results showed that the proposed algorithm has the strong global search capability and can effectively solve the practical lot splitting problems with equipment capacity constraints.

1. Introduction

In a practical production system, with much change of market and diversification of customer needs, variety and small batch production mode has gradually become the main way of manufacturing. In this mode, batch splitting scheduling problem has become an urgent issue addressed in the actual production. Batch splitting, also called lot streaming in many researches, is the process of splitting given jobs, each consisting of a batch of identical parts, into many smaller subbatches to allow their overlapping processing on alternative machines to get a better performance. An overview of the models and methods for batch splitting has been reported in the literature [1–3]. Job shop scheduling problem with batch splitting is commonly more complex than traditional job shop scheduling problem and usually more close to the actual manufacturing system.

In recent years, a growing body of literature suggests the research of job shop scheduling problems with batch splitting" for the sake of correctness. Low et al. [4] and Buscher and Shen [5] adopted integer programming methods to solve the batch splitting in job shop scheduling problem. Buscher and Shen (2009) presented consistent-sized batch splitting strategy under the conditions of given subbatches and put forward three-phase algorithm to solve lot streaming problem in job shops, which consists of predetermination of sublot sizes, the determination of schedules based on tabu search, and the variation of sublot sizes [6].

To minimize makespan of a batch splitting scheduling problem with setup time and alternative machines, Pan and Zhu (2004) split an original batch into equal sized subbatches, with the batch size for each subbatch fixed in advance, and used genetic algorithm to find the optimal sequence in which operations have to be machined [7]. Sun et al. [8] and An [9] put forward a batch splitting algorithm and a sublot scheduling algorithm based on genetic algorithm, optimized the number of subbatches of each job, and determined the size of each subbatch based on splitting strategy of equal-sized subbatches. Considering the production batch and the setup time of machine, Lin et al. (2007) proposed a hybrid genetic algorithm and combined the heuristic rules with simulated annealing algorithm to solve the flexible job shop scheduling problem by fixing the number of subbatches [10]. Ju and Zhu (2007) regarded each job as a batch, merged the same type of jobs sorted which are adjacent to each other into a subbatch in the evolution, combined particle swarm optimization and

genetic algorithm, and proposed the strategy of job shop scheduling optimization of batch production [11].

Bai et al. [12] proposed a flexible size lot splitting approach based on "cursors," designed a novel particle coding scheme combining the lot splitting with the sublot scheduling in particle swarm algorithm to solve the multiobjective flexible job shop scheduling with lot splitting. Huang [13] discussed multiobjective job shop scheduling with lot splitting production when the number of lots split is 1, 2, 3, and 4, in which subbatches are equal, and used ant colony algorithm to optimize the processing sequences. However, these recent researches focused on that the batch sizes of subbatch and sub-subbatches are determined in advance, and these research results have provided an important basis for intensive study solving the lot splitting problem in job shop scheduling.

Differential evolution algorithm (DE) [14] introduced by Storn and Price (1995) is a novel and efficient parallel direct search method, which has good robustness as well as fast convergence. Consequently, over the past few years, several researchers have demonstrated the applicability of DE in job shop scheduling problem [15–20]. In particular, Sang et al. (2010) gave an equal-sized batch scheme in advance and used DE to optimize the sequence of nonintermingling subbatches for the flow shop problems with the objective to minimize the total weighted earliness and tardiness [17]. Zhao et al. (2010) used the strategy of small production batch to shorten the production cycle. Considering mutation operator with the characteristics of the same sum of the individual values to meet the batch splitting constraints, DE was firstly applied in the lot splitting scheduling problem, and job shop scheduling problem with consistent-sized batch splitting [18], variable-sized batch splitting [19], and lot streaming under multiple-resource constraints [20] were, respectively, studied. However, the above researches did not involve the constraints of equipment capacity.

For solving the lot splitting with equipment capacity constraints in flexible job shop scheduling, it needs to split an original lot into many appropriate smaller processing subbatches, find the optimal sequence of those subbatches, and also consider equipment processing capacity constraints. It is a more complex NP hard problem which is motivated by practical engineering applications. In this paper, we proposed a hybrid discrete differential evolution algorithm for lot splitting with equipment capacity constraints in flexible job shop scheduling problem and used a two-level coding scheme to map the solution space into the chromosome space. To improve the performance of DE and avoid its trapping into local optimal solution, a local search strategy with dynamic random searching based on the critical path and random mutation operator is developed. The simulation results indicated that the algorithm has strong global search capability and can effectively solve the practical batch scheduling problem with the equipment capacity constraints.

The remainder of this paper is organized as follows. Section 2 provides the formulation of lot splitting with equipment capacity constraints in job shop scheduling problem, including subbatch size constraints, product and requirement constraints, setup costs, and setup time constraints and delivery duration constraints. Section 3 proposes a hybrid parallel algorithm combining with a local search strategy with dynamic random searching based on the critical path and random mutation operator to solve both the batch splitting problem and the batch scheduling problem. Section 4 gives the computational experiments and analyzes the results, and Section 5 makes the conclusion.

2. Problem Description and Formulation

2.1. Problem Description. Given a job shop production system with N kinds of products and R kinds of machines, where the capacity of each kind of machine is different, that is, each kind of machine has minimum and maximum output, the assumptions are that (1) any product can be divided into S batches processed individually, (2) each product has several operations that are waiting to be processed, (3) each machine can process different operation of any type of product, and (4) the setup time and costs for each operation cannot be omitted. Lot splitting with equipment capacity constraints in job shop scheduling problem is how to merge and split orders received, determine the number of sublots and sublot sizes of each product, the operational sequencing and working equipment of each batch, and minimize the objective function to satisfy with the constraints.

2.2. Notations

j: Machine index

o: Order index

p: Product type index

L: Number of orders

M: Number of machines

N: Number of production types

n_p: Number of operations of product p

J: A set of machines

O: A set of orders

P: A set of products

D_p: Demand of product p

$BD_{j,p}$: Sublot size of product p processed on machine j

BN_p: Number of sublots of product p

$\min V_j$: Minimum lot size allowed on machine j, that is, minimum output of machine j

$\max V_j$: Maximum lot size allowed on machine j, that is, maximum output of machine j

$\text{maxCost}_{j,p}$: Processing cost of product p on machine j with the maximum output

$\text{conCost}_{j,p}$: Constant cost of product p processed on machine j

$\text{unitCost}_{j,p}$: Unit cost of product p processed on machine j

$\text{maxTime}_{j,p}$: Processing time of product p processed on machine j with the maximum output

$\text{conTime}_{j,p}$: Constant time of product p processed on machine j

$\text{unitTime}_{j,p}$: Unit Time of product p processed on machine j

$\text{trCost}_{j,p,p'}$: Setup costs of product p and p' successively processed on machine j

$\text{trTime}_{j,p,p'}$: Setup time of product p and p' successively processed on machine j

$\beta_{p,o}$: Tardiness penalty coefficient of product p in order o

$d_{p,o}$: Due date of product p in order o

$TS_{j,p}$: Start time of product p processed on machine j

$TE_{j,p}$: Finishing time of product p processed on machine j

$Pe_{p,o}$: Completion time of product p in order o

$\text{poCost}_{p,o}$: Tardiness penalty cost of product p in order o.

2.3. Formulation. Lot splitting with equipment capacity constraints in job shop scheduling problem has the following constraints.

(1) Subbatch Size Constraint. In general, several products are taken in a batch as a unit to be processed during batch production. For the problems with equipment capacity constraints, each machine has its minimum output ($\min V_j$) and maximum output ($\max V_j$). The subbatch size ($BD_{j,p}$) of product p processed on machine j must satisfy the formula (1) and achieve the maximum output as much as possible

$$\min V_j \le BD_{j,p} \le \max V_j, \quad \forall j \in J,\ p \in P. \tag{1}$$

(2) Product Requirement Constraint. The number of the pth type of product processed on all machines should be equal to the total number of its requirement (D_p)

$$\sum_{j \in J} BD_{j,p} = D_p, \quad \forall p \in P. \tag{2}$$

(3) Setup Costs and Setup Time Constraint. In batch production, each machine is a flexible production line as it can process any type of product. Setup between two different products on the same machine usually generates setup costs. For example, the dye vat commonly must be cleaned in order to guarantee the quality of the next product, so it results in setup costs, that is, cleaning costs

$$TS_{j,p'} - TS_{j,p} \ge \text{trTime}_{j,p,p'}, \quad \forall j \in J,\ p \in P,\ p' \in P. \tag{3}$$

Formula (3) represents that the start time difference ($TS_{j,p'} - TS_{j,p}$) between the earlier product p and the later product p' processed on machine j should be bigger than the setup time ($\text{trTime}_{j,p,p'}$). Setup costs are affected by two successive products processed and equipment capacity; that is, there are the setup costs between the earlier task and the later task on the same equipment.

(4) Delivery Duration Constraint. Delivery dates of the orders from different customers always are different. In the same order, the customer sometimes even may have different requirements that some productions need to be processed urgently, so the tardiness penalty coefficient ($\beta_{p,o}$) of production p is different. Generally, $\beta_{p,o}$ of important customers is bigger than that of general customers. If the completion time ($Pe_{p,o}$) is beyond the delivery time ($d_{p,o}$), there will be the tardiness penalty costs ($\text{poCost}_{p,o}$) just like the following:

$$\text{poCost}_{p,o} = \beta_{p,o} \times \max\left(0, Pe_{p,o} - d_{p,o}\right), \quad \forall p \in P,\ o \in O. \tag{4}$$

Based on above constraints, the batches splitting of all products must be determined according to the received orders, moreover the number of sublots and sublot sizes of each procedure for each product must be arranged on different machine to minimize the total production costs of enterprises under the precondition of meeting the orders requirements as far as possible. Objective function of the lot splitting with equipment capacity constraints in flexible job shop scheduling problem can be seen as follows:

$$\begin{aligned} \text{minimize}\quad f = & \sum_{j \in J,\ p \in P} \left(\text{conCost}_{j,p} + \text{unitCost}_{j,p} \times BD_{j,p}\right) \\ & + \sum_{j \in J,\ p \in P,\ p' \in P} \text{trCost}_{j,p,p'} \\ & + \gamma \sum_{p \in P,\ o \in O} \beta_{p,o} \times \max\left(0, Pe_{p,o} - d_{p,o}\right). \end{aligned} \tag{5}$$

In the above formula, the total production costs include processing costs ($\sum_{j \in J, p \in P}(\text{conCost}_{j,p} + \text{unitCost}_{j,p} \times BD_{j,p})$), setup costs ($\sum_{j \in J, p \in P, p' \in P} \text{trCost}_{j,p,p'}$), and tardiness penalty costs ($\sum_{p \in P, o \in O} \beta_{p,o} \times \max(0, Pe_{p,o} - d_{p,o})$). γ is the ratio coefficient between tardiness penalty costs and the other two costs. Processing costs are made up of constant costs ($\sum_{j \in J, p \in P} \text{conCost}_{j,p}$) and variable costs ($\sum_{j \in J, p \in P} \text{unitCost}_{j,p} \times BD_{j,p}$). In the practical production, processing costs and processing time of product p processed on machine j with maximum output can be obtained by experience.

3. A Hybrid Discrete Differential Evolution Algorithm

3.1. Encoding and Decoding. In the lot splitting scheduling problem, the number of sublots and sublot sizes of each product must be determined, and the working machine and operational sequencing of each sublot must be also arranged. In this paper, a two-level coding scheme is developed to map the solution space into the chromosome space, where

the first level code represents lot splitting chromosome (chrome_1), and the second level code denotes lot scheduling chromosome (chrome_2). chrome_1 is used to determine the number of subbatches and subbatch sizes of each product, and chrome_2 is used to prioritize the operational sequencing arrangement of each subbatch.

For the problem with equipment constraints, a new coding scheme is proposed to describe chrome_1 considering the requirements of maximum output as far as possible. chrome_1 is divided into two parts, where the first part is the lot splitting scheme number of product p, the second part describes sublot sizes of product p, and the meaningless "0" is as the interval between them. The number of lot splitting scheme for product p (K_p) is less than or equal to the smaller one between its demand (D_p) and the number of machines (M). For example, there are two products divided into two subbatches. The batch size of each product is 8, and it can be divided into four schemes such as (1,7), (2,6), (3,5), and (4,4). So one of the lot splitting chromosomes may be [3, 1|0|3, 5, 1, 7]. As shown in Figure 1, BP_{P,BN_p} represents sublot size of the BN_pth batch of product p.

Based on the coding scheme of processing procedure, chrome_2 represents the operational sequencing of each procedure in each batch and the length of each chromosome $\text{len}_1 = \sum_{p=1}^{N} BN_p \times n_p$. For example, chrome_2 is [21, 11, 22, 12, 12, 21, 22, 11, 11, 22, 21, 12], where "21" means the first batch of the second product and three "21" indicate three different successive procedures of the second product.

Based on chrome_1 and chrome_2, the decoding scheme is as follows.

Step 1. Set $l = 1$ and the allowed start time of each machine $MST_j = 0$, where $j = 1, \ldots, M$.

Step 2. Select the lth gene of chrome_2 from left to right, acquire the processing product number p and the subbatch number k, and then obtain the subbatch size ($BP_{p,k}$) of the kth batch of production p combining with chrome_1 based on the gene location.

Step 3. Allocate working machines for the unscheduled tasks according to Substeps 3.1 and 3.2 and execute preparation and processing operation.

Substep 3.1. Fully considering the utilization of equipments, select the appropriate type of machine according to the subbatch size ($BP_{p,k}$). If $BP_{p,k}$ equals the maximum output of machine, then arrange the kth batch of production p to the type of equipment. Otherwise, arrange the task to process on the type of equipment whose maximum output is the nearest to $BP_{p,k}$.

Substep 3.2. After confirming the type of equipment, select a specific equipment according to the dispatching rules that first completed equipment is firstly arranged.

Firstly calculate the processing operation number s corresponding to the lth gene of chrome_2 and compare the finish time of the preceding operation with the earliest allowed start time of the processing equipment to obtain the start time of the kth batch of production p. According to the operation of the production p and the preceding task processed on the machine, determine whether they need setup. If this type of machine is the only one, then directly arrange the kth batch of production p to process on this machine. Otherwise, select the machine which can earliest complete the other tasks. Finally, calculate start time and finish time of the kth batch of production p.

Step 4. Refresh the earliest allowed start time of each machine.

Step 5. Set $l = l + 1$. If $l \leq \text{len}_1$, then go to Step 2. Otherwise, quit.

3.2. Global Evolution Procedure. DE is a parallel evolutionary algorithm based on population, which utilizes N_p parameter vectors as a population $X^t = [X_1^t, X_2^t, \ldots, X_{N_p}^t]$ for each generation t. Each vector called each individual of the population represents potential solution for the optimization problem. DE generates a mutant vector V_i^{t+1} for each vector X_i^t in the population by adding the weighted difference between two randomly selected vectors ($X_{r_2}^t$ and $X_{r_3}^t$) to a third one $X_{r_1}^t$. A trial vector U_i^{t+1} is then generated by using the crossover operator which mixes the components of the mutant vector V_i^{t+1} and the original one X_i^t. In the last step of each iteration, the selection operator with greedy strategy chooses the better one for the next generation by comparing U_i^{t+1} with X_i^t.

Since the DE algorithm was originally designed to work with continuous variables, rather most engineering problems have either an integer objective function or discrete objective function. Here, we adopt the following method as the global search strategy of a hybrid discrete differential evolution algorithm (HDDE) in this paper, which makes the individuals evolve directly in the discrete domain and can effectively speed up the algorithm.

(1) Consider

$$V_i^{t+1} = F \otimes G\left(F \otimes G\left(X_{r_2}^t, X_{r_3}^t\right), X_{r_1}^t\right), \tag{6}$$

where $i, r_1, r_2, r_3 \in [1, N_p]$, randomly selected; expect $r_1 \neq r_2 \neq r_3 \neq i$. $X_{r_1}^t$, $X_{r_2}^t$, and $X_{r_3}^t$ are three different randomly selected individuals, V_i^{t+1} is a mutant individual, and $F \in [0, 1]$. Formula (6) consists of two parts as follows.

(a) Consider

$$\begin{aligned} Y_i^{t+1} &= F \otimes G\left(X_{r_2}^t, X_{r_3}^t\right) \\ &= \begin{cases} G\left(X_{r_2}^t, X_{r_3}^t\right), & \text{if rand} < F \\ X_{r_2}^t, & \text{else.} \end{cases} \end{aligned} \tag{7}$$

In formula (7), rand represents a uniformly distributed random value that ranges from zero to one. Here generate a random number (rand), if rand $< F$, then set the subsequent individual $Y_i^{t+1} = G(X_{r_2}^t, X_{r_3}^t)$. Otherwise, $Y_i^{t+1} = X_{r_2}^t$. For the lot scheduling chromosome, $Y_i^{t+1} = G(X_{r_2}^t, X_{r_3}^t)$ means

$$\text{chrome}_1 \quad K_1, \ldots, K_p, \ldots, K_N | 0 | BP_{1,1}, \ldots, BP_{1,BN_1}, \ldots, BP_{p,1}, \ldots, BP_{p,BN_p}, \ldots, BP_{N,1}, \ldots, BP_{N,BN_N}$$

Figure 1: Lot splitting chromosome example of two products ($N = 2$).

to use POX crossover operator [21] that only crosses the processing sequence of the operations of parent chromosomes. For the lot splitting chromosome, $Y_i^{t+1} = G(X_{r_2}^t, X_{r_3}^t)$ indicates a simple two-point crossover that only operates the first part of the lot splitting chromosome.

(b) Consider

$$V_i^{t+1} = F \otimes G\left(Y_i^{t+1}, X_{r_1}^t\right) = \begin{cases} G\left(Y_i^{t+1}, X_{r_1}^t\right), & \text{if rand} < F \\ Y_i^t, & \text{else.} \end{cases} \tag{8}$$

Formula (8) describes that a mutant individual V_i^{t+1} is generated by the randomly selected individual $X_{r_1}^t$ and the above individual Y_i^{t+1}. Operation $G(Y_i^{t+1}, X_{r_1}^t)$ is the same as $G(X_{r_2}^t, X_{r_3}^t)$.

(2) Consider

$$U_i^{t+1} = Cr \otimes G\left(V_i^{t+1}, X_i^t\right) = \begin{cases} G\left(V_i^{t+1}, X_i^t\right), & \text{if rand} < Cr \\ V_i^{t+1}, & \text{else.} \end{cases} \tag{9}$$

Formula (9) means that a trial individual U_i^{t+1} is generated by the crossover of the original individual X_i^t and the mutant individual V_i^{t+1}, where Cr is the crossover probability ranged in [0, 1]. Operation $G(V_i^{t+1}, X_i^t)$ is the same as above.

3.3. Local Search Strategy. The above operators of global evolution procedure are directly carried out in discrete domain, which can effectively improve the search efficiency of the algorithm, but it is easy to fall into local extremum. In this paper, the local search strategy is developed to enhance the local exploration ability of the algorithm, which includes dynamic random search strategy based on the critical path used to adjust the lot scheduling chromosome and search neighborhood solutions of the operational sequencing in the case of the same subbatch divided and random mutation operator used to adjust the lot splitting chromosome and change the subbatch sizes again under the condition of the unchanged operational sequence of the sublots.

Dynamic random search [26] (DRS) is appropriate for continuous optimization problems and cannot be directly applied to the problem in this paper. So we improve DRS strategy to make it applicable for discrete operators. Furthermore, DRS involves the selection of the neighborhood, which has a larger influence on search quality and efficiency of the algorithm. Here the longest path of no time intervals between the operations is called the critical path of a feasible scheduling. The change of any operation on the critical path is the key to the change of the maximum finish time, and it is also an important part of the neighborhood structure constructed. Neighborhood based on the critical path can effectively avoid unnecessary search and improve the performance of algorithm. Therefore, the local search with two neighborhood structures including INTERCHANGE and INSERT based on the critical path is embedded in the framework of HDDE, which can adjust the operational sequencing under the condition of the lot splitting chromosome unchanged. The procedure of DRS is as follows.

Step 1. Set the initial values: *epoch* = 0 and Xcurrent = X, where Xcurrent and X are the current individual and the candidate individual, respectively.

Step 2. Reset the iteration counter, $n = 0$.

Step 3. Generate a new individual Xnew = INTERCHANGE (Xcurrent); that is, Xnew is generated by using exchange between the kth component and the qth component of Xcurrent and keeping the else components unchanged.

Step 4. Calculate the objective value of Xnew. Judge the dominance relation among Xnew, X, and Xcurrent. If Xnew dominates X, then update X = Xnew and go to Step 7. Otherwise, if Xnew dominates Xcurrent, then update Xcurrent = Xnew and go to Step 7.

Step 5. Generate a new individual, Xnew = INSERT(Xcurrent); that is, Xnew is generated by shifting one of the components from the kth component to the $q - 1$th component of Xcurrent in turn backward, and moving the qth component to the kth component.

Step 6. Calculate the objective value of Xnew. Judge the dominance relation among Xnew, X, and Xcurrent. If Xnew dominates X, then update X = Xnew and go to Step 7. Otherwise, if Xnew dominates Xcurrent, then update Xcurrent = Xnew and go to Step 7.

Step 7. Increase iteration number by one, $n = n+1$. If iteration counter is less than its maximum value ($n < Ns$), then go to Step 3.

Step 8. Update *epoch*, *epoch* = *epoch* + 1. If local search stop criterion is not reached (*epoch* < *Epoch*), then go to Step 2. Otherwise, quit.

At the same time, random mutation operator is introduced in the local search of lot splitting chromosome, which is used to search the neighborhood solution of the subbatches splitting under the unchanged operational sequence of the subbatches. The specific procedure is as follows: randomly select a chromosome and its mutant bit and then divide the subbatch sizes again.

Table 1: Results of HDDE and the other algorithms.

Problem	HDDE	KBACO [22]		PVNS [23]		hGA [24]		eGA [25]	
	C^*	$C_{\max}$	dev (%)	$C_{\max}$	dev (%)	$C_{\max}$	dev (%)	$C_{\max}$	dev (%)
C_1	14	14	0	14	0	14	0	- -	- -
C_2	11	11	0	- -	- -	- -	- -	- -	0
C_3	7	7	0	7	0	7	0	- -	- -
C_4	11	11	0	12	8.33	- -	- -	- -	- -
C_5	40	39	−2.56	40	0	40	0	40	0
C_6	26	29	10.34	26	0	26	0	26	0
C_7	204	204	0	204	0	204	0	204	0
C_8	60	65	7.69	60	0	60	0	60	0
C_9	173	173	0	173	0	173	0	173	0
C_{10}	63	67	5.97	60	−5.00	60	−5.00	58	−8.62
C_{11}	139	144	3.47	141	1.42	140	0.71	144	3.47
C_{12}	523	523	0	523	0	523	0	523	0
C_{13}	311	311	0	307	−1.30	307	−1.30	307	−1.30
C_{14}	221	229	3.49	208	−6.25	205	−7.80	198	−11.62

3.4. Algorithm Procedure. In this paper, the HDDE algorithm is presented based on the combination of the DE algorithm and the local search strategy. For simplicity, population evolution of DE is denoted as operation Oper_1, and the local search strategy is denoted as operation Oper_2. The procedure of HDDE solving the lot splitting with equipment capacity constraints in flexible job shop scheduling problem is as follows.

Step 1. Set parameters, such as Np, F, Cr, $G_{\max}$, *Epoch,* and Ns.

Step 2. Population initialization. Use random initialization to generate each individual including the lot splitting chromosome and the lot scheduling chromosome in the population. Moreover set generation $t = 0$.

Step 3. Operation Oper_1. Execute mutation operator and crossover operator on the lot splitting chromosome and the lot scheduling chromosome successively and then generate the mutant individual (V_i^{t+1}) and the trial individual (U_i^{t+1}), respectively.

Step 4. Calculate the fitness values, select to generate the next population (X^{t+1}), and update the optimal value and the optimal individual.

Step 5. Operation Oper_2. Carry out the local search of 10% individuals in the population X^{t+1} to find their neighborhood solution. Then calculate the fitness function and update the optimal value and the optimal individual.

Step 6. If the optimal value of the population has no continuous change for the appointed iterations, then reinitialize 10% individuals of the population and have the global optimal solution instead of one solution in the population.

Table 2: Information of orders received.

Order number	Product number	Color depth	Demand (kg)	Delivery (hour)	Weight of tardiness penalty
o_1	1	24	90	26	2
	2	60	40	30	1.5
	3	89	100	27	1
o_2	1	24	50	26	1.1
	3	89	90	35	1.6
o_3	1	24	130	48	1.3
	3	89	90	40	1.2
o_4	2	60	110	60	1.9
o_5	4	25	60	9	8
	5	62	120	16	8.5
o_6	6	88	500	48	1
o_7	7	98	60	18	1.4
	8	100	80	36	1.2
o_8	7	98	88	50	2.3
	8	100	100	48	2.1
	9	120	135	64	1.8
o_9	10	145	90	72	3.5

Step 7. Set $t = t + 1$. If $t < G_{\max}$, then go to Step 3. Otherwise, quit.

3.5. Analysis of Complexity of the Proposed Algorithm. As shown in the above procedure, the computation time of HDDE spent is almost in the iterative process. Each individual in the population goes through main operators including difference evolution, decoding, selection, and local search every iteration. Considering a batch splitting scheduling

Table 3: Cleaning time (unit: hour) while each product of different color depth is processed in the vat with the maximum output.

Dye vat number/quantity	Minimum capacity, maximum capacity (kg)	Color depth of each product									
		24	60	89	25	62	88	98	100	120	145
v_1/3	1, 10	0.1	0.15	0.25	0.15	0.5	0.25	0.3	0.3	0.35	0.45
v_2/1	5, 30	0.1	0.15	0.3	0.35	0.25	0.25	0.4	0.45	0.5	0.55
v_3/5	5, 50	0.2	0.3	0.45	0.2	0.3	0.45	0.5	0.5	0.6	0.65
v_4/1	10, 85	0.25	0.35	0.45	0.25	0.35	0.45	0.55	0.6	0.65	0.75
v_5/1	10, 90	0.3	0.45	0.55	0.3	0.45	0.55	0.65	0.65	0.8	0.8
v_6/7	20, 100	0.35	0.5	0.65	0.35	0.5	0.65	0.75	0.85	1.0	1.05
v_7/2	50, 500	0.45	0.65	0.8	0.45	0.65	0.8	0.95	0.95	1.15	1.2

Table 4: Processing time (unit: hour) and cost of different product on different vat with its maximum output.

Dye vat	Color depth of each product														
	p1	p2	p3	p1	p2	p3	p1	p2	p3	p1	p2	p3	p1	p2	p3
		24			60			89			25			62	
v_1	1/10	1.9/25	1.2/12	1.5/18	2.8/30	1.6/20	2/20	4/42	2/22	1/10	1.9/25	1.2/12	1.5/18	2.8/30	1.6/20
v_2	1/12	2/30	1.3/15	1.5/20	3.5/35	1.8/23	2.2/24	4.5/50	2/25	1/12	2/30	1.3/15	1.5/20	3.5/35	1.8/23
v_3	1.5/15	2.5/35	1.8/19	1.8/23	3.5/40	2/28	2.5/28	4.8/55	2.3/30	1.5/15	2.5/35	1.8/19	1.8/23	3.5/40	2/28
v_4	2/18	3/40	2.3/25	2.2/25	4/50	2.6/30	2.5/30	5.2/60	3/33	2/18	3/40	2.2/25	2.2/25	4/50	2.5/30
v_5	2/20	3/45	2.3/28	2.2/30	4/55	2.6/35	2.5/35	5.2/68	3/38	2/20	3/45	2.3/28	2.2/30	4/55	2.6/35
v_6	2.2/25	3.5/50	2.5/30	2.5/35	4.5/60	2.8/40	2.8/40	5.5/70	3.5/45	2.2/25	3.5/50	2.5/30	2.5/35	4.5/60	2.8/40
v_7	2.5/25	4.5/60	3.5/40	3.5/45	5.5/70	3.8/50	3.8/50	6.5/80	4.5/55	2.7/25	4.7/60	3.7/40	3.6/45	5.7/70	3.9/50
		88			98			100			120			145	
v_1	2.2/20	4/42	2/22	2.4/21	4.5/48	2.4/24	2.5/21	4.6/50	2.4/24	3/25	5/68	2.8/26	3.5/26	5.5/75	3.2/27
v_2	2.2/24	4.5/50	2/25	2.5/25	4.7/55	2.5/26	2.6/25	4.8/58	2.5/27	3.2/28	5.5/75	3/30	3.8/30	6.2/82	3.5/32
v_3	2.5/28	4.8/55	2.3/30	2.8/30	5/62	2.6/32	2.9/30	5.2/65	2.7/33	3.5/32	6/78	3.2/34	4/35	6.5/88	3.7/36
v_4	2.5/30	5/60	2.8/33	2.8/32	5.2/70	3.1/35	3/32	5.4/75	3.2/35	3.6/35	6.2/85	3.8/38	4.2/38	6.8/100	4.4/40
v_5	2.5/35	5.2/68	3/38	2.8/38	5.4/80	3.2/40	3/38	5.5/85	3.3/40	3.7/40	6.3/95	3.9/45	4.4/45	7/120	4.5/58
v_6	2.8/40	5.5/70	3.5/45	3/45	5.6/100	3.2/50	3.3/48	6.2/10	3.4/53	3.8/55	6.8/120	4/60	4.5/62	7.2/135	4.6/65
v_7	3.8/50	6.5/80	4.5/55	4/50	6.2/110	4.2/55	4.2/50	6.5/115	4.4/55	4.5/60	7/145	4.8/65	4.8/65	7.5/165	5/70

problem with population size (N_p), the maximal iteration number ($G_{\max}$), dynamic random search times (*Epoch*), and the length of the lot scheduling chromosome (len_1), the time complexity of difference evolution operator is $O(N_p \times \text{len}_1)$, decoding operator is $O(N_p \times \text{len}_1)$, selection operator is $O(N_p)$, and local search operator is $O(\text{len}_1 \times Epoch \times 0.1)$. Therefore, the time complexity of whole algorithm is

$$\begin{aligned} &O\left(N_p, G_{\max}\right) \\ &\quad = G_{\max} \times \left(O\left(N_p \times \text{len}_1\right) + O\left(N_p \times \text{len}_1\right)\right. \\ &\qquad \left. + O\left(N_p\right) + O\left(\text{len}_1 \times Epoch \times 0.1\right)\right) \quad (10) \\ &\quad = G_{\max} \times \left(2O\left(N_p \times \text{len}_1\right)\right. \\ &\qquad \left. + O\left(\text{len}_1 \times Epoch \times 0.1\right) + O\left(N_p\right)\right). \end{aligned}$$

Obviously, the time complexity of HDDE algorithm is related to len_1, N_p, $G_{\max}$, and *Epoch*.

4. Simulation Results and Analysis

4.1. Test Problems and Comparison of Results. We evaluate performance of the HDDE algorithm on 14 test problems with different complexity, in which $C_1 \sim C_4$ designed by Kacem et al. [27, 28] are 8 × 8, 10 × 7, 10 × 10, and 15 × 10 problem, respectively, and $C_5 \sim C_{14}$ designed by Brandimarte [29] are from MK1 to MK10 problem. The proposed algorithm has been coded with VC++ 6.0 and runs on a PC with Intel Pentium CPU 2.66 GHz processor and 1 G of memory. The parameters of HDDE are set as follows: $N_p = 50$, $G_{\max} = 500$, $F = 0.5$, $Cr = 0.7$, $Epoch = 50$, and $Ns = 10$.

The results of HDDE, KBACO [22], PVNS [23], *hGA* [24], and eGA [25] are compared in Table 1, where C^* is the optimal value obtained by HDDE, $C_{\max}$ is the optimal value obtained by the algorithms from literatures, dev denotes the relative error percent between HDDE and the other algorithms, that is, $\text{dev} = (C_{\max} - C^*)/C_{\max} \times 100\%$, and "- -" means that corresponding experiment was not conducted in those literatures. As shown in Table 1, the HDDE algorithm is not worse than any of those algorithms, and it can obtain the global optimal solution for most of the problems, which

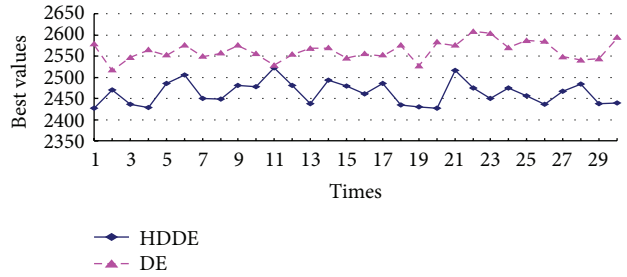


Figure 2: Best values of HDDE and DE obtained over 30 runs.

Table 5: Results of DE and HDDE.

Algorithm	Best	Worst	Average	Standard deviation	Runtime (s)
DE	2518.21	2607.86	2563.29	21.65	1.3
HDDE	2426.6	2521.89	2463.16	26.76	3.2

indicates that the proposed algorithm has better global searching ability. Each algorithm has its own advantages, but it also has certain disadvantage on large scale problems. For instance, the HDDE algorithm has lower global search for C_{13} and C_{14}.

4.2. Practical Dye Vat Scheduling Problem and Discussion of Results. In this paper, we get a certain amount of data samples from actual production in some dyeing workshop. Suppose that the enterprise receives 9 orders, where 10 types of products need to be processed. The products, color depth, demand, delivery duration, and weight of tardiness penalty are shown in Table 2. There are 7 types of dye vats in the dyeing workshop, all of which have different capacity. Information of each vat, such as the number, quantity, minimum capacity, and maximum capacity of each dye vat, is shown in Table 3. There is also the cleaning time (unit: hour), that is, setup time, while the product of different color depth is processed in the vat with the maximum output in Table 3. The processing time (unit: hour) and cost of different product on different vat with its maximum output are shown in Table 4, where p1, p2, and p3 denote three main operations of dyeing process, that is, preprocessing, dyeing, and postprocessing, respectively. The setup time and cost of preprocessing operation of any product are 0. Moreover, suppose that switching cost factors $\alpha = 0.9$ and $\beta = 0.7$.

To evaluate the effectiveness of local search strategies, DE and HDDE run 30 times, independently, and the parameters are set as follows: $N_p = 50$, $G_{\max} = 200$, $F = 0.5$, $Cr = 0.7$, $Epoch = 10$, and $Ns = 3$. Figure 2 and Table 5 show the comparison of results obtained by DE and HDDE over 30 runs and Figure 3 is convergent graph of the optimal value obtained by DE and HDDE. As seen in Figure 2 and Table 5, the local search strategy leads to a significant improvement in the global searching ability of HDDE. The results obtained by HDDE over 30 times are better than those of DE. However, HDDE also consumes a certain amount of computation time. In addition, As we can see from Figure 3, DE does not evolve essentially in later stage with the increasing of iteration.

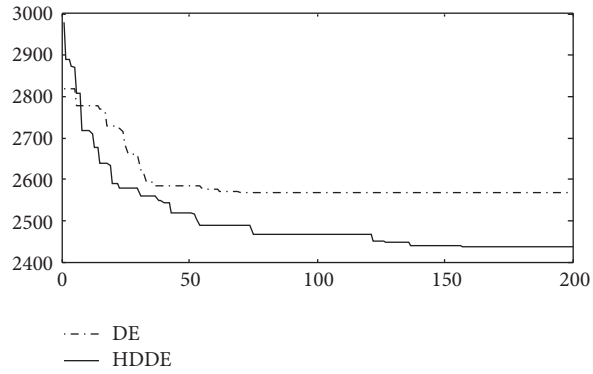


Figure 3: Convergent graph of the optimal solution for DE and HDDE.

On the contrary, HDDE can avoid trapping into the local optimum effectively, and it has higher optimization efficiency than DE.

In general, a larger standard deviation e represents that there is a greater difference between the most value and the average value of solutions, and a smaller standard deviation shows that the most value is close to the average value of solution. As we can see from Figure 3, no more changes are occurred in the optimal value obtained by DE in later iterations; that is, DE easily gets into the local extremum in the later evolution period. Furthermore, as shown in Figure 2 and Table 5, most value of solutions calculated for 30 times is close to the average value. Because of a local search strategy with dynamic random searching based on the critical path and random mutation operator, HDDE can evolve further in later iterations as seen in Figure 3; that is, the local search strategy helps HDDE being out of local minimum. On the other hand, randomness of the local search strategy also results in the bigger difference between some value of solution and the average value as shown in Figure 2 and Table 5.

Finally, the batch number of every product is [3, 2, 3, 1, 2, 5, 2, 2, 2, 1], the optimal lot splitting chromosome is [85, 100, 85, 30, 120, 10, 10, 260, 60, 100, 20, 10, 10, 10, 50, 420, 10, 138, 10, 170, 10, 125, 90] and the optimal lot scheduling chromosome is [52, 81, 32, 64, 101, 101, 101, 31, 63, 32, 82, 33, 51, 92, 71, 91, 22, 12, 22, 41, 32, 63, 81, 21, 41, 81, 22, 72, 72, 31, 21, 62, 12, 72, 63, 31, 13, 64, 21, 65, 51, 62, 52, 12, 62, 41, 71, 33, 71, 13, 61, 82, 52, 11, 64, 91, 33, 82, 11, 61, 65, 65, 13, 61, 92, 11, 51, 92, 91]. As seen from the lot splitting chromosome, most of the subbatches satisfy the requirement of the maximum output of certain dye vat, and each product only has at most one batch unsatisfied. Simulation results show the feasibility of batch splitting and validity of the proposed algorithm.

5. Conclusion

Aiming at lot splitting with equipment capacity constraints in flexible job shop scheduling problem, a two-level coding scheme including the lot splitting chromosome and the lot scheduling chromosome is presented. The new encoding method of the lot splitting chromosome is applied to solve

the batch splitting optimization problem satisfying the maximum processing output of machine. Moreover, according to the faultiness that DE is apt to be trapped into the local optima, the local search strategy with dynamic random searching based on the critical path and random mutation operator is developed, which is, respectively, used to adjust the neighborhood solutions of the lot splitting chromosome and the lot scheduling chromosome. Based on the analysis and comparison of results of benchmark problems, it is shown that the proposed algorithm has better optimization performance beyond most of other global algorithms. Finally, the proposed algorithm is applied to solve the practical scheduling problem of some dyeing workshop. Simulation results show that the proposed algorithm has good global search ability and it can effectively solve the practical lot splitting with equipment capacity constraints in flexible job shop scheduling problem.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC 61070043, 61105073, and 61203371), the Subproject of National Science and Technology Support Plan (2012BAD10B0101), and the Open Fund Project of Zhejiang University of Technology (20120814).

References

[1] D. Trietsch and K. R. Baker, "Basic techniques for lot streaming," *Operations Research*, vol. 41, no. 6, pp. 1065–1076, 1993.

[2] W. Huang and S. Chen, "Epidemic metapopulation model with traffic routing in scale-free networks," *Journal of Statistical Mechanics*, vol. 2011, Article ID P12004, 19 pages, 2011.

[3] W. Han, J. Chen, and X. Bu, "Batch splitting activity in scheduling of virtual cell with capacity constraints based on Bi-level mathematical model," *Applied Mechanics and Materials*, vol. 263–266, pp. 1257–1264, 2012.

[4] C. Low, C. M. Hsu, and K. I. Huang, "Benefits of lot splitting in job-shop scheduling," *International Journal of Advanced Manufacturing Technology*, vol. 24, no. 9-10, pp. 773–780, 2004.

[5] U. Buscher and L. Shen, "An integer programming formulation for the lot streaming problem in a job shop environment with setups," in *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS '11)*, pp. 1343–1348, Hong Kong, March 2011.

[6] U. Buscher and L. Shen, "An integrated tabu search algorithm for the lot streaming problem in job shops," *European Journal of Operational Research*, vol. 199, no. 2, pp. 385–399, 2009.

[7] Q. K. Pan and J. Y. Zhu, "Optimization method for a job-shop scheduling problem with alternative machines in the batch process," *Chinese Journal of Mechanical Engineering*, vol. 40, no. 4, pp. 36–39, 2004.

[8] Z. J. Sun, J. An, and W. Q. Huang, "Lot scheduling with multiple process routes in job shop," *China Mechanical Engineering*, vol. 19, no. 2, pp. 183–187, 2008.

[9] J. An, *Optimization of Job-Shop Scheduling Problem in the Batch Process*, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 2005.

[10] N. Lin, B. Meng, and Y. Fan, "Hybrid genetic algorithm for multiple process and batch scheduling in job-shop," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 33, no. 12, pp. 1471–1476, 2007.

[11] Q. Y. Ju and J. Y. Zhu, "Multi-objective flexible job shop scheduling of batch production," *Chinese Journal of Mechanical Engineering*, vol. 43, no. 8, pp. 148–154, 2007.

[12] J. J. Bai, Y. G. Gong, N. S. Wang, and D. B. Tang, "Multi-objective flexible Job Shop scheduling with lot-splitting," *Computer Integrated Manufacturing Systems*, vol. 16, no. 2, pp. 396–403, 2010.

[13] R. H. Huang, "Multi-objective job-shop scheduling with lot-splitting production," *International Journal of Production Economics*, vol. 124, no. 1, pp. 206–213, 2010.

[14] R. Storn and K. Price, *Differential Evolution—A Simple And Efficient Adaptive Scheme For global optimization Over Continuous spaces*, University of California, Berkeley, Calif, USA, 1995.

[15] Q. K. Pan, L. Wang, and B. Qian, "A novel differential evolution algorithm for bi-criteria no-wait flow shop scheduling problems," *Computers and Operations Research*, vol. 36, no. 8, pp. 2498–2511, 2009.

[16] H. Y. Wang, Y. W. Zhao, X. L. Xu, and W. L. Wang, "Scheduling batch and continuous process production based on an improved differential evolution algorithm," *System Engineering Theory and Practice*, vol. 29, no. 11, pp. 157–167, 2009.

[17] H. Y. Sang, Q. K. Pan, Y. X. Pan, and L. Wu, "A discrete differential evolution algorithm for lot-streaming flow shop scheduling problem," *Computer Simulation*, vol. 27, no. 7, pp. 292–295, 2010.

[18] Y. W. Zhao, H. Y. Wang, X. L. Xu, and W. L. Wang, "A new hybrid parallel algorithm for consistent-sized batch splitting job shop scheduling on alternative machines with forbidden intervals," *International Journal of Advanced Manufacturing Technology*, vol. 48, no. 9–12, pp. 1091–1105, 2010.

[19] Y. Zhao, H. Wang, W. Wang, and X. Xu, "New hybrid parallel algorithm for variable-sized batch splitting scheduling with alternative machines in job shops," *Chinese Journal of Mechanical Engineering*, vol. 23, no. 4, pp. 484–495, 2010.

[20] H. Y. Wang, Y. W. Zhao, W. L. Wang, and X. L. Xu, "New parallel algorithm based on DE for batch splitting job shop scheduling under multiple-resource constraints," *Control and Decision*, vol. 25, no. 11, pp. 1635–1644, 2010.

[21] G. Y. Shi, "A genetic algorithm applied to a classic job-shop scheduling problem," *International Journal of Systems Science*, vol. 28, no. 1, pp. 25–32, 1997.

[22] L. N. Xing, Y. W. Chen, P. Wang, Q. S. Zhao, and J. Xiong, "A knowledge-based ant colony optimization for flexible job shop scheduling problems," *Applied Soft Computing Journal*, vol. 10, no. 3, pp. 888–896, 2010.

[23] M. Yazdani, M. Amiri, and M. Zandieh, "Flexible job-shop scheduling with parallel variable neighborhood search algorithm," *Expert Systems with Applications*, vol. 37, no. 1, pp. 678–687, 2010.

[24] J. C. Tang, G. J. Zhang, B. B. Lin, and B. X. Zhang, "A hybrid algorithm for flexible Job-shop scheduling problem," in *Proceedings of the International Conference on Advanced in Control Engineering and Information Science*, pp. 3678–3683, Guangzhou, China, 2011.

[25] G. Zhang, L. Gao, and Y. Shi, "An effective genetic algorithm for the flexible job-shop scheduling problem," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3563–3573, 2011.

[26] H. Coskun and F. Kutay, "Continuous functions minimization by dynamic random search technique," *Applied Mathematical Modelling*, vol. 31, no. 10, pp. 2189–2198, 2007.

[27] I. Kacem, S. Hammadi, and P. Borne, "Pareto-optimality approach for flexible job-shop scheduling problems: hybridization of evolutionary algorithms and fuzzy logic," *Mathematics and Computers in Simulation*, vol. 60, no. 3–5, pp. 245–276, 2002.

[28] I. Kacem, S. Hammadi, and P. Borne, "Approach by localization and multiobjective evolutionary optimization for flexible job-shop scheduling problems," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 32, no. 1, pp. 1–13, 2002.

[29] P. Brandimarte, "Routing and scheduling in a flexible job shop by tabu search," *Annals of Operations Research*, vol. 41, no. 3, pp. 157–183, 1993.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 369056, 7 pages
http://dx.doi.org/10.1155/2013/369056

Research Article

Cloud Computing Platform for an Online Model Library System

Mingang Chen,[1,2] Wenjun Cai,[3] and Lizhuang Ma[1]

[1] *Department of Computer Science and Technology, Shanghai Jiao Tong University, Shanghai 200240, China*
[2] *Shanghai Key Laboratory of Computer Software Testing and Evaluating, Shanghai 20112, China*
[3] *Autodesk ACRD, Shanghai 201804, China*

Correspondence should be addressed to Mingang Chen; miles_cmg@sjtu.edu.cn

Received 22 January 2013; Accepted 22 March 2013

Academic Editor: Carlo Cattani

The rapid developing of digital content industry calls for online model libraries. For the efficiency, user experience, and reliability merits of the model library, this paper designs a Web 3D model library system based on a cloud computing platform. Taking into account complex models, which cause difficulties in real-time 3D interaction, we adopt the model simplification and size adaptive adjustment methods to make the system with more efficient interaction. Meanwhile, a cloud-based architecture is developed to ensure the reliability and scalability of the system. The 3D model library system is intended to be accessible by online users with good interactive experiences. The feasibility of the solution has been tested by experiments.

1. Introduction

It is generally known that the design of 3D virtual scene is the foundation for the production of animations and games, and 3D models are the most important components to construct 3D scenes. Although there have been some quick 3D model construction method [1, 2], the design and modeling of 3D models are extremely time consuming and labor intensive. Therefore, some online 3D model material libraries have been constructed, for example, the Google 3D Warehouse and other commercial 3D libraries like 3D Export, 3D Kingdom, and so forth.

Functions as browsing, storage, sharing, and reusing of models are provided by these web-based online 3D model libraries. Users can reuse models, save their time of modeling, and improve their efficiency of production with the application of 3D model libraries. Therefore, the construction of real-time 3D interactive, full-featured, and reliable system is the trends of the development of 3D model library.

At the same time, image-based proxy is usually used in most of the existing online 3D model libraries to display 3D model, which can only provide limited 3D information and lacking of human-computer interaction. Lately, the panorama technology [3] provides interactive features for the display of 3D model, but it cannot provide the function of model rotation at any angle and real-time rendering. Google has also launched 3D Warehouse, which provides interactive browsing of 3D models. However, it still has limitations. One is about the rotation of interaction, which tests to be effective only in the horizontal direction. The other is that only Google SketchUp model format is supported.

Web-based 3D model library requires users to contribute new models to the library, while it provides users with model sharing services, which is a typical Web 2.0 pattern. Cloud computing is a fresh paradigm for Web 2.0 that will replace the current dominant web application. So we deploy our Web 3D model library applications onto the cloud, for the advantages of a cloud platform like the scalability of computing and storage, load balancing, failover, and on demand service.

In this paper, a Web 3D model library is designed and implemented based on the cloud computing platform. And a front-end efficient Web 3D and back-end reliable cloud computing solutions are proposed. The paper makes the following contributions. (1) A model format converter is designed to support a wide range of model file formats in the system. (2) The model simplification and optimization

strategies are adopted to improve the efficiency of rich interaction such as scaling and rotation of arbitrary angle. (3) A model size adaptive adjustment module is proposed to fit the model browser quite well. (4) The Web 3D model library is implemented and deployed onto the cloud platform to ensure the stability and scalability of the system.

2. Related Work

Related technologies used in this paper mainly include Web 3D technology and cloud computing technology.

2.1. Web 3D Technology. There have been a large number of Web 3D systems and solutions over the last 20 years. Recently, there have been broadly two categories of Web 3D technologies: one is the browser plug-in-based approach, and the other is a small number of systems, which are trying to integrate the rendering system directly into the browser [4].

In developing content rich Internet applications, Adobe Flash is widely used as a multimedia platform, and it has become a popular solution used to add animation and interactivity to web pages. Flash supports 3D transformations and simulates 3D effects by displaying 2D vector shapes on the screen. According to Adobe, Flash Player has reached over 99% of the Internet users. Therefore, the flash-based technologies are the mainstream Web 3D solution.

Away3D is an open-source 3D graphics engine, written for the Adobe Flash platform in ActionScript 3, and runs in modern web browsers that utilize Adobe Flash Player. It can be used to render 3D models and perform various other 3D computations. It supports hierarchical object transformation with features such as translation, rotation and scaling, rendering of bitmap textures, and real-time lighting effects [5]. In this paper, Away3D technology is used in the Web 3D model library based on the following two reasons. Firstly, 3D effects are highly impressive and the performance is relatively good compared with other Flash 3D engines. The newest version, Away3D 4, has been upgraded to support the latest Flash Player with fully GPU-accelerated graphics. Secondly, Away3D supports multiple external model formats, such as the COLLADA, 3DS, and OBJ.

Other frameworks as Microsoft Silverlight and Oracle Java 3D are also used for developing rich Internet applications in recent years. However, users of Silverlight or Java 3D are much less compared with those using Adobe Flash.

The WebGL solution does not rely on a plug-in and it is a cross-platform web standard for low-level 3D graphics APIs based on OpenGL ES 2.0, exposed through the HTML5 Canvas element as Document Object Model interfaces [6]. The development of HTML5 will greatly promote the WebGL solution. Further work has been done with the X3DOM project [7], which overcomes WebGL limitations by providing an adaptation of the rendering technique. However, WebGL is a low-level interface, in which programmers have to deal with every triangle and transformation. Because of the above complex development process, it is not appropriate for the Web 3D model library.

X3D is an XML-based 3D computer graphics file [8]. The X3D specification includes graphics APIs and a full runtime that enables picking, viewing, navigation, scripting, and manipulating the scene graph. X3D supports DOM and Ajax programming and focuses on the visualization of 3D assets. However, so far the mainstream modeling software such as 3ds Max and Maya cannot transform 3D models directly into X3D format.

COLLAborative Design Activity (COLLADA) is an intermediate format with an open standard XML schema whose primary goal is to represent rich data in multiple forms, to exchange digital assets among various graphics software applications [9]. COLLADA defines the format for 3D assets but not their runtime semantics. In this paper, COLLADA is adopted as an interchange model format for the Web 3D model library because of its compatibility with the mainstream 3D modeling software and the well support by Away3D.

2.2. Cloud Computing Technology. In recent years, the cloud computing paradigm is becoming more popular for hosting applications, as it provides businesses with pay-as-you-go computing and storage services [10]. Growing numbers of corporations are involved in this wave to provide public cloud computing services, such as Amazon AWS, Google AppEngine, Microsoft Azure, and Rackspace CloudServer. These cloud providers offer a variety of options in pricing, performance, and feature set. For example, some offer platform as a service (PaaS), where a cloud customer builds applications using the APIs provided by the cloud; others offer infrastructure as a service (IaaS), where a customer runs applications inside virtual machines (VMs), using the APIs provided by their chosen guest operating systems [11].

In this paper, Amazon Web Services (AWSs) [12] are adopted as the cloud computing platform for Web 3D model library mainly due to flexibility of services and its great success in the business compared to Google AppEngine and Microsoft Azure. The Amazon cloud computing platform provides the necessary computing environment, including CPUs, storage, memory, networking, and operating system. The Web 3D model library makes use of three AWSs: Elastic Compute Cloud (EC2), Elastic Block Storage (EBS), and Simple Storage Service (S3). EC2 contains a variety of user selectable virtual machine types, also called instances, that range in computing power and cost. An instance is booted within a few minutes, and the user is given root or administrator access.

3. The Architecture Overview of the Web 3D Model Library System

A classic three-layer architecture is adopted in the Web 3D model library system, that is, data layer, business layer, and presentation layer, as shown in Figure 1.

The data layer deployed in Amazon S3 stands for model files storage. S3 is composed of buckets that are geographically distributed across Amazon's multiple data centers in different locations. So S3 is an extremely reliable persistent storage

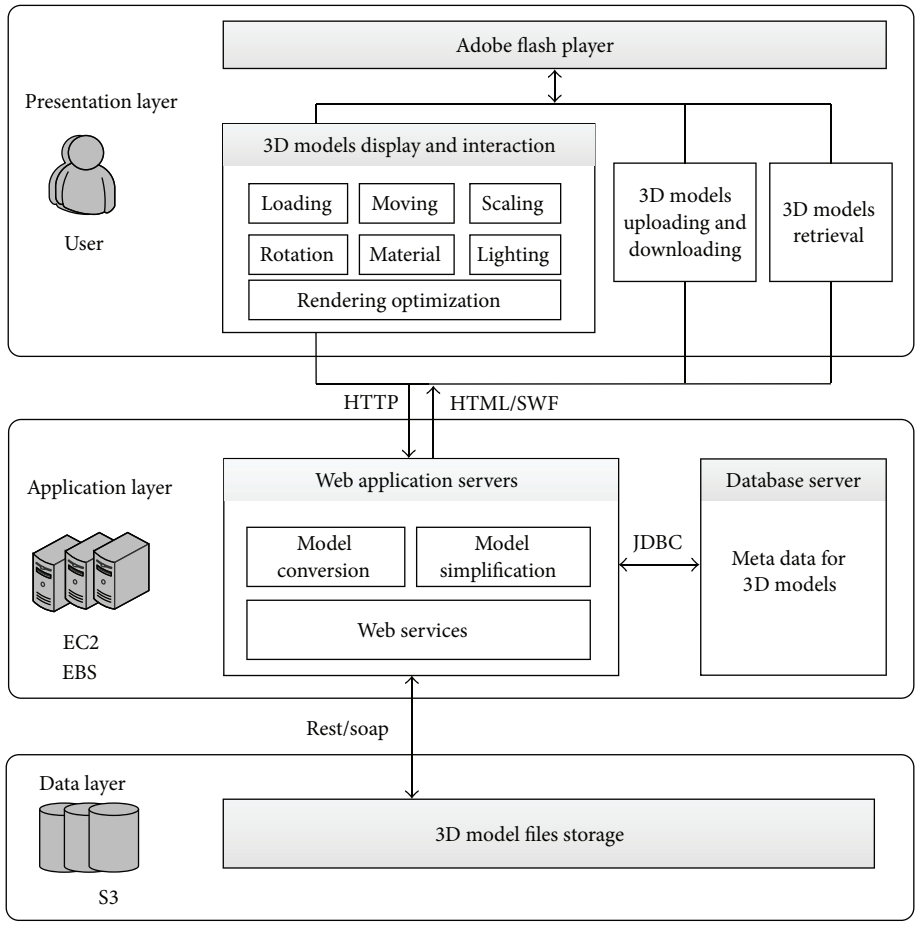


Figure 1: The system architecture of Web 3D model library.

system for files storage. When the storage capacity of model files is larger than expectation, storage expansion service of S3 can be easily requested.

Web application servers and models metadata database are deployed on Amazon EC2 that mainly provides business logic of the system, such as user's requests, conversion of model formats, and model's retrieval. Amazon EC2 allows users to deploy scalable resources on demand.

The presentation layer in the user's browser provides 3D model display and interaction, model uploading and downloading functions, and so forth.

This paper will focus on how the system supports 3D model display and interaction, including model format conversion, model mesh simplification, model size adaptive adjustment, and the design of model browser.

4. Model Format Conversion and Model Simplification

The most widespread model files in our model library are 3DS Max and Maya formats. However, these two model formats cannot be rendered in the Away3D engine directly. So in this paper, we design a model format conversion service on the server side, and this service is for converting various model formats into the COLLADA format.

When models are uploaded to the system, the model format conversion service will first analyze the format of models and then move the COLLADA files into the "COLLADA" folder, while it moves models of other formats into the "NON-COLLADA" folder. The service will get files in the "NON-COLLADA" folder one by one and open it in the relevant modeling software which has installed COLLADA plug-in, and then the format conversion command script is executed. On the server side, the model format conversion service runs periodically. At last, the converted files will be moved to the "COLLADA" folder. Away3D engine will load and render the COLLADA model file for the model's display and interaction.

Furthermore, taking into account complex models may be difficult to real-time 3D interaction, so we also implement a model simplification service on the server side. An iterative decimation algorithm [13] is adopted for the model simplification. This algorithm applies the pair collapse operation iteratively, which replaces two vertices with one, causing neighboring faces to be degenerated. The algorithm defines the quadric error metric for selecting a collapsed pair of vertices from the potential candidates. For a vertex v,

the quadric error metric is defined as the sum of point-plane distance from v to a set of associated planes, as shown in the following formula:

$$\begin{aligned}\Delta(v) &= \sum_{p\in\text{planes}(v)} \left(p^T v\right)^2 \\ &= v^T \left(\sum_{p\in\text{planes}(v)} p^T p\right) v \\ &= v^T Q_v v,\end{aligned} \tag{1}$$

where planes(v) denotes the triangle faces adjacent to v, and Q_v is the error quadric matrix. In the model simplification process, we assign planes(v_k) = planes(v_i) ∪ planes(v_j) by iteratively applying the pair collapse $(v_i, v_j) \rightarrow v_k$.

5. Model Size Adaptive Adjustment

When models are just loaded onto the scene, some are too large to be contained by the screen while others are so small that customers have to pay great efforts to zoom them up. To solve this problem, we have deeply researched into the perspective projection principle of Away3D engine and proposed a method to adjust related parameter automatically and display models of various sizes properly.

The perspective projection matrix P in Away3D can be represented by the following formula:

$$\begin{aligned}P &= \begin{pmatrix} \dfrac{f}{\text{aspect}} & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & -\dfrac{N+F}{N-F} & \dfrac{2NF}{N-F} \\ 0 & 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \dfrac{N}{W} & 0 & 0 & 0 \\ 0 & \dfrac{N}{H} & 0 & 0 \\ 0 & 0 & -\dfrac{N+F}{N-F} & \dfrac{2NF}{N-F} \\ 0 & 0 & 1 & 0 \end{pmatrix},\end{aligned} \tag{2}$$

where aspect = W/H, $f = 1/\tan(\text{FOV}/2) = N/H$. W and H denote the half of the width and height of the viewport, respectively. N and F denote the distance from the camera to the near plane and the far plane, respectively. FOV is illustrated in Figure 2.

When a 3D point $p(x, y, z)$ is projected onto the projection plane, its 2D coordinate is $p' = (Nx/W, Ny/N)$. So the 3D model bounding box's width and height can be calculated by the following formula:

$$\begin{aligned}\text{model Width} &= \frac{N}{W}\left(x_{\max} - x_{\min}\right), \\ \text{model Height} &= \frac{N}{H}\left(y_{\max} - y_{\min}\right),\end{aligned} \tag{3}$$

where $x_{\max}$ and $x_{\min}$ denote the maximum and minimum x-coordinate value of all points in a model. Similarly, $y_{\max}$ and $y_{\min}$ denote the maximum and minimum y-coordinate value of all points in a model. model Width and model Height denote the model's width and height when it is projected onto the model browser's screen. According to formula (3), we can adjust the distance from the camera to near plane N to a proper value adaptively when models are loaded. Figure 3 compares the display effect of loading models before and after the model size adaptive adjustment.

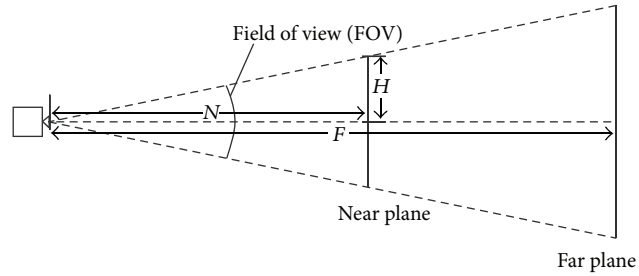


Figure 2: An illustration of field of view.

6. Design of the Model Browser

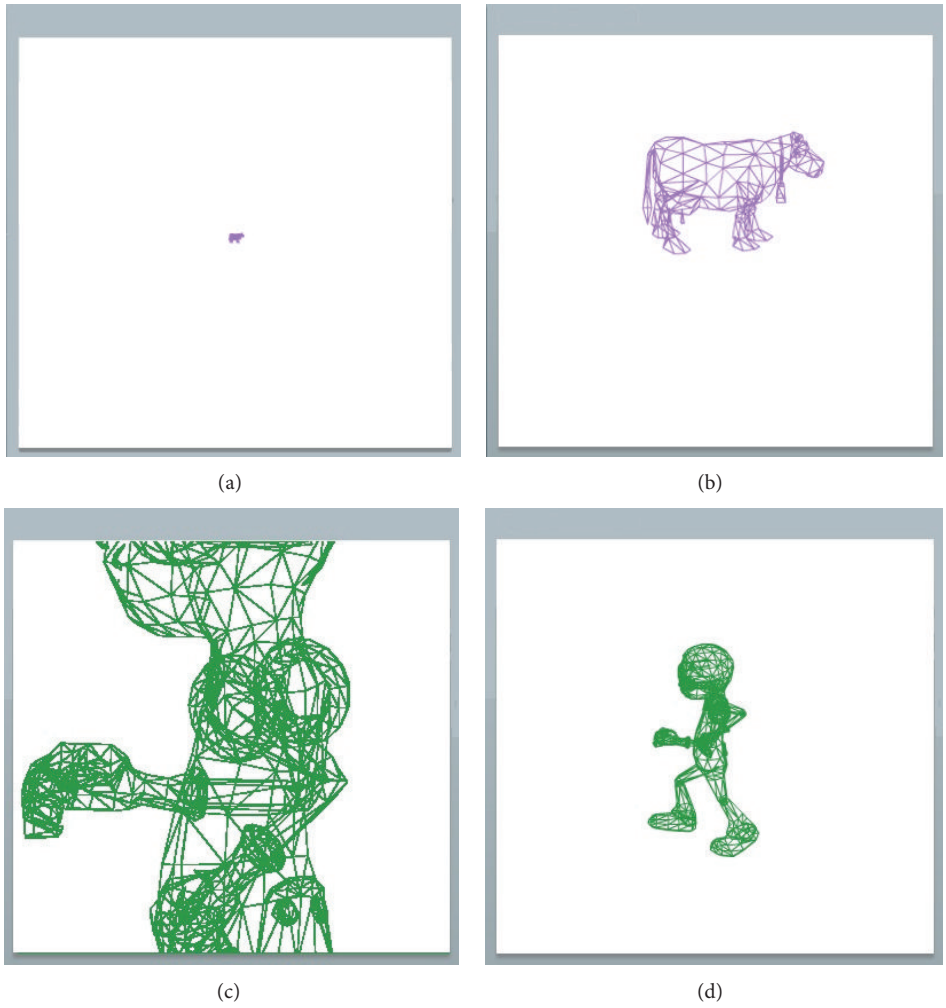
The model browser based on Away3D is a window for displaying 3D models with multiangle visibility and rich human-computer interaction. The browser supports not only the model rotation, scaling, and translation, but also advanced operations such as material replacement and light source adjustment.

Away3D engine provides rich APIs to process real-time rendering of 3D models. Model's rotation, scaling, and translation operations are implemented by updating the attribute of camera position; model's material replacement and adjustment of light source by updating the attribute of material and light. Away3D engine supports point light, ambient light, or directional light for 3D scene and also specifies ambient shader or diffuse Phong shader, and so forth.

In the following part, we take model's rotation as an example to describe model real-time rendering process in the model browser. Firstly, the model browser program loads 3D model by calling Away3D Collada class and parses model files described with XML. After parsing process, we can get model's information, such as vertex's coordinates and model material. Secondly, the model browser program adds 3D model to the scene and computes the camera's position parameter by formula (3). Thirdly, the event listener functions are registered in the scene to response users real-time interaction. For model's rotation, we just compute the angle of model's rotation. Finally, the model browser program calls Away3D rendering function to realize model's real-time rendering.

7. Application and Performance Analysis

The Web 3D model library system is intended to be accessed by online users and provide users with good interactive experiences, as shown in Figure 4.



(a) (b) (c) (d)

Figure 3: (a) and (c) are effects before the model size adaptive adjustment. (b) and (d) are effects after the model size adaptive adjustment.

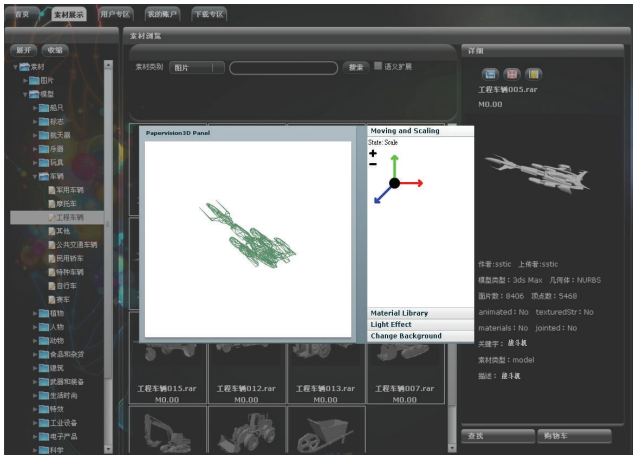


Figure 4: Web 3D model library system.

Interactions with 3D models such as rotation, scaling, and translation in the model browser can be realized easily by using a mouse, as shown in Figure 5.

The model browser also supports users to upload images as texture and material and assign them to a model or part of a model, as shown in Figure 6. Figure 6(a) is an original car model, and Figure 6(b) shows that certain parts of car model are selected and are assigned with new materials. Figure 6(c) shows the effect of changing the material of the car model.

Figure 7 shows the light effects of a plane model. The interface is designed for user to change the position of light source, light color, and ambient color. Figure 7(a) shows a plane model without light effects, and Figures 7(b) and 7(c) are the high light and ambient light effects, respectively.

The model rendering time and model browser response time are the important indicators for the model browser's performance. In order to enhance the model browser's performance, in this paper the following optimizing techniques are adopted: (1) choosing suitable Flash stage quality, (2) applying camera's frustum culling process, that is, removing objects that lie outside the viewing frustum from the rendering process [14], and (3) keeping the amount of pixels that need to be rendered as low as possible.

Figure 5: 3D model's rotation, scaling, and translation.

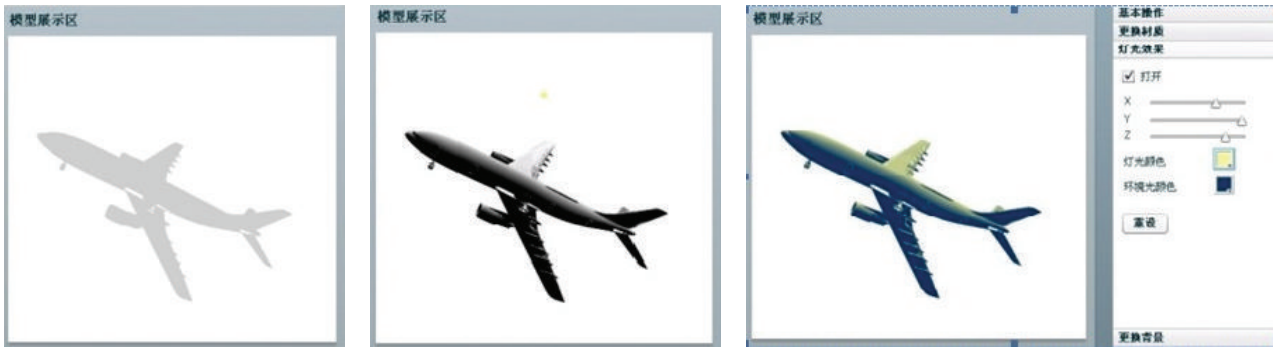
(a) (b) (c)

Figure 6: The effect of changing the material of a car model.

We choose two models with different complexity to test the model browser's performance, as shown in Table 1. All performance analysis was executed on a laptop with Intel i3 CPU at 2.4 GHz and 4 G memory. From the performance analysis we observe that (1) the more model complex is, the more memory SWF file takes up, and the longer rendering time is, the lower frame rate is and (2) after optimization, memory cost and rendering time are reduced significantly.

8. Conclusion

A web 3D model library system solution and complete design are proposed in this paper. The system is deployed in the Amazon cloud computing platform to ensure its reliability and scalability. In the system, a 3D model conversion service is developed to support a variety of model formats. Furthermore, a model simplification algorithm and a model size



(a) (b) (c)

Figure 7: The light effects of a plane model.

Table 1: Performance analysis of the model browser.

	Windmill model		Plane model	
	Vertex number: 1362		Vertex number: 2724	
	Facet number: 1218		Facet number: 2325	
	Before optimization	After optimization	Before optimization	After optimization
Memory (MB)	17.82	12.56	24.38	18.32
Rendering time (ms)	6	4	12	8
Frame rate (fps)	24	32	16	24

adaptive method are designed to improve user's 3D interaction experience. This Web 3D model library solution of design and deployment can be applied to other web applications, such as online virtual tour and design. In further studies, we are planning to design a real-time model simplification method to improve the display accuracy of the complex models.

Acknowledgments

The authors would like to thank professor Jinyuan Jia for his support and valuable suggestions during the project research. This work was funded by the National Natural Science Foundation of China (61133009) and Science and Technology Commission of Shanghai Municipality program (11511500200, 12QB1402300, and 12DZ2290700).

References

[1] M. Pollefeys, D. Nistér, J. M. Frahm et al., "Detailed real-time urban 3D reconstruction from video," *International Journal of Computer Vision*, vol. 78, no. 2-3, pp. 143–167, 2008.

[2] H. Liu, S. Chen, and N. Kubota, "Special section on intelligent video systems and analytics," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, p. 90, 2012.

[3] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.

[4] J. Behr, P. Eschler, Y. Jung, and M. Zöllner, "X3DOM—a DOM-based HTML5/ X3D integration model," in *Proceedings of the14th International Conference on 3D Web Technology (Web3D '09)*, pp. 127–135, June 2009.

[5] C. Matthew, *Away3D 3.6 Essentials*, Packt Publishing, 2011.

[6] C. Marrin, *WebGL Specification*, Khronos WebGL Working Group, 2011.

[7] J. Behr, Y. Jung, J. Keil et al., "A scalable architecture for the HTML5/X3D integration model X3DOM," in *Proceedings of the 15th International Conference on Web 3D Technology (Web3D '10)*, pp. 185–194, July 2010.

[8] L. Daly and D. Brutzman, "X3D: extensible 3D graphics standard [Standards in a Nutshell]," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 130–135, 2007.

[9] A. Rémi and P. Tony, "Developing web application with COLLADA and X3D: a whitepaper," 2007.

[10] M. Cusumano, "Cloud computing and SaaS as new computing platforms," *Communications of the ACM*, vol. 53, no. 4, pp. 27–29, 2010.

[11] A. Li, X. Yang, S. Kandula, and M. Zhang, "CloudCmp: comparing public cloud providers," in *Proceedings of the 10th Internet Measurement Conference (IMC '10)*, pp. 1–14, November 2010.

[12] Amazon Services, http://aws.amazon.com/.

[13] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 209–216, August 1997.

[14] U. Assarsson and T. Möller, "Optimized view frustum culling algorithms for bounding boxes," *Journal of Graphics Tools*, vol. 5, no. 1, pp. 9–22, 2000.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 313868, 6 pages
http://dx.doi.org/10.1155/2013/313868

Research Article

The Inverse 1-Median Problem on Tree Networks with Variable Real Edge Lengths

Longshu Wu,[1] Joonwhoan Lee,[2] Jianhua Zhang,[3] and Qin Wang[1]

[1] *College of Science, China Jiliang University, Hangzhou 310018, China*
[2] *Division of Computer Science and Engineering, Chonbuk National University, Jeonju, Jeonbuk 561-756, Republic of Korea*
[3] *College of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Qin Wang; wq@cjlu.edu.cn

Received 25 January 2013; Accepted 22 March 2013

Academic Editor: Carlo Cattani

Location problems exist in the real world and they mainly deal with finding optimal locations for facilities in a network, such as net servers, hospitals, and shopping centers. The inverse location problem is also often met in practice and has been intensively investigated in the literature. As a typical inverse location problem, the inverse 1-median problem on tree networks with variable real edge lengths is discussed in this paper, which is to modify the edge lengths at minimum total cost such that a given vertex becomes a 1-median of the tree network with respect to the new edge lengths. First, this problem is shown to be solvable in linear time with variable nonnegative edge lengths. For the case when negative edge lengths are allowable, the NP-hardness is proved under Hamming distance, and strongly polynomial time algorithms are presented under l_1 and l_∞ norms, respectively.

1. Introduction

Location problems have found important applications in operations research, computer science, and logistics management. The p-median problem is a typical location problem and has been intensively considered in the literature. The p-median problem on an undirected graph is to locate p facilities on edges or vertices of the graph such that the sum of the weighted distances of the vertices to the closest facility is minimized. This problem possesses the vertex optimality property which asserts that there exists an optimal solution such that all facilities are located at the vertices of the graph. Kariv and Hakimi [1] showed that the p-median problem is NP-hard even if the graph G is a planar graph of maximum degree 3. Therefore, many investigations are dedicated to special graph classes which allow a more efficient algorithm. In the case of tree networks, the 1-median problem is solvable in linear time by Hua [2] and Goldman [3].

However, in an established network, the locations of facilities have already been fixed. The changing environment might make the existing facilities deviate from the median places of the network. Instead of finding optimal locations, the task is to modify parameters at minimum cost such that the prespecified locations become optimal. This arises what we call the inverse p-median problem. In recent years, inverse p-median problem found an increasing interest [4–8].

Burkard et al. [9] considered inverse p-median problems with variable vertex weights and showed that discrete inverse p-median problems with real weights can be solved in polynomial time, if p is fixed and not an input parameter.Burkard et al. [9, 10] investigated the inverse 1-median problem with variable vertex weights and showed that the problem is solvable by a greedy-type algorithm in $O(n \log n)$ time if the underlying network is a tree or the location problem is defined in the plane (if the distances between the points are measured in the Manhattan or maximum metric) and in $O(n^2)$ time on cycles. The inverse Fermat-Weber problem was studied by Burkard et al. [11]. They derived a combinatorial approach which solves the problem in $O(n \log n)$ time for unit cost and under the assumption that the prespecified point that should become a 1-median does not coincide with a given point in the plane. Galavii [12] proposed linear time

algorithms for the inverse 1-median problem with variable vertex weights on trees (with nonnegative weights) and on paths (with negative weights while the weights of endpoints are arbitrary). Under weighted Hamming distance, Guan and Zhang [13] presented polynomial time algorithm for the inverse 1-median problem with variable vertex weights on trees. Bonab et al. [14] showed that the inverse p-median problem with variable edge lengths is strongly NP-hard on general graphs and weakly NP-hard on series-parallel graphs. In [14], the authors also presented a polynomial time algorithm for the inverse 2-median problem with positive variable edge lengths on trees and a linear time algorithm for the special case of star graphs. In this paper, we first provide a linear time algorithm to solve the inverse 1-median problem on tree networks with variable nonnegative edge lengths. For the case when negative edge lengths are allowable, we show that this problem is NP-hard under Hamming distance and we propose strongly polynomial time algorithms under l_1 and l_∞ norms, respectively.

The paper is organized as follows. In Section 2, we define the inverse 1-median problem on trees with variable edge lengths, then we show some properties of this problem. Based on these properties, a linear time algorithm for variable nonnegative edge lengths is presented; when negative edge lengths are allowable, the NP-hardness of the problem under Hamming distance is proved and the strongly polynomial time algorithms of the problem under l_1 and l_∞ norms are proposed in Section 3. Some final remarks are given in Section 4.

2. Preliminaries

Let $T = (V, E)$ be an undirected tree graph with vertex set $V = \{v_1, \ldots, v_n\}$ and edge set E with $|E| = n - 1$. The weight of a vertex $v \in V$ is denoted by $w_v \geq 0$, and the length of each edge $e \in E$ is denoted by $l_e > 0$. The 1-median problem on trees is to locate a facility on V such that the sum of the weighted distances of the vertices to this facility is minimized.

Let $C(a) : R^E \to R$ be a criterion function (having length vectors as argument) such that $C(0) = 0$ and $C(a) \leq C(b)$ if $a \leq b$. (For $a, b \in R^E$, $a \leq b$ means that $a_e \leq b_e$, for all $e \in E$.) Suppose that the unique path from $s \in V$ to $t \in V$ is denoted by $P(s,t)$. Let v_0 be a specific vertex in V. Denote the adjusted edge length vector by x. Let $d_x(s,t)$ denote the length of $P(s,t)$ with respect to x. The lower bound of x_e is denoted by $\underline{l}_e \leq 0$ for each $e \in E$. Readers can refer to [15] for other graph theory terms not defined here. The inverse 1-median problem on tree networks with variable real edge lengths can be formally described as follows:

$$\begin{aligned} \min \quad & C(|l - x|) \\ \text{s.t.} \quad & \sum_{v \in V \setminus \{v_0\}} w_v d_x(v_0, v) - \sum_{v \in V \setminus \{\overline{v}\}} w_v d_x(\overline{v}, v) \leq 0, \\ & \qquad \forall \overline{v} \in V \setminus \{v_0\}, \\ & x_e \geq \underline{l}_e, \quad e \in E. \end{aligned} \tag{1}$$

If we delete an edge e from T, the tree T will be partitioned into two subtrees, one of which containing the prespecified vertex v_0 is denoted by $T'_e = (V'_e, E'_e)$, and the other is denoted by $T''_e = (V''_e, E''_e)$, where V'_e, V''_e denote the vertex sets and E'_e, E''_e denote the edge sets of T'_e, T''_e, respectively. Let $D_x(t) = \sum_{v \in V \setminus \{t\}} w_v d_x(t, v)$, and let $B_e = \sum_{v \in V''_e} w_v - \sum_{v \in V'_e} w_v$ for each $e \in E$; then for any $\overline{v} \in V \setminus \{v_0\}$,

$$\begin{aligned} D_x(v_0) - D_x(\overline{v}) &= \sum_{v \in v \setminus \{v_0\}} w_v d_x(v_0, v) - \sum_{v \in v \setminus \{\overline{v}\}} w_v d_x(\overline{v}, v) \\ &= \sum_{e \in P(v_0, \overline{v})} \left(\sum_{v \in V''_e} w_v - \sum_{v \in V'_e} w_v \right) x_e \\ &= \sum_{e \in P(v_0, \overline{v})} B_e x_e, \end{aligned} \tag{2}$$

and the first constraint of (1) can be changed to

$$\sum_{e \in P(v_0, \overline{v})} B_e x_e \leq 0, \quad \forall \overline{v} \in V \setminus \{v_0\}. \tag{3}$$

So, we can simplify (1) as

$$\begin{aligned} \min \quad & C(|l - x|) \\ \text{s.t.} \quad & \sum_{e \in P(v_0, \overline{v})} B_e x_e \leq 0, \quad \forall \overline{v} \in V \setminus \{v_0\}, \\ & x_e \geq \underline{l}_e, \quad e \in E. \end{aligned} \tag{4}$$

For any $v_r \in V \setminus \{v_0\}$, without loss of generality, suppose the sequence of vertices and edges in $P(v_0, v_r)$ is v_0, e_1, v_1, e_2, $v_2, \ldots, v_{r-1}$, e_r, v_r. We have

$$\begin{aligned} \sum_{v \in V''_{e_i}} w_v &\leq \sum_{v \in V''_{e_{i-1}}} w_v - w_{v_{i-1}}, \\ \sum_{v \in V'_{e_i}} w_v &\geq \sum_{v \in V'_{e_{i-1}}} w_v + w_{v_{i-1}}. \end{aligned} \tag{5}$$

Therefore,

$$\begin{aligned} \sum_{v \in V''_{e_i}} w_v - \sum_{v \in V'_{e_i}} w_v &\leq \left(\sum_{v \in V''_{e_{i-1}}} w_v - w_{v_{i-1}} \right) - \left(\sum_{v \in V'_{e_{i-1}}} w_v + w_{v_{i-1}} \right) \\ &= \sum_{v \in V''_{e_{i-1}}} w_v - \sum_{v \in V'_{e_{i-1}}} w_v - 2w_{v_{i-1}}. \end{aligned} \tag{6}$$

Then, for $i = 1, 2, \ldots, r$, by (2), we have

$$D_x(v_0) - D_x(v_i) = \sum_{e \in P(v_0, v_i)} B_e x_e = \sum_{j=1}^{i} B_{e_j} x_{e_j},$$

$$\begin{aligned} B_{e_i} - B_{e_{i-1}} &= \left(\sum_{v \in V''_{e_i}} w_v - \sum_{v \in V'_{e_i}} w_v \right) - \left(\sum_{v \in V''_{e_{i-1}}} w_v - \sum_{v \in V'_{e_{i-1}}} w_v \right) \\ &\leq \left(\left(\sum_{v \in V''_{e_{i-1}}} w_v - \sum_{v \in V'_{e_{i-1}}} w_v \right) - 2w_{v_{i-1}} \right) \\ &\quad - \left(\sum_{v \in V''_{e_{i-1}}} w_v - \sum_{v \in V'_{e_{i-1}}} w_v \right) \\ &= -2w_{v_{i-1}} \\ &\leq 0. \end{aligned} \tag{7}$$

According to (7), we further obtain the following two formulas:

$$D_x(v_0) - D_x(v_j) = D_x(v_0) - D_x(v_i) + \sum_{e \in P(v_i, v_j)} B_e l_e, \quad 0 \leq i < j \leq r, \tag{8}$$

$$B_{e_{i-1}} \geq B_{e_i}, \quad i = 2, \ldots, r. \tag{9}$$

From the above analysis, we have the following.

Theorem 1. *If there exists a vertex $v_t \in V$ satisfying $D_l(v_0) - D_l(v_t) \leq 0$, then $B_{e_t} \leq 0$ and $D_l(v_0) - D_l(v) \leq 0$ for each $v \in V''_{e_t}$, where e_t denotes the edge of the path $P(v_0, v_t)$ incident with vertex v_t.*

Proof. By contradiction. If $B_{e_t} > 0$, then from (9) we have $B_e > 0$ for each edge $e \in P(v_0, v_t)$. By (2), $D_l(v_0) - D_l(v_t) = \sum_{e \in P(v_0, v_t)} B_e l_e > 0$, which contradicts the fact that $D_l(v_0) - D_l(v_t) \leq 0$. So $B_{e_t} \leq 0$. As $B_e \leq B_{e_t} \leq 0$, for each $e \in P(v_t, v)$ and $v \in V''_{e_t}$, by (8), we can deduce that $D_l(v_0) - D_l(v) = D_l(v_0) - D_l(v_t) + \sum_{e \in P(v_t, v)} B_e l_e \leq 0$. This completes the proof. □

Based on the above properties, we will present some algorithms of the problem under different norms and analyses their complexities in the next section.

3. Algorithms and Complexities

3.1. The Problem with Variable Nonnegative Edge Lengths (I1MT-N). In this section, we consider the following problem:

> *I1MT-N.* When $\underline{l}_e = 0$ for each $e \in E$, we call (4) Problem I1MT-N, which is the inverse 1-median problem on trees with variable nonnegative edge lengths.

Denote the optimal solution of Problem I1MT-N by l^*. We present Algorithm 2 to solve this problem. particulary, Algorithm 2 can solve Problem I1MT-N under all kinds of norms only by edge length reduction.

Algorithm 2.

Step 1. Let $\overline{E} = \emptyset$, $\overline{V} = \{v_0\}$.

Step 2. Let $E' = \{e' \mid e' = (v, \overline{v}) \in E \setminus \overline{E}, v \in V \setminus \overline{V}, \overline{v} \in \overline{V}\}$. For each $e' \in E'$,

(1) if $B_{e'} > 0$, then let $l^*_{e'} = 0$ and $\overline{E} = \overline{E} \cup \{e'\}$;

(2) if $B_{e'} \leq 0$, then for each $e \in \{e'\} \cup E''_{e'}$, let $l^*_e = l_e$ and $\overline{E} = \overline{E} \cup \{e'\} \cup E''_{e'}$.

Step 3. Let $\overline{V} = \{v \mid (v, \overline{v}) = e' \in E', \overline{v} \in \overline{V} \text{ and } B_{e'} > 0\}$. If $\overline{E} = E$, stop. Otherwise, turn to Step 2.

Theorem 3. *Algorithm 2 correctly finds the optimal solution l^* of Problem I1MT-N in linear time.*

Proof. First, we show that l^* obtained by Algorithm 2 is a feasible solution of Problem I1MT-N. It is clear that $l^*_e = 0$ or $l^*_e = l_e$; so it satisfies $l^*_e \geq 0$ for each $e \in E$. For each vertex $v \in V \setminus \{v_0\}$ and $e \in P(v_0, v)$, from Step 2 of Algorithm 2, when $B_e > 0$, $l^*_e = 0$; and when $l^*_e = l_e > 0$, $B_e \leq 0$. Hence, $\sum_{e \in P(v_0, v)} B_e l^*_e \leq 0$ for each $v \in V \setminus \{v_0\}$. Therefore, l^* is a feasible solution of Problem I1MT-N. □

Suppose l^* is not the optimal solution of Problem I1MT-N and suppose that the optimal solution is l'. Then under l' there exists at least one edge $\overline{e} \in E$ satisfying one of the following two cases:

Case 1 ($B_{\overline{e}} > 0$ and $l'_{\overline{e}} > 0$). Without loss of generality, suppose $\overline{e} = (v_{i-1}, v_i) \in P(v_0, v_i) = v_0, e_1, v_1, e_2, v_2, \ldots, v_{i-1}, e_i, v_i$. By (9), $B_{e_j} \geq B_{\overline{e}} > 0$ for $j = 1, \ldots, i-1$. Notice that $l'_e \geq 0$ for each $e \in E$. We have $\sum_{j=1}^{i} B_{e_j} l'_{e_j} > 0$, which contradicts the feasibility of l'.

Case 2 ($B_{\overline{e}} \leq 0$ and $l'_{\overline{e}} \geq 0$, $l'_{\overline{e}} \neq l_{\overline{e}}$). We define a new vector $l'' \in R^E$ such that

$$l''_e = \begin{cases} l_e, & \text{if } e = \overline{e}, \\ l'_e, & \text{otherwise.} \end{cases} \tag{10}$$

From the result of Case 1, we know that when $B_e > 0$, $l''_e = l'_e = 0$. And it is easy to see that $l''_e \geq 0$ for each $e \in E$. Then, we have $\sum_{e \in P(v_0, v)} B_e l''_e = \sum_{e \in P(v_0, v), B_e \leq 0} B_e l''_e \leq 0$ for each vertex $v \in V$. So l'' is a feasible solution of Problem I1MT-N. But $C(|l - l''|) < C(|l - l'|)$, a contradiction to the optimality assumption of l'.

Therefore, Algorithm 2 correctly finds the optimal solution l^* of Problem I1MT-N.

The main computation of Algorithm 2 is to compute B_e for each edge $e \in E$, then to modify the corresponding edge length of e according to B_e from Step 2. Notice that

$$\begin{aligned} B_e &= \sum_{v \in V_e''} w_v - \sum_{v \in V_e'} w_v \\ &= 2 \sum_{v \in V_e''} w_v - \sum_{v \in V} w_v, \end{aligned} \tag{11}$$

where $\sum_{v \in V} w_v$ is a constant. Let v_0 be the root vertex and conduct the breadth-first method to T. We can compute B_e iteratively from the edges incident with the leaves at the bottom layer up to v_0 in the rooted tree. So Problem I1MT-N can be solved in $O(|E|)$ time by using Algorithm 2.

From Algorithm 2, we know that for Problem I1MT-N it is not necessary to increase the edge lengths of the tree network. So to solve Problem I1MT-N is equivalent to solving the following program:

$$\begin{aligned} \min \quad & C(l - x) \\ \text{s.t.} \quad & \sum_{e \in P(v_0, \overline{v})} B_e x_e \le 0, \quad \forall \overline{v} \in V \setminus \{v_0\}, \\ & 0 \le x_e \le l_e, \quad e \in E. \end{aligned} \tag{12}$$

3.2. The Problem with Variable Real Edge Lengths

3.2.1. The Model Under l_1 Norm with Variable Real Edge Lengths (I1MT-R-l_1). In this section, we consider the following inverse 1-median problem on trees with variable real edge lengths under l_1 norm when edge length increase is forbidden.

> *I1MT-R-l_1.* When $\underline{l}_e < 0$ for each $e \in E$ and the criterion function $C(a) : R^E \to R$ is a function such that $C(a) = ca$, where c is the cost vector and $c_e > 0$ is the cost of decreasing l_e by one unit for each $e \in E$, we call (4) Problem I1MT-R-l_1, which is the inverse 1-median problem on trees with variable real edge lengths under l_1 norm.

Problem I1MT-R-l_1 can be formally described as follows:

$$\begin{aligned} \min \quad & \sum_{e \in E} c_e (l_e - x_e) \\ \text{s.t.} \quad & \sum_{e \in P(v_0, \overline{v})} B_e x_e \le 0, \quad \forall \overline{v} \in V \setminus \{v_0\}, \\ & \underline{l}_e \le x_e \le l_e, \quad e \in E. \end{aligned} \tag{13}$$

For a path $P(v_0, v_r) = v_0, e_1, v_1, e_2, v_2, \ldots, v_{r-1}, e_r, v_r$, suppose that $B_{e_{i-1}} > 0$ and $B_{e_i} \le 0$, where $2 \le i \le r$. Denote the optimal solution of Problem I1MT-R-l_1 by l^{opt}. We have the following.

Theorem 4. *$l_{e_j}^{\text{opt}} = l_{e_j}$ for $j = i, i+1, \ldots, r$.*

Proof. By (9), clearly $B_{e_j} > 0$ for $j = 1, 2, \ldots, i-1$ and $B_{e_i} \le 0$ for $j = i, i+1, \ldots, r$. Suppose to the contrary that there exists an edge $e_k \in P(v_{i-1}, v_r)$ such that $\underline{l}_{e_k} \le l_{e_k}^{\text{opt}} < l_{e_k}$. Similar to the proof of Case 2 in Theorem 3, we define a new feasible solution l'' such that $l_{e_k}'' = l_{e_k}$ and $l_{e_j}'' = l_{e_j}^{\text{opt}}$ when $j \ne k$. Then we have $\sum_{e \in P(v_0, v)} B_e l_e'' \le \sum_{e \in P(v_0, v)} B_e l_e^{\text{opt}} \le 0$ for each vertex $v \in V$. But $c_e(l_e - l_e'') < c_e(l_e - l_e^{\text{opt}})$, and therefore $c(l - l'') < c(l - l^{\text{opt}})$, a contradiction to the fact that l^{opt} is the optimal solution of Problem I1MT-R-l_1. This completes the proof. □

From Theorem 4, we only need to determine the optimal adjusted edge lengths l_e^{opt} with $B_e > 0$. We construct the subtree $T^S = (V^S, E^S)$ of T with $E^S = \{e \mid B_e > 0, e \in E\}$, $T^S = T[E^S]$, and $V^S = V(T^S)$. Clearly, $v_0 \in V^S$. Regard T^S as a directed tree with the orientation from the root v_0 to all leaves (e.g., by conducting the breadth-first method). For each $v \in V^S$, let $P(v)$ denote the path from v_0 to v on the tree T^S. We can simplify (13) as

$$\begin{aligned} \min \quad & \sum_{e \in E^S} c_e (l_e - x_e) \\ \text{s.t.} \quad & \sum_{e \in P(v)} B_e x_e \le 0, \quad \forall v \in V^S \setminus \{v_0\}, \\ & \underline{l}_e \le x_e \le l_e, \quad \forall e \in E^S. \end{aligned} \tag{14}$$

In [16, 17], the authors considered the shortest path improvement problem, which is how to shorten the lengths of edges with as less cost as possible such that the distances between specified sources and terminals are reduced to the required bounds. It is not difficult to see that (14) is in fact a shortest path improvement problem for tree networks with single source v_0 and the required bound 0 under l_1 norm. Let $\theta_e = B_e(l_e - x_e) \ge 0$, $b_e = B_e(l_e - \underline{l}_e) > 0$, and $p_e = c_e / B_e > 0$ for each $e \in E^S$. Let $r_v = \sum_{e \in P(v)} B_e l_e > 0$ for each $v \in V^S$. Equation (14) can be formulated as the following linear program:

$$\begin{aligned} \min \quad & \sum_{e \in E^S} p_e \theta_e \\ \text{s.t.} \quad & \sum_{e \in P(v)} \theta_e \ge r_v, \quad \forall v \in V^S \setminus \{v_0\}, \\ & 0 \le \theta_e \le b_e, \quad \forall e \in E^S. \end{aligned} \tag{15}$$

Equation (15) is a combinatorial linear program and can be solved by a strongly polynomial time algorithm [18]. Moreover, (15) can be solved much faster [16, 17]. In [16], the authors presented a steepest descent-type combinatorial algorithm to solve this problem, which runs in $O(|V|^2)$ time. In [17], the authors formulated the dual problem of (15) as a minimum cost flow problem on a two-terminal serial-parallel graph and then gave an algorithm to obtain the primal optimal solution of (15), with the whole procedure running in $O(|V| \log |V|)$ time.

From the above analysis, we have the following theorem.

Theorem 5. *Problem I1MT-R-L_1 can be solved in $O(|V|\log|V|)$ time.*

3.2.2. The Model under Hamming Distance with Variable Real Edge Lengths (I1MT-R-H). Now we consider the model under Hamming distance. The definition of Hamming distance will be given in this section.

I1MT-R-H. When $\underline{l}_e < 0$ for each $e \in E$ and $c_e > 0$ is the cost of decreasing l_e for each $e \in E$, we call the following program (16) Problem I1MT-R-H, which is the inverse 1-median problem on trees with variable real edge lengths under Hamming distance:

$$\begin{aligned} \min \quad & \sum_{e\in E} c_e H(x_e, l_e) \\ \text{s.t.} \quad & \sum_{e\in P(v_0,\bar{v})} B_e x_e \le 0, \quad \forall \bar{v} \in V \setminus \{v_0\}, \\ & \underline{l}_e \le x_e \le l_e, \quad e \in E, \end{aligned} \tag{16}$$

where the Hamming distance $H(x_e, l_e)$ is defined as

$$H(x_e, l_e) = \begin{cases} 0, & x_e = l_e, \\ 1, & x_e \ne l_e. \end{cases} \tag{17}$$

From a similar discussion in the former Section 3.2.1, Problem I1MT-R-H can be formulated as follows:

$$\begin{aligned} \min \quad & \sum_{e\in E^S} c_e H(x_e, l_e) \\ \text{s.t.} \quad & \sum_{e\in P(v)} B_e x_e \le 0, \quad \forall v \in V^S \setminus \{v_0\}, \\ & \underline{l}_e \le x_e \le l_e, \quad \forall e \in E^S. \end{aligned} \tag{18}$$

Equation (18) is a shortest path improvement problem for tree networks with single source v_0 and the required bound 0 under Hamming distance [19]. In [19], the authors proved that even if the network is a chain network, the problem is still NP-hard. So, we have the following.

Theorem 6. *Problem I1MT-R-H is NP-hard.*

3.2.3. The Model under l_∞ Norm with Variable Real Edge Lengths (I1MT-R-l_∞). We consider the model under l_∞ norm as follows:

I1MT-R-l_∞. When $\underline{l}_e < 0$ for each $e \in E$ and the criterion function $C(a) : R^E \to R$ is a function such that $C(a) = ||ca||_\infty$, where c is the cost vector, $c_e > 0$ is the cost of decreasing l_e by one unit for each $e \in E$, and $\|u\|_\infty = \max_{e\in E}|u_e|$ stands for l_∞ norm of vector u, we call (4) Problem I1MT-R-l_∞, which is the inverse 1-median problem on trees with variable real edge lengths under l_∞ norm.

Also from a similar discussion in Section 3.2.1, Problem I1MT-R-l_∞ can be formally described as follows:

$$\begin{aligned} \min \ & \max_{e\in E^S} c_e (l_e - x_e) \\ \text{s.t.} \ & \sum_{e\in P(v)} B_e x_e \le 0, \quad \forall v \in V^S \setminus \{v_0\}, \\ & \underline{l}_e \le x_e \le l_e, \quad \forall e \in E^S. \end{aligned} \tag{19}$$

Let $t'_e = B_e(x_e - \underline{l}_e)$, $t_e = B_e(l_e - \underline{l}_e)$, $h_e = c_e/\mathrm{B_e}$, $L_y(P(v)) = \sum_{e\in P(v)} y_e$, and $U(v) = -\sum_{e\in P(v)} B_e \underline{l}_e$ for each $e \in E^S$ and $v \in V^S \setminus \{v_0\}$. First, we suppose that $h_e = 1$ for each $e \in E^S$. Equation (19) can be formulated as the following program:

$$\begin{aligned} \min \ & \max_{e\in E^S} \left(t_e - t'_e\right) \\ \text{s.t.} \ & L_{t'}(P(v)) \le U(v), \quad \forall v \in V^S \setminus \{v_0\}, \\ & 0 \le t'_e \le t_e, \quad \forall e \in E^S. \end{aligned} \tag{20}$$

In [20], the authors presented a strongly polynomial time algorithm to solve this type of problem, which runs in $O(|V|^3)$ time. Because the network concerned here is tree network, we now give a simplified procedure to solve (20) which runs in $O(|V|^2)$ time.

On tree T^S, if $\max\{L_t(P(v)) - U(v) \mid v \in V^S \setminus \{v_0\}\} \le 0$, we need do nothing. Hence, we only consider the case that $\max\{L_t(P(v)) - U(v) \mid v \in V^S \setminus \{v_0\}\} > 0$. Clearly, the optimal solution t^* satisfies $\max\{L_{t^*}(P(v)) - U(v) \mid v \in V^S \setminus \{v_0\}\} = 0$. Denote $\theta^* = \|t - t^*\|_\infty$, then for any edge $e \in E^S$, we have $t^*_e = \max\{t_e - \theta^*, 0\}$.

As the procedure in [20], first we sort the edges by their lengths in an increasing order. Beginning with the shortest edge, say e', set $t'_e = t_e - t_{e'}$ for each $e \in E^S$, and check if $\max\{L_{t'}(P(v)) - U(v) \mid v \in V^S \setminus \{v_0\}\} < 0$. If yes, we know that the lengths are overreduced and $\theta^* < t_{e'}$. If $\max\{L_{t'}(P(v)) - U(v) \mid v \in V^S \setminus \{v_0\}\} = 0$, $t^* = t'$ is the optimal solution. If $\max\{L_{t'}(P(v)) - U(v) \mid v \in V^S \setminus \{v_0\}\} > 0$, then $\theta^* > t_{e'}$, and in this case we contract the edges e with $t'_e = 0$ into vertices and repeat the process on the resulted network with respect to t'. Since at least one vertex disappears in each contraction, the procedure has at most $|V|$ iterations. For the time complexity, to find the shortest edge e' can be completed in $O(|E|)$ time, and in each iteration, the time for checking if $\max\{L_{t'}(P(v)) - U(v) \mid v \in V^S \setminus \{v_0\}\} < 0$ is $O(|V|)$. Hence, the complexity of this procedure is $O(|V|^2)$. In the following, we assume that $\theta^* < t_{e'}$, which means that the preliminary procedure is finished while the optimality has not been reached.

Because the path between any two vertices on tree network is unique, let Q denotes the set of vertices which does not satisfy the upper bound demand. Let $|P|$ be the number of edges of path P. Let

$$\theta = \max\left\{\frac{L_t(P(v)) - U(v)}{|P(v)|} \mid v \in Q\right\}, \tag{21}$$

then $\theta^* = \theta$ is the optimal reduction.

Remark 7. Note that when $h_e \neq 1$ for each $e \in E^S$, (20) can be formulated as the following program:

$$\begin{aligned} \min \max_{e \in E^S} \quad & h_e \left(t_e - t'_e\right) \\ \text{s.t.} \quad & L_{t'}\left(P\left(v\right)\right) \leq U\left(v\right), \quad \forall v \in V^S \setminus \{v_0\}, \\ & 0 \leq t'_e \leq t_e, \quad \forall e \in E^S. \end{aligned} \tag{22}$$

In this case, we denote $\theta^* = \|h(t - t^*)\|_\infty$; then for any edge $e \in E^S$, we have $t_e^* = \max\{t_e - \theta^*/h_e, 0\}$. We sort the edges by $t_e h_e$ ($e \in E^S$) in an increasing order to conduct the preliminary procedure to ensure that $\theta^* < t_{e'} h_{e'}$, where e' is the edge satisfying $t_{e'} h_{e'} = \min\{t_e h_e \mid e \in E^S\}$. Let

$$\theta = \max\left\{\frac{L_t\left(P\left(v\right)\right) - U\left(v\right)}{\sum_{e \in P(v)} 1/h_e} \mid v \in Q\right\}, \tag{23}$$

then $\theta^* = \theta$ is the optimal reduction. Similar discussions show that the above procedure correctly solves program (22), and its computational complexity is $O(|V|^2)$.

From the above analysis, we have the following theorem.

Theorem 8. *Problem I1MT-R-l_∞ can be solved in $O(|V|^2)$ time.*

4. Conclusion

In this paper, we discussed the inverse 1-median problem with variable edge lengths on tree networks. We showed that this problem is solvable in linear time with variable nonnegative edge lengths under all kinds of norms. For the case when negative edge lengths are allowable, by formulating this problem to some known combinatorial optimization problems, we showed that this problem is NP-hard under Hamming distance and we proposed strongly polynomial time algorithms under l_1 and l_∞ norms, respectively.

As a future research topic, it will be meaningful to consider the problem on other networks such as the interval graph, the cactus graph, and the block graph, especially under the Hamming distance. We know that Algorithm 2 can solve Problem I1MT-N under all kinds of norms. But when negative edge lengths are allowable, we only dealt with the problems under the constraint that the edge length increase is forbidden. So we can further study the complexity of the problems such that the edge lengths can either be reduced or increased. In addition, for the inverse p-median problems for $p \geq 2$, designing the optimal or approximation algorithms and analysing their computational complexities are also promising.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (NSFC nos. 11171316 and 61173002), the Grant of the Second Stage of Brain Korea 21, and Zhejiang Provincial Natural Science Foundation of China (no. Y6090472).

References

[1] O. Kariv and S. L. Hakimi, "An algorithmic approach to network location problems. II. The p-medians," *SIAM Journal on Applied Mathematics*, vol. 37, no. 3, pp. 539–560, 1979.

[2] L. K. Hua, "Applications of mathematical models to wheat harvesting," *Chinese Mathematics*, no. 2, pp. 77–91, 1962.

[3] A. J. Goldman, "Optimal center location in simple networks," *Transportation Science*, no. 5, pp. 212–221, 1971.

[4] H. Shi, W. Wang, N. M. Kwok, and S. Y. Chen, "Game theory for wireless sensor networks: a survey," *Sensors*, vol. 12, no. 7, pp. 9055–9097, 2012.

[5] O. Berman and J. Wang, "The network p-median problem with discrete probabilistic demand weights," *Computers & Operations Research*, vol. 37, no. 8, pp. 1455–1463, 2010.

[6] M. Chrobak, C. Kenyon, J. Noga, and N. E. Young, "Incremental medians via online bidding," *Algorithmica*, vol. 50, no. 4, pp. 455–478, 2008.

[7] R. Tadei, N. Ricciardi, and G. Perboli, "The stochastic p-median problem with unknown cost probability distribution," *Operations Research Letters*, vol. 37, no. 2, pp. 135–141, 2009.

[8] S. Chen, Y. Zheng, C. Cattani, and W. Wang, "Modeling of biological intelligence for SCM system optimization," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 769702, 10 pages, 2012.

[9] R. E. Burkard, C. Pleschiutschnig, and J. Zhang, "Inverse median problems," *Discrete Optimization*, vol. 1, no. 1, pp. 23–39, 2004.

[10] R. E. Burkard, C. Pleschiutschnig, and J. Z. Zhang, "The inverse 1-median problem on a cycle," *Discrete Optimization*, vol. 5, no. 2, pp. 242–253, 2008.

[11] R. E. Burkard, M. Galavii, and E. Gassner, "The inverse Fermat-Weber problem," Technical Report 2008-14, Graz University of Technology, Graz, Austria, 2008.

[12] M. Galavii, "The inverse 1-median problem on a tree and on a path," *Electronic Notes in Discrete Mathematics*, no. 36, pp. 1241–1248, 2010.

[13] X. C. Guan and B. W. Zhang, "Inverse 1-median problem on trees under weighted Hamming distance," *Journal of Global Optimization*, vol. 54, no. 1, pp. 75–82, 2012.

[14] F. B. Bonab, R. E. Burkard, and E. Gassner, "Inverse p-median problems with variable edge lengths," *Mathematical Methods of Operations Research*, vol. 73, no. 2, pp. 263–280, 2011.

[15] R. Diestel, *Graph Theory*, Springer, Heidelberg, Germany, 3rd edition, 2005.

[16] J. Z. Zhang and Y. X. Lin, "Computation of the reverse shortest-path problem," *Journal of Global Optimization*, vol. 25, no. 3, pp. 243–261, 2003.

[17] J. Z. Zhang, X. G. Yang, and M. C. Cai, "Inapproximability and a polynomially solvable special case of a network improvement problem," *European Journal of Operational Research*, vol. 155, no. 1, pp. 251–257, 2004.

[18] E. Tardos, "A strongly polynomial algorithm to solve combinatorial linear programs," *Operations Research*, vol. 34, no. 2, pp. 250–256, 1986.

[19] B. W. Zhang, J. Z. Zhang, and L. Q. Qi, "The shortest path improvement problems under Hamming distance," *Journal of Combinatorial Optimization*, vol. 12, no. 4, pp. 351–361, 2006.

[20] J. Z. Zhang, X. G. Yang, and M. C. Cai, "A network improvement problem under different norms," *Computational Optimization and Applications*, vol. 27, no. 3, pp. 305–319, 2004.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 217213, 8 pages
http://dx.doi.org/10.1155/2013/217213

Research Article

Stationarity Testing of Accumulated Ethernet Traffic

Zhiping Lu,[1] Ming Li,[2] and Wei Zhao[3]

[1] *School of Finance and Statistics, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China*
[2] *School of Information Science and Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China*
[3] *Department of Computer and Information Science, University of Macau, Av. Padre Tomas Pereira, Taipa, Macau*

Correspondence should be addressed to Zhiping Lu; camping9@gmail.com

Received 26 January 2013; Revised 12 March 2013; Accepted 13 March 2013

Academic Editor: Shengyong Chen

We investigate the stationarity property of the accumulated Ethernet traffic series. We applied several widely used stationarity and unit root tests, such as Dickey-Fuller test and its augmented version, Phillips-Perron test, as well as the Kwiatkowski-Phillips-Schmidt-Shin test and some of its generalizations, to the assessment of the stationarity of the traffic traces at the different time scales. The quantitative results in this research provide evidence that when the time scale increases, the accumulated traffic series are more stationary.

1. Introduction

Stationarity testing is essential for time series. According to the meaning of nonstationarity discussed in [1, 2], investigating time-varying spectra turns to be a natural way of testing nonstationarity of time series [3–7]. However, one may encounter difficulties in nonstationarity testing of long-range dependent (LRD) network traffic (traffic for short) because the LRD property implies $1/f$ noise, which is divergent at the zero frequency [8]. In addition to time-frequency distributions, there are many approaches in this regard, such as bootstrap testing [9] and so on [10–12]. However, conventional approaches may not be properly used for testing the nonstationarity of LRD traffic as can be seen from Abry and Veitch [13], Grossglauser and Bolot [14]. That difficulty may also refer to a paper by Mandelbrot in 1976 [15]. Abry and Veitch [13] proposed a test method for LRD traffic by investigating the time invariability of the Hurst parameter, H. However, in terms of engineering, time-varying H may not always imply nonstationarity but multifractal in general, see [16, 17] for details. Therefore, further research of nonstationarity of LRD traffic is desired.

The paper aims at providing the following contributions.

(i) Different scales are taken into account for investigating the stationarity of the accumulated traffic data through unit root tests.

(ii) The quantitative description of the large scales for the Ethernet traffic to be stationary is given.

The remainder of this paper is organized as follows. In Section 2, the dataset is described. In Section 3, several widely used unit root tests and stationarity tests are presented. In Section 4, the numerical results of the statistical tests are included and the discussion is followed. Finally, Section 5 presents some concluding remarks.

2. Datasets

This research utilizes four real Ethernet traffic traces, listed in Table 1, which were measured on an Ethernet at the Bellcore Morristown Research and Engineering facility [18] in 1989. We may refer to [19] for more details about the datasets.

Denote by $x(t(i))$ a traffic time series, where $t(i)$ ($i = 0, 1, \ldots$) is the time-stamp series, which indicates the time stamp of the ith packet. Note that $x(t(i))$ represents the packet size of the ith packet at time $t(i)$. This research uses $x(i)$ to represent the packet size of the ith packet on a packet-by-packet basis. Further, on an interval-by-interval basis, we consider the accumulated traffic, denoted by $y(n)$. It is given by

$$y(n) = \sum_{i=nT}^{(n+1)T} x(i), \tag{1}$$

Table 1: Four Ethernet traffic traces.

Series name	Starting time	Duration	Series length
pAug.TL	11:25 AM, 29 Aug 89	52 minutes	1 million
pOct.TL	11:00 AM, 05 Oct 89	29 minutes	1 million
OctExt.TL	11:46 PM, 03 Oct 89	34.111 h	1 million
OctExt4.TL	2:37 PM, 10 Oct 89	21.095 h	1 million

where T is the interval width. It is in fact the accumulation scale. Without causing confusion, we still call it time scale (or scale for short) throughout the paper. Thus, $y(n)$ stands for the accumulated bytes of arrival traffic in the nth interval with the time scale T. We shall exhibit an interesting phenomenon that the stationarity of $y(n)$ of the Ethernet traffic traces considerably relies on the time scale T. More precisely, qualitatively speaking, $y(n)$ is nonstationary if T is small while it is stationary when T is large. The key point of this research is to exhibit the quantitative descriptions of T at which $y(n)$ changes from the non-stationary case to the stationary one.

3. Statistical Tools

Computer scientists concern about the stationarity property of traffic, see for example, [13]. Nevertheless, the consensus of that property of traffic may not be achieved due to different attributes of traffic. For instance, our research investigates the stationarity property of traffic at different time scales using the attribute of accumulation scale, which is rarely reported, to the best of our knowledge. For that reason, we explain our results in what follows.

On one hand, from Figures 1, 2, 3, and 4, we could observe the trajectories of the accumulated $y(n)$ under different time scales. Visually, it seems that there exists some nonstationarity in all the four series, especially under the large time scales. However, we need some more objective methods to verify the reliability of this opinion. On the other hand, we present in Figures 5, 6, 7, and 8 the spectrum of the accumulated $y(n)$ under different time scales, which clearly indicates the existence of long range dependence in the series.

It is important to learn about the statistical property of the data before quantitatively using the accurate models. In statistics, one of the most investigated branches is that of unit root and stationarity testing. Suppose a discrete time stochastic process can be written as an autoregressive process. If 1 is a root of the characteristic equation, the stochastic process is said to have a unit root or, alternatively, is integrated of order one, denoted $I(1)$ and is non-stationary. If the other roots of the characteristic equation lie inside the unit circle—that is, have a modulus (absolute value) less than one—then the first difference of the process will be stationary. This is the original idea of the unit root testing proposed in 1979, and hereafter many extensions and modifications have been developed. In the following we concentrate on the existing stationarity testing as well as some of their variants and generalizations to identify the existence of unit root in the series.

The standard Dickey-Fuller test (DF) [20] is based on i.i.d. errors and has as the null hypothesis the unit root. The DF test is valid if the time series is well characterized by an Auto-Regressive (AR(1)) process with white noise errors. Many time series, however, have a more complicated dynamic structure than is captured by a simple AR(1) model. Said and Dickey (1984) [21] augment the basic autoregressive unit root test to accommodate Auto-Regressive and Moving-Average (ARMA(p, q)) models with unknown orders and their test is referred to as the augmented Dickey-Fuller (ADF) test. On the other hand, the Phillips-Perron test (PP) [22] is non-parametric and allows for some heterogeneity and serial correlation in the innovations, which is different from the ADF tests. Some class of so-called efficient unit root tests was proposed by Elliott et al. (1996) (hereafter ERS) [23] whose test statistics come very close to the power envelope for a wide range of alternatives, and they can have substantially higher power than the ADF or PP unit root tests. Unlike AR(MA) unit root tests, stationarity tests are far less numerous and have as null hypothesis the stationarity assumption and as alternative the unit root. Among the most well known stationarity tests, the most extensively used one is the Kwiatkowski et al. (1992) test, also known as KPSS [24], which is intended to complement unit root tests and can be used to distinguish short memory and long memory stationary processes.

There exist many other unit root and stationarity tests as well as generalizations and combinations of the ones mentioned above. For supplementary material on unit root and stationarity testing, see [10, 25–29], and so forth, for more details.

However, one fundamental problem is that there is still no consensus on the optimal choice of the stationarity test. Therefore, to be more objective, we take advantage of several common tests to make a comprehensive study of the Ethernet traffic data in the following empirical study.

4. Discussion

In this section, we present the statistical testing results of the four series, that is, pAug.TL, pOct.TL, OctExt.TL, and OctExt4.TL data. Our interest lies in the identification of the existence of the unit root in the series. We finally adopt the most widely used ADF, PP, ERS, and KPSS tests in the empirical application. Based on the software R, we mainly utilize the functions of the packages "urca" and "fUnitRoots" to realize the statistical tests. The null hypothesis is that the series is $I(0)$ stationary, while the alternative is that it is $I(1)$. The P value of each test under the time scales $T = 2^n$, $n = 9, \ldots, 17$ are presented in Tables 2, 3, 4, and 5. It is worth mentioning that we have also tried to utilize the Robinson (1994) test [30] to test the stationarity of the series due to their identified long range dependence features. However, according to the result of Ferrara et al. (2010) [31], this test is appropriate to use only for series whose sample size is at least 3000, which could never be satisfied in our case.

As seen in Tables 2–5 as time scale increases, the accumulated traffic series $y(n)$ is more likely to be stationary,

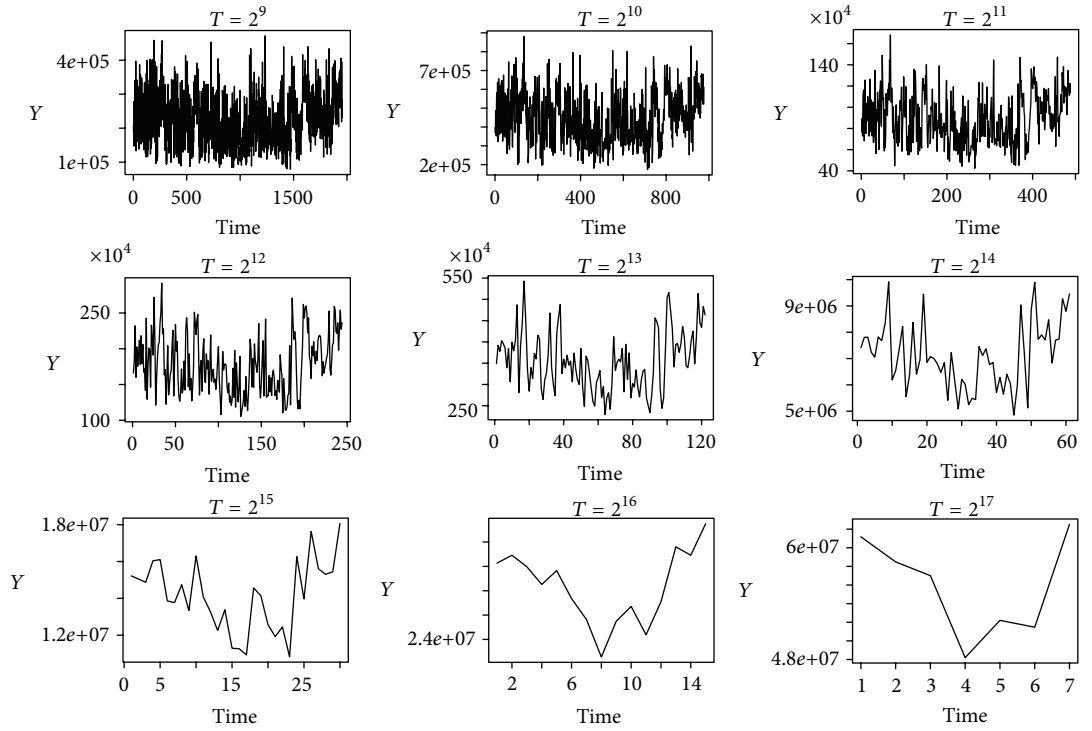


Figure 1: Trajectory of the accumulated pAug.TL data under different time scales.

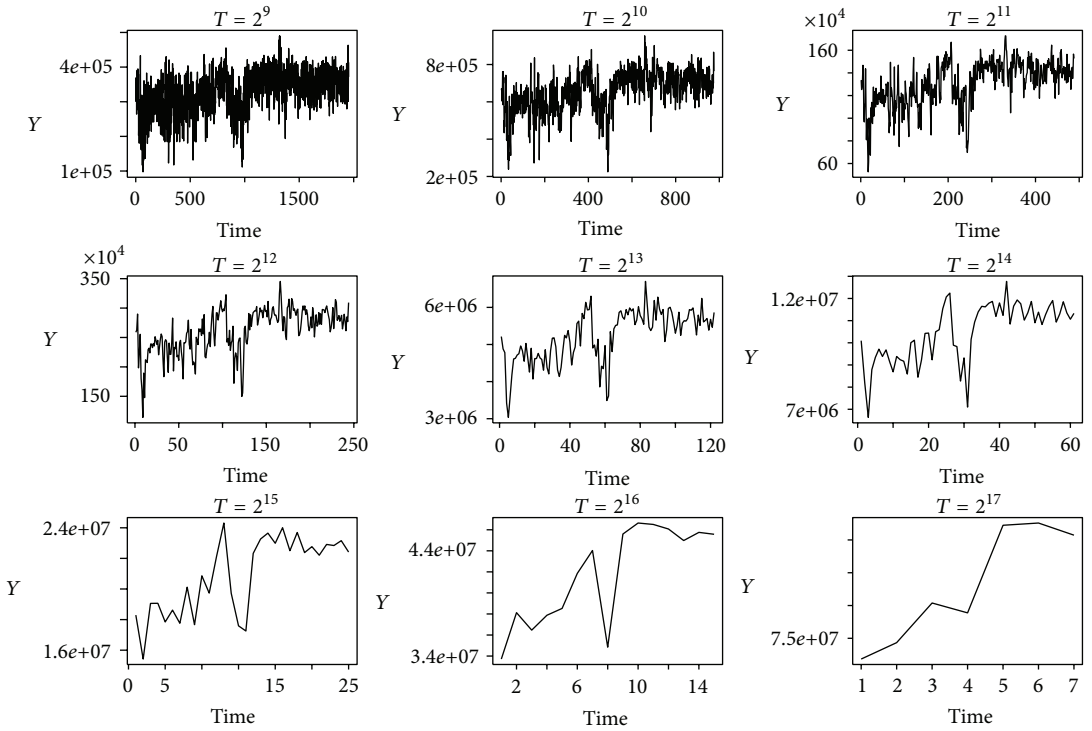


Figure 2: Trajectory of the accumulated pOct.TL data under different time scales.

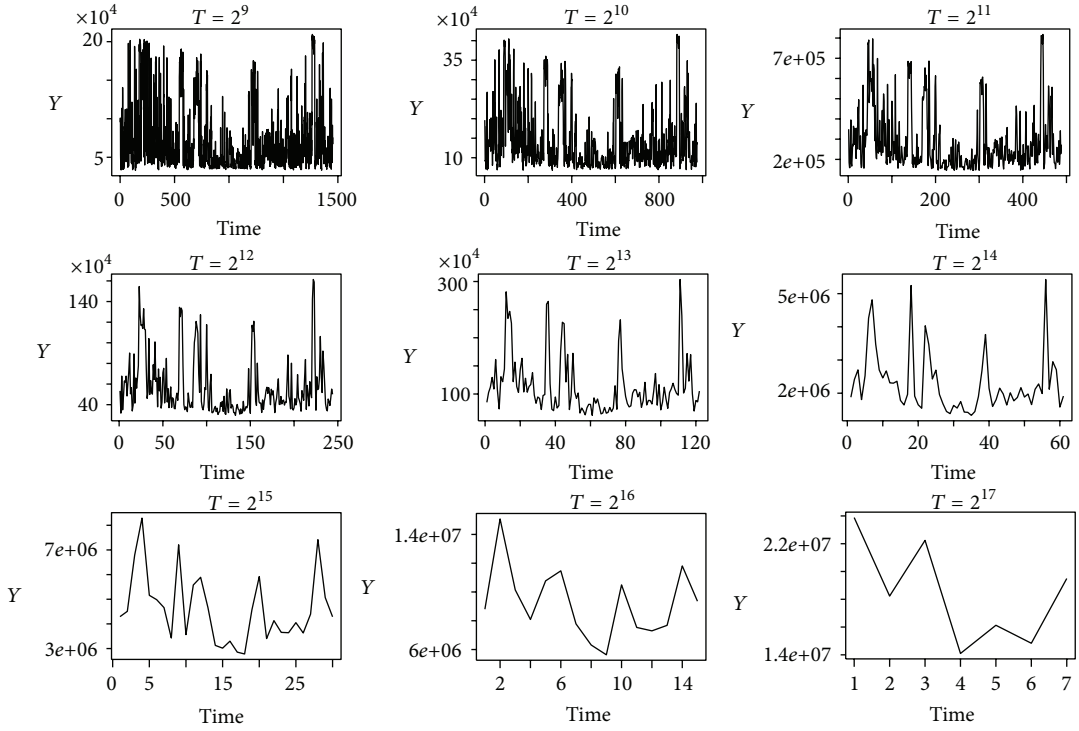


Figure 3: Trajectory of the accumulated Octxt.TL data under different time scales.

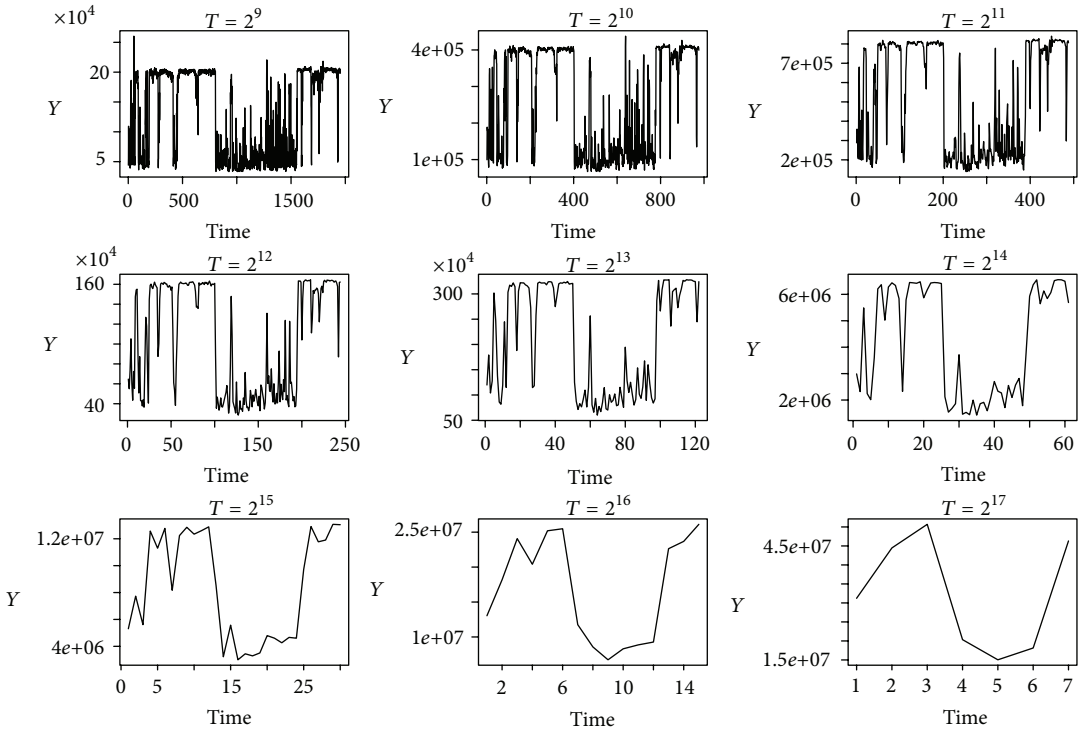


Figure 4: Trajectory of the accumulated OctExt4.TL data under different time scales.

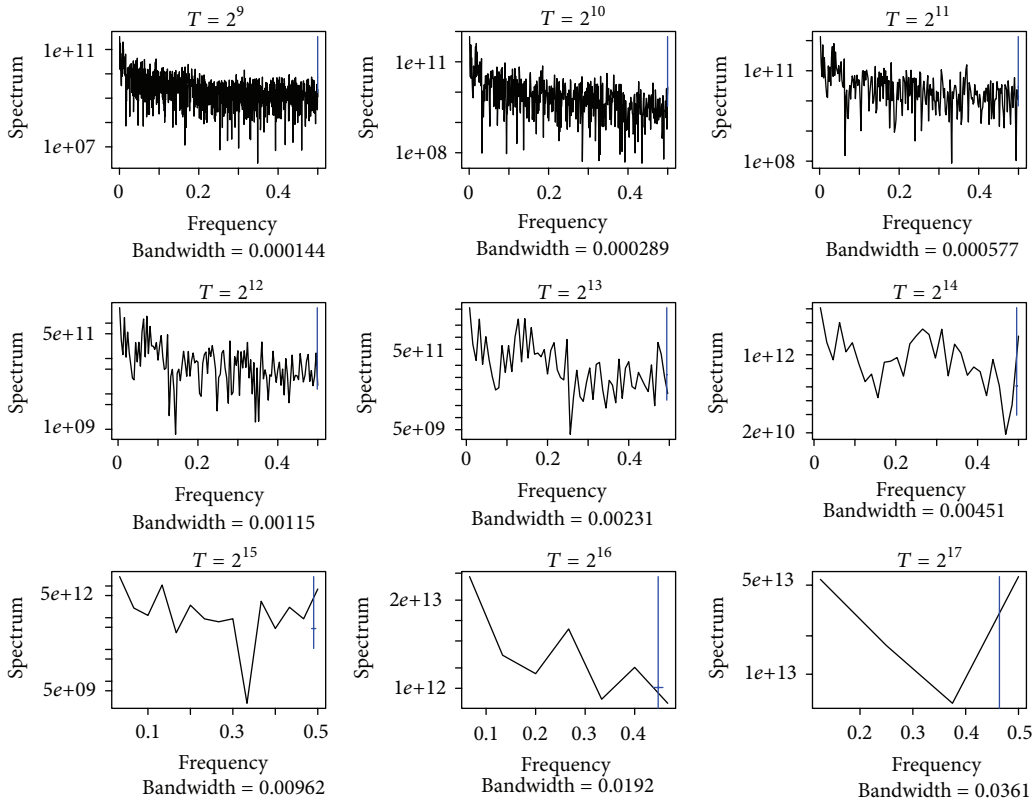


Figure 5: Spectrum of the accumulated pAug.TL data under different time scales.

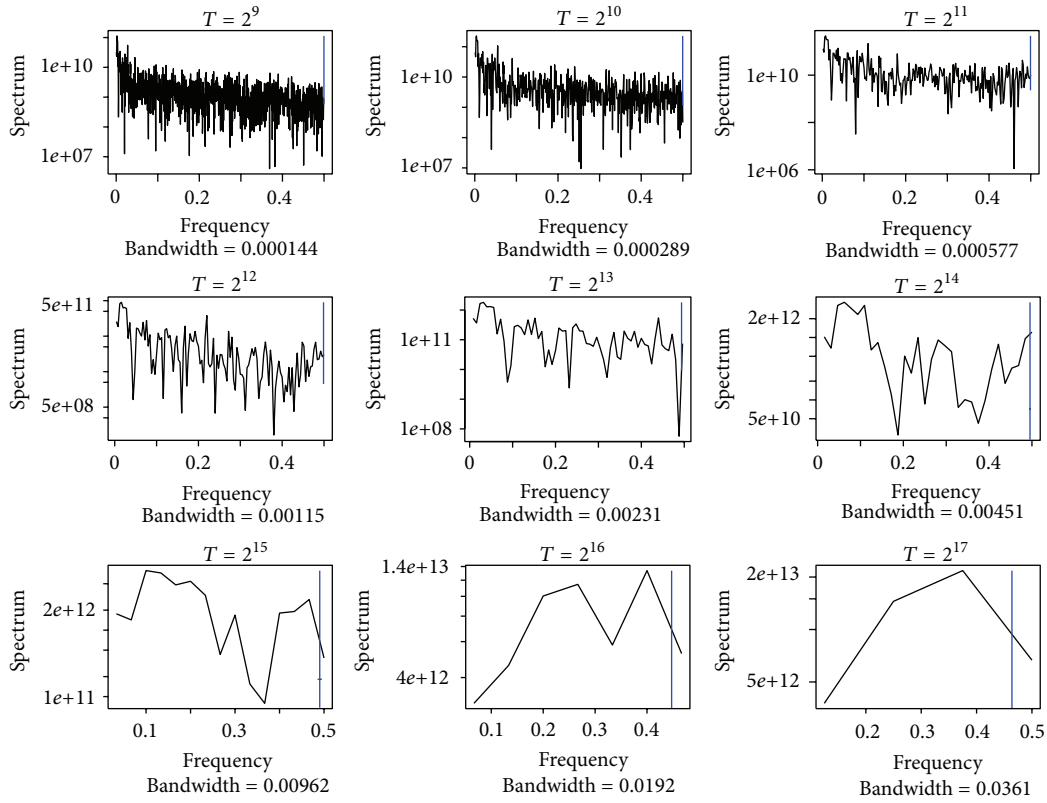


Figure 6: Spectrum of the accumulated pOct.TL data under different time scales.

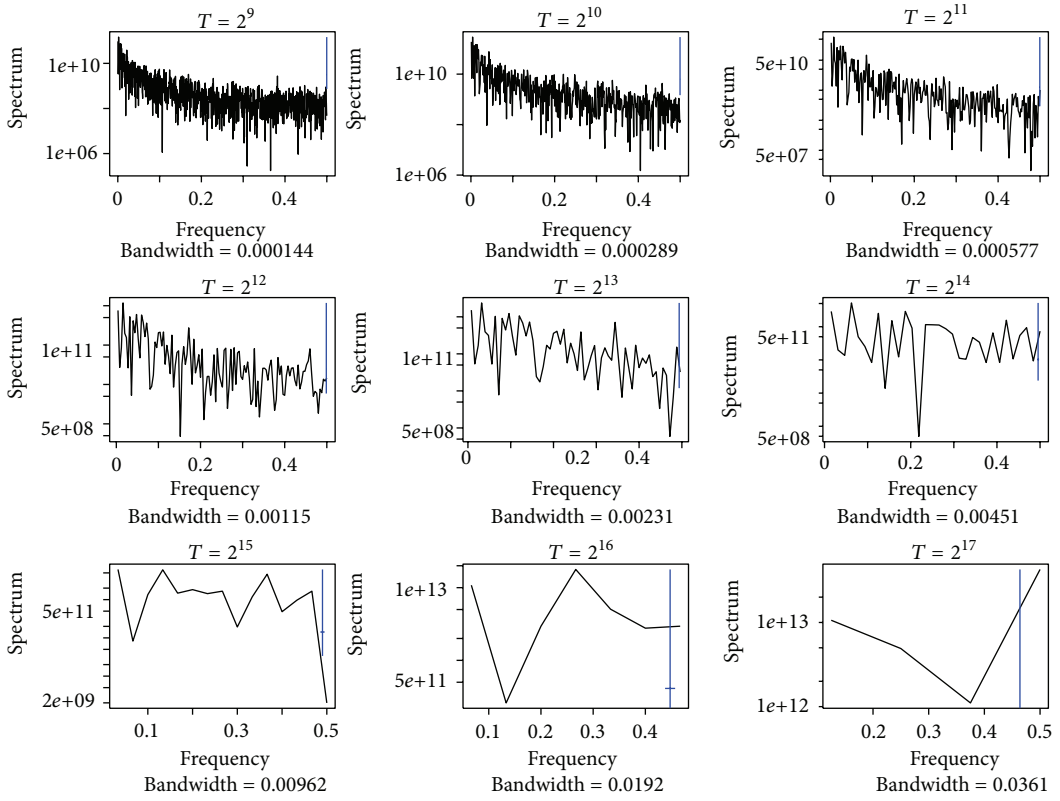


Figure 7: Spectrum of the accumulated Octxt.TL data under different time scales.

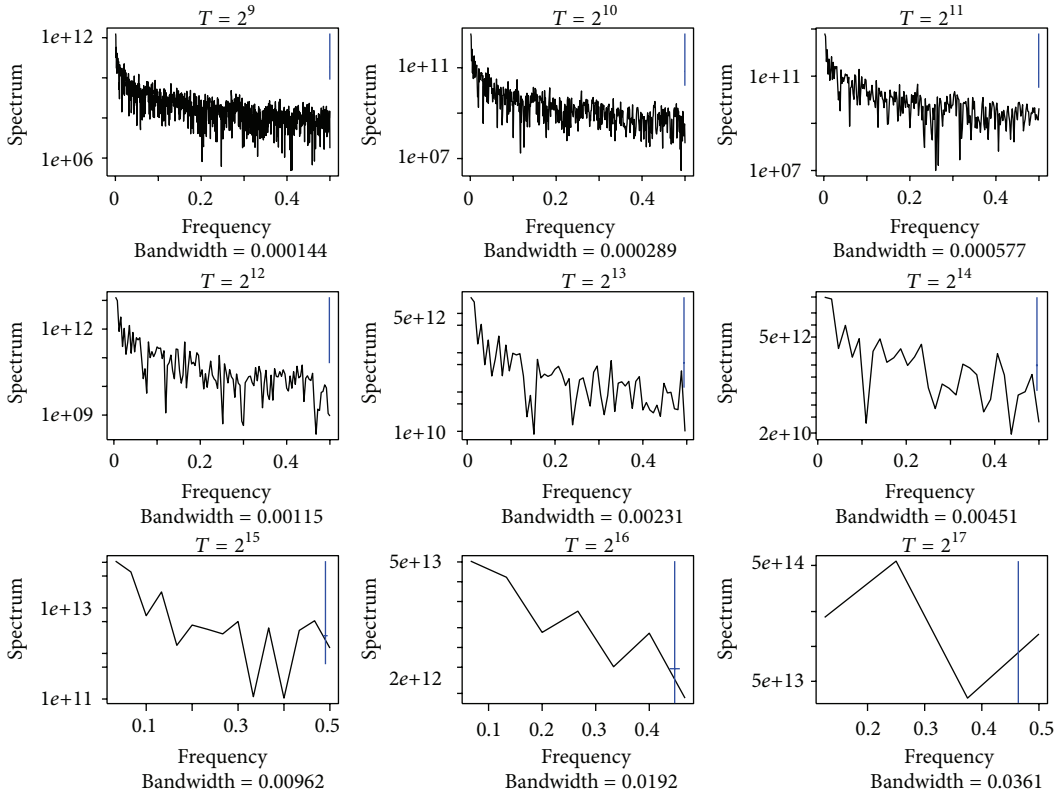


Figure 8: Spectrum of the accumulated OctExt4.TL data under different time scales.

Table 2: Test results for pAug.TL series.

Test	KPSS	ADF	PP	ERS
$T = 512$	<0.01	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 1024$	<0.01	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 2048$	0.01531	$6.622e-14$	$<2.2e-16$	$<2.2e-16$
$T = 4096$	0.02589	$8.969e-08$	$<2.2e-16$	$1.736e-13$
$T = 8192$	0.05159	0.1222	$1.076e-11$	$6.255e-07$
$T = 16384$	0.07265	0.03636	0.003181	$1.563e-05$
$T = 32768$	>0.1	0.02883	0.0157	0.1148
$T = 65536$	>0.1	0.9794	0.007272	0.3533
$T = 131072$	>0.1	0.9433	0.8024	NA

Table 3: Test results for pOct.TL series.

Test	KPSS	ADF	PP	ERS
$T = 512$	<0.01	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 1024$	<0.01	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 2048$	<0.01	$7.589e-15$	$<2.2e-16$	$<2.2e-16$
$T = 4096$	<0.01	0.0002329	$<2.2e-16$	$2.749e-09$
$T = 8192$	<0.01	0.005249	$<2.2e-16$	$1.107e-05$
$T = 16384$	<0.01	0.07512	$1.609e-10$	0.01228
$T = 32768$	<0.01	0.294	$8.265e-05$	0.1865
$T = 65536$	<0.01	0.197	0.03797	0.7795
$T = 131072$	0.01835	0.4773	0.07518	NA

Table 4: Test results for OctExt.TL series.

Test	KPSS	ADF	PP	ERS
$T = 512$	<0.01	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 1024$	0.01112	$3.262e-11$	$<2.2e-16$	$<2.2e-16$
$T = 2048$	0.02846	$3.615e-05$	$<2.2e-16$	$3.718e-13$
$T = 4096$	0.0487	0.000314	$<2.2e-16$	$1.216e-08$
$T = 8192$	0.08652	0.04858	$5.691e-11$	$1.65e-05$
$T = 16384$	>0.1	0.01177	0.01622	0.0006824
$T = 32768$	>0.1	0.1303	0.1326	0.0247
$T = 65536$	>0.1	0.2777	0.6618	0.1359
$T = 131072$	>0.1	0.2846	0.8794	NA

Table 5: Test results for OctExt4.TL series.

Test	KPSS	ADF	PP	ERS
$T = 512$	<0.01	0.0001743	$<2.2e-16$	$<2.2e-16$
$T = 1024$	<0.01	0.001538	$<2.2e-16$	$6.507e-11$
$T = 2048$	<0.01	0.05514	$<2.2e-16$	$3.832e-05$
$T = 4096$	0.01045	0.1265	$<2.2e-16$	0.000242
$T = 8192$	0.03833	0.1103	$<2.2e-16$	0.04042
$T = 16384$	0.07791	0.3414	$1.6e-14$	0.4146
$T = 32768$	>0.1	0.5884	$1.004e-06$	0.3668
$T = 65536$	>0.1	0.885	0.008174	0.314
$T = 131072$	>0.1	0.8827	0.6979	NA

which is on the contrast with the intuitive observation from the trajectory figures. Moreover, given the significance level at 1%, when the time scale is small, the accumulated traffic could be judged as non-stationary and possesses a unit root. Specifically, we have the following findings.

(i) The accumulated traffic of pOct.TL series could be regarded as stationary until the time scale reaches 17, while the other three series cannot be rejected to be stationary once the time scale is greater than 15. Thus, among the four series, the pOct.TL series seems to possess the least possibility to be stationary under each time scale.

(ii) Considering the series of pAug.TL, pOct.TL, OctExt.TL, and OctExt4.TL, it seems that their stationarity behaviors are quite similar and hard to distinguish from each other under the corresponding time scales.

(iii) For all these four series, when the time scale is small, the accumulated traffic has a unit root and exhibits at the same time some long range dependence behavior, which indicates after first order difference we could obtain a stationary process.

The previous discussions are for the Ethernet traffic but the methods may also be a reference for other types of time series, for instance, those in [32–35]. Our future research will work on the general description of network traffic rather than the Ethernet one alone.

5. Conclusions

In this paper, we have carried out several widely used tests of stationarity in order to study the stationary property of the accumulated Ethernet series at different time scales. The quantitative results reveal that when the time scale increases, the investigated accumulated Ethernet traffic is more likely to be stationary, which coincides with the normality investigation results for the same series and provide a useful empirical evidence for the traffic data modeling under large time scale.

Acknowledgments

This work was in part supported by the 973 Plan under Project Grant no. 2011CB302800, the National Natural Science Foundation of China under Project Grant nos. 11101158, 61272402, 61070214, and 60873264, and "the Fundamental Research Funds for the Central Universities". The authors acknowledge W. Willinger, W. Leland, and D. Wilson with Bellcore, Morristown, who provided the data in this research.

References

[1] R. M. Loynes, "On the concept of the spectrum for non-stationary processes," *Journal of the Royal Statistical Society Series B*, vol. 30, pp. 1–30, 1968.

[2] M. B. Priestley, "Evolutionary spectra and non-stationary processes," *Journal of the Royal Statistical Society Series B*, vol. 27, pp. 204–237, 1965.

[3] Z. Guo, L. G. Durand, and H. C. Lee, "Time-frequency distributions of nonstationary signals based on a Bessel kernel," *IEEE*

Transactions on Signal Processing, vol. 42, no. 7, pp. 1700–1706, 1994.

[4] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, no. 1971, pp. 903–995, 1998.

[5] A. R. Messina and V. Vittal, "Nonlinear, non-stationary analysis of interarea oscillations via Hilbert spectral analysis," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1234–1241, 2006.

[6] C. R. Pinnegar and L. Mansinha, "Time-local spectral analysis for nonstationary time series: the s-transform for noisy signals," *Fluctuation and Noise Letters*, vol. 3, no. 3, pp. L357–L364, 2003.

[7] J. Xiao and P. Flandrin, "Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2851–2860, 2007.

[8] M. Li and W. Zhao, "On 1/f noise," *Mathematical Problems in Engineering*, vol. 2012, Article ID 673648, 23 pages, 2012.

[9] Z. Psaradakis, "Blockwise bootstrap testing for stationarity," *Statistics and Probability Letters*, vol. 76, no. 6, pp. 562–570, 2006.

[10] S. Ling, "Estimation and testing stationarity for double-autoregressive models," *Journal of the Royal Statistical Society Series B*, vol. 66, no. 1, pp. 63–78, 2004.

[11] P. M. M. Rodrigues and A. Rubia, "A note on testing for non-stationarity in autoregressive processes with level dependent conditional heteroskedasticity," *Statistical Papers*, vol. 49, no. 3, pp. 581–593, 2008.

[12] R. Von Sachs and M. H. Neumann, "A wavelet-based test for stationarity," *Journal of Time Series Analysis*, vol. 21, no. 5, pp. 597–613, 2000.

[13] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic patrice abry and darryl veitch," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 2–15, 1998.

[14] M. Grossglauser and J. C. Bolot, "On the relevance of long-range dependence in network traffic," *IEEE/ACM Transactions on Networking*, vol. 7, no. 5, pp. 629–640, 1999.

[15] B. B. Mandelbrot, "Note on the definition and the stationarity of fractional Gaussian noise," *Journal of Hydrology*, vol. 30, no. 4, pp. 407–409, 1976.

[16] M. Li and W. Zhao, "Quantitatively investigating locally weak stationarity of modified multifractional Gaussian noise," *Physica A*, vol. 391, no. 24, pp. 6268–6278, 2012.

[17] B. B. Mandelbrot, *Multifractals and 1/f Noise*, Springer, New York, NY, USA, 1999.

[18] http://ita.ee.lbl.gov/html/contrib/BC.html.

[19] Z. Lu, M. Li, and W. Zhao, "Normality of Ethernet traffic at large time scales," *Mathematical Problems in Engineering*, vol. 2013, Article ID 471963, 2013.

[20] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 427–431, 1979.

[21] S. E. Said and D. A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, vol. 71, no. 3, pp. 599–607, 1984.

[22] P. C. B. Phillips and P. Perron, "Testing for a unit root in time series regression," *Biometrika*, vol. 75, no. 2, pp. 335–346, 1988.

[23] G. Elliott, T. J. Rothenberg, and J. H. Stock, "Efficient tests for an autoregressive unit root," *Econometrica*, vol. 64, no. 4, pp. 813–836, 1996.

[24] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root?" *Journal of Econometrics*, vol. 54, no. 1–3, pp. 159–178, 1992.

[25] P. J. Brockwell and R. A. Davis, "Nonstationary and seasonal time series models," in *Introduction to Time Series and Forecasting*, pp. 179–196, Springer, New York, NY, USA, 2002.

[26] C. B. Foster, *Economics and Business Forecasting—Lectures 4, 5 and 6*, Lecture notes-Ec 178, University of California, San Diego, Calif, USA, 2005.

[27] M. Lanne and P. Saikkonen, "Reducing size distortions of parametric stationarity tests," *Journal of Time Series Analysis*, vol. 24, no. 4, pp. 423–439, 2003.

[28] R. Refinetti, "Non-stationary time series and the robustness of circadian rhythms," *Journal of Theoretical Biology*, vol. 227, no. 4, pp. 571–581, 2004.

[29] E. Zivot, "Economic Theory II: Time Series Analysis—Unit Root Tests," Lecture Notes-Econ 584. University of Washington, 2005, http://faculty.washington.edu/ezivot/econ584/notes/unitroot.pdf.

[30] P. M. Robinson, "Efficient tests of nonstationary hypotheses," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1420–1437, 1994.

[31] L. Ferrara, D. Guegan, and Z. Lu, "Testing fractional order of long memory processes: a monte carlo study," *Communications in Statistics*, vol. 39, no. 4, pp. 795–806, 2010.

[32] E. G. Bakhoum and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.

[33] C. Cattani, "On the existence of wavelet symmetries in archaea DNA," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 673934, 21 pages, 2012.

[34] C. Cattani, G. Pierro, and G. Altieri, "Entropy and multifractality for the myeloma multiple TET 2 gene," *Mathematical Problems in Engineering*, vol. 2012, Article ID 193761, 14 pages, 2012.

[35] C. Toma, "Advanced signal processing and command synthesis for memory-limited complex systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 927821, 13 pages, 2012.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 324945, 9 pages
http://dx.doi.org/10.1155/2013/324945

Research Article

Adaptive Multilevel Kernel Machine for Scene Classification

Junlin Hu,[1] Liang Wang,[2] Fuqing Duan,[3] and Ping Guo[1]

[1] *Image Processing and Pattern Recognition Laboratory, Beijing Normal University, Beijing 100875, China*
[2] *College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China*
[3] *College of Information Science and Technology, Beijing Normal University, Beijing 100875, China*

Correspondence should be addressed to Fuqing Duan; fqduan@bnu.edu.cn and Ping Guo; pguo@ieee.org

Received 23 January 2013; Accepted 24 February 2013

Academic Editor: Carlo Cattani

Scene classification is a challenging problem in computer vision applications and can be used to model and analyze a special complex system, the internet community. The spatial PACT (Principal component Analysis of Census Transform histograms) is a promising representation for recognizing instances and categories of scenes. However, since the original spatial PACT only simply concatenates compact census transform histograms at all levels together, all levels have the same contribution, which ignores the difference among various levels. In order to ameliorate this point, we propose an adaptive multilevel kernel machine method for scene classification. Firstly, it computes a set of basic kernels at each level. Secondly, an effective adaptive weight learning scheme is employed to find the optimal weights for best fusing all these base kernels. Finally, support vector machine with the optimal kernel is used for scene classification. Experiments on two popular benchmark datasets demonstrate that the proposed adaptive multilevel kernel machine method outperforms the original spatial PACT. Moreover, the proposed method is simple and easy to implement.

1. Introduction

With the rapid development of the smart phone and internet, millions of users can now share their photos or videos taken by these smart phones on the internet and then construct a special complexity system, an internet community. On one hand, these photos and videos of each internet community can be directly used to model the real world by 3D reconstruction, scene retrieval, navigation, and augmented reality [1–7]. On the other hand, they can provide information to model and analyze the complexity system by scene classification and face recognition [4–6]. Great advances have been made on the face recognition, which is widely used in real applications. However, scene classification still is a great challenging problem due to the clutter and the occlusion. Related techniques for associating scenes with semantic labels have a high potential for improving the performance of other computer vision applications such as image browsing, retrieval, and categorization [2, 4, 8]. Motivated by these applications, many image representation models have been proposed for addressing this problem, such as probabilistic latent semantic analysis model [9], part-based model [10], and BoW (bag of visual words) model [3, 6]. Among these existing models, the BoW model has shown an excellent performance and been widely used in many real applications due to its robustness to scale, translation, and rotation transform, including image classification, image annotation [11, 12].

Generally, the BoW model treats an image as a collection of unordered appearance descriptors extracted from local image patches, quantizes them into discrete visual words, and computes a compact histogram representation for semantic image classification. However, the BoW method discards the spatial information of local descriptors, which makes the representation power of the descriptors severely limited. To include information about the spatial layout of the descriptors, Lazebnik et al. [13] propose an extension of BoW model, named as SPM (spatial pyramid matching), which has achieved excellent results on several challenging image classification tasks. For each image, spatial pyramid partitions an image into increasingly finer patches at different levels, and a pyramid match kernel is used to estimate the overall perceptual similarity between images. In their work, the k-means technique is used to generate the codebook, and each local descriptor is only quantized to its nearest

visual word. However, such a hard assignment may result in severe quantization error. To reduce quantization error, Yang et al. [14] propose an improved version of SPM called ScSPM (sparse coding SPM), which computes a spatial pyramid image representation based on sparse codes of the SIFT (scale invariant feature transform) [15]. In this way, the ScSPM can automatically learn the optimal codebook by replacing k-means with sparse coding, and concurrently search for the optimal weights to be assigned to the visual words for each local descriptor. By adopting sparse coding to generate the codebook, their model achieves an excellent performance in image classification [14, 16].

Although the BoW approach has achieved good results for scene classification, the most important procedures in this method, codebook generation and feature quantization, have a high complexity both in time and space. Recently, Wu and Rehg [17] introduce PACT, an effective representation that fulfills the need for recognizing both instances and categories of places and scenes. The CT (Census Transform) encodes local shape information with significant relevance between neighboring CT values, while PCA (Principal Component Analysis) can extract the global shape information in an image patch to represent image compactly. To capture spatial information, they also propose spatial PACT scheme, spatial pyramid representation of PACT, which encodes roughly global spatial arrangement of subblocks in an image and tries to seek the tradeoff between discriminative power and invariance for scene recognition tasks. Furthermore, they conclude that the spatial PACT has several important advantages in comparison to previous state-of-the-art feature representations for scene classification, such as superior recognition performance, no parameter to tune, and extremely fast evaluation speed.

The spatial PACT, however, only concatenates compact histograms at all levels together and ignores the difference among different levels. In this paper, to alleviate the impact of this problem, we propose an adaptive multilevel kernel machine approach, which firstly computes a set of base kernels corresponding to each level in the pyramid of PACT, then a promising adaptive weight learning scheme is used to seek a set of optimal weights for best fusing all base kernels for scene categorization. Lastly, support vector machine (SVM) classifier [18] with the optimal kernel is employed to perform the scene classification task.

The remainder of this paper is organized as follows. Section 2 briefly reviews some related works, mainly including Census Transform, CENTRIST (CENsus TRansform hISTogram), PACT, and spatial PACT, respectively. Section 3 presents the proposed multilevel kernel machine approach and introduces an adaptive weight learning scheme. Section 4 presents experiment results on two public datasets. Finally, we conclude the paper in Section 5.

2. Related Works

In this section, we describe Census Transform, PACT, and spatial PACT in detail, respectively.

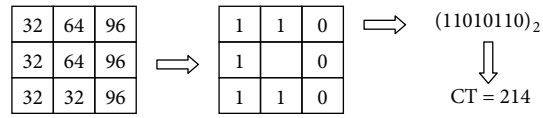


Figure 1: The CT operation.

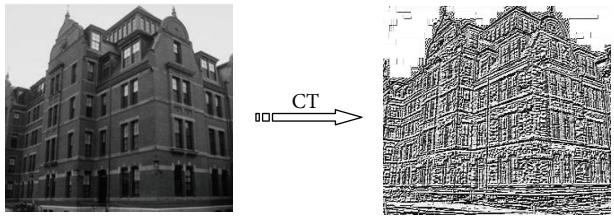
2.1. Census Transform and CENTRIST. According to descriptions in [17, 19], Census Transform is a nonparametric local transform, which is originally designed to establish the correspondence between image patches. Census Transform compares the intensity value of a pixel with its eight neighboring pixels (see illustration in Figure 1). If the center pixel is bigger than (or equal to) one of its neighbors, a bit 1 is set at the corresponding location. Otherwise a bit 0 is set.

Usually the eight bits generated from intensity comparisons can be put together in any order. In our work, we adopt the same order as that in [17] and collect these bits from left to right and top to bottom. Consequently the eight bits are converted to a base-10 number, which is the CT value for this center pixel. Another important property of the transform is that CT values of neighboring pixels are highly correlated. What is more, similar to other nonparametric local transforms which are based on intensity comparisons, Census Transform is robust to illumination changes, gamma variations.

As an intuitive representation, a Census Transformed image is created by replacing a pixel with its CT value. An example in Figure 2 shows that the Census Transform retains global structures of the image (e.g., especially discontinuities), meanwhile, captures the local structures as what is expected. Thus, a histogram of the CT values in an image (or image patch) encodes both local and global information.

2.2. PACT and Spatial PACT. CT values, denoted in the base-2 format, represent homogeneous regions with small variations (e.g., $(00001000)_2$). That is, there exist strong correlations between pairs of CT values. In order to remove these correlation effects and get a more compact image representation, Wu and Rehg [17] propose PACT by performing the PCA operation on the CT histograms. When computing histograms and PCA, they remove two bins with CT = 0 and CT = 255 and further normalize the CT histograms and eigenvectors such that they have zero mean and unit norm.

Since PACT only encodes global shape structure in a small image patch [17], to capture the global structure of an image in larger scales, Wu and Rehg [17] also propose a spatial PACT representation based on the SPM approach [20]. The spatial pyramid scheme divides an image into patches and concatenates corresponding PACT features in these patches. Since it encodes roughly spatial structure of an image, the recognition accuracy usually can be improved. As described in [20], the level 2 split in a spatial pyramid divides the image into $2^2 \times 2^2 = 16$ blocks. We also shift the division (dash line blocks) to avoid artifacts produced by the nonoverlapping division, which makes a total of 25 blocks in level 2. Similarly, level 1 and level 0 have 5 blocks and 1 block, respectively. It is noted that the image is resized at different levels so that all blocks contain the same number of pixels. These blocks



(a) Example image (b) Transformed image

Figure 2: An example of Census Transformed image.

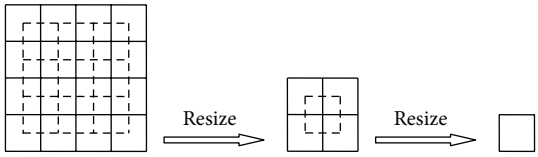


Figure 3: An intuitive illustration of the level 2, 1, and 0 of spatial pyramid split of an image, respectively.

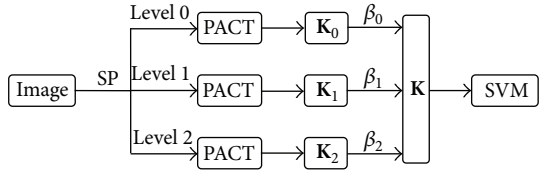


Figure 4: The proposed multilevel kernel machine scheme.

in different levels are shown in Figure 3. PACT features in all blocks are then concatenated to form an overall long feature vector. For example, if 40 eigenvectors are selected in PACT, the $L = 3$ levels spatial pyramid will generate a long feature vector whose dimension is $40 \times (25 + 5 + 1) = 1240$.

After spatial PACT feature vectors are extracted from images, SVM classifier is used to perform scene image classification task. In addition, the CT values are only based on pixel intensity comparisons, so it might be helpful to include a few image statistics, such as average value and standard deviation of pixels in an image block. This statistical information is appended to spatial PACT in the input to SVM classifiers for scene classification problem.

3. Proposed Approach

The original spatial PACT, however, only simply concatenates compact histograms at all levels together and discards the difference among different pyramid levels. In this section, we will introduce an adaptive multilevel kernel machine scheme. It firstly computes a set of base kernel matrixes corresponding to each level of the pyramid of PACT, and then a promising adaptive weight learning scheme is used to seek a set of optimal weights for best fusing all base kernels for scene image categorization.

3.1. Multilevel Kernel Machine. Let $\mathbf{x} = [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{L-1}]$ and $\mathbf{y} = [\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{L-1}]$ denote two spatial PACT feature vectors, where L is the total number of pyramid levels and $\mathbf{x}_l$ (or $\mathbf{y}_l$) is PACT feature vector located at the level l ($l = 0, \ldots, L-1$). Generally speaking, PACT features from various levels have different contributions to image representation. In order to fully exploit the descriptive capacity of each level, we compute a set of base kernels corresponding to each pyramid level of PACT. The aim is to learn a SVM classifier [18, 21], where the kernel is learnt to be a linear combination of given base kernels [22].

Here there are several ways to estimate similarity between $\mathbf{x}_l$ and $\mathbf{y}_l$, such as RBF kernel and linear kernel. By coupling each representation with its corresponding distance function, we obtain a set of similarity-based kernel matrices $\{\mathbf{K}_l\}_{l=0}^{L-1}$, for RBF kernel:

$$\mathbf{K}_l(\mathbf{x}_l, \mathbf{y}_l) = k_l(\mathbf{x}_l, \mathbf{y}_l) = \exp\left(-\gamma \left\|\mathbf{x}_l - \mathbf{y}_l\right\|^2\right) \tag{1}$$

or for linear kernel:

$$\mathbf{K}_l(\mathbf{x}_l, \mathbf{y}_l) = k_l(\mathbf{x}_l, \mathbf{y}_l) = \mathbf{x}_l \mathbf{y}_l^T. \tag{2}$$

Here, $\gamma > 0$ and $\mathbf{K}_l$ is the similarity kernel corresponding to level l. Our multilevel kernel machine approach is to find an optimal linear combination of given base kernels. Suppose we have obtained a set of base kernel functions $\{k_l\}_{l=0}^{L-1}$ (or base kernel matrices $\{\mathbf{K}_l\}_{l=0}^{L-1}$), an ensemble kernel function k (or an ensemble kernel matrix $\mathbf{K}$) is then defined by

$$\begin{aligned} k(\mathbf{x}_i, \mathbf{y}_i) &= \sum_{l=0}^{L-1} \beta_l k_l(\mathbf{x}_i, \mathbf{y}_i) \quad \left(\sum_{l=0}^{L-1} \beta_l = 1,\ \beta_l \geq 0\right), \\ \mathbf{K} &= \sum_{l=0}^{L-1} \beta_l \mathbf{K}_l \quad \left(\sum_{l=0}^{L-1} \beta_l = 1,\ \beta_l \geq 0\right). \end{aligned} \tag{3}$$

With (1), (2), and (3), determining a set of optimal coefficients $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_{L-1}\}$ can now be interpreted as finding appropriate weights for best fusing these spatial pyramid levels of PACT feature representations. Lastly, SVM classifier with kernel matrix $\mathbf{K}$ can be used to perform scene image classification task.

The flowchart of our proposed multilevel kernel machine scheme is shown in Figure 4 for three levels. Note that SP means spatial pyramid scheme.

Figure 5: Some example images from the 15-class scene category dataset (two images per category).

3.2. Adaptive Weight Learning Scheme. In our previous work [23], the weight parameters $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_{L-1}\}$ are empirically set by a large number of experiments, which is a time consuming procedure and does not always achieve reasonable results. To address this problem, in this paper, we adopt an adaptive weight learning scheme; our aim is to find the optimum weights to best fuse the entire various PACT features at different levels by a promising method.

According to the descriptions in [24], their basic idea of weight learning is that finding optimal weight is actually equivalent to the problem of fusing predefined kernels according to the label information, which is a standard kernel matrix learning problem.

As for kernel matrix learning, it is always believed that the target similarity should be close to the label similarity. Therefore, we need to seek $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_{L-1}\}$, such that the matrix $\mathbf{K}$ is close to the label similarity matrix $\mathbf{C}$. Here, the label similarity is defined as follows:

$$C_{i,j} = \begin{cases} 1 & \text{if } c_i = c_j \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where c_i is the corresponding label information of target images $\mathbf{I}_i$.

Then, the following criterion is used to minimize the distance between $\mathbf{K}$ and $\mathbf{C}$, which is essentially a quadratic programming problem:

$$\|\mathbf{K} - \mathbf{C}\|_F^2 = \sum_i \sum_j \left(K_{i,j} - C_{i,j}\right)^2. \tag{5}$$

Here $\|\mathbf{X}\|_F^2 = \mathrm{tr}(\mathbf{X}\mathbf{X}^T)$, where $\mathrm{tr}(\cdot)$ means trace operation. More intuitively, $\|\mathbf{K} - \mathbf{C}\|_F^2$ is the sum of element-wise distance between $\mathbf{K}$ and $\mathbf{C}$, where the subscript "F" denotes Frobenius norm operation.

In practice, if we use (5) directly, we may suffer a scale problem. As the elements in K are not scaled in $[0, 1]$, minimizing $\|\mathbf{K} - \mathbf{C}\|_F^2$ may not guarantee a reasonable result. In order to address the scaling problem, all the elements in $\mathbf{K}$ should be constrained within $[0, 1]$ by following method:

$$\begin{aligned} \mathbf{K} &= \sum_{l=0}^{L-1} \beta_l \overline{\mathbf{K}_l}, \\ m_l &= \max\left(\left|\mathbf{K}_l\right|\right), \\ \overline{\mathbf{K}_l} &= \frac{\mathbf{K}_l}{m_l}, \quad l = 0, \ldots, L-1. \end{aligned} \tag{6}$$

	Suburb	Coast	Forest	Highway	Inside city	Mountain	Open country	Street	Tall building	Office	Bedroom	Industrial	Kitchen	Living room	Store
Suburb	99.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0
Coast	0.4	85.0	0.4	1.9	0.0	1.9	9.6	0.0	0.0	0.4	0.4	0.0	0.0	0.0	0.0
Forest	0.0	0.0	92.5	0.0	0.0	4.8	2.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4
Highway	0.0	2.5	0.0	87.5	1.9	1.3	3.1	1.3	0.0	0.0	0.6	0.0	0.6	0.6	0.6
Inside city	1.9	0.0	0.0	0.0	83.2	0.0	0.0	3.4	7.7	0.0	0.0	1.4	0.0	0.5	1.9
Mountain	0.0	1.8	2.9	0.4	0.0	87.2	6.9	0.0	0.4	0.4	0.0	0.0	0.0	0.0	0.0
Open country	0.3	9.7	3.9	2.3	0.0	6.8	76.5	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Street	1.0	0.0	0.0	1.0	4.2	0.0	0.0	90.1	2.1	0.0	0.0	0.5	0.0	1.0	0.0
Tall building	0.0	0.0	0.0	0.4	5.9	2.3	0.0	0.4	86.7	0.0	0.4	2.7	0.0	0.4	0.8
Office	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	93.3	0.9	0.0	2.6	1.7	0.0
Bedroom	0.0	0.0	0.0	0.0	0.9	1.7	0.0	0.9	0.0	2.6	69.8	0.9	5.2	16.4	1.7
Industrial	0.0	0.5	0.9	0.5	2.8	0.5	0.5	0.5	1.4	0.0	1.9	74.9	1.9	0.0	13.7
Kitchen	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	4.5	3.6	2.7	77.3	0.9	10.0
Living room	0.0	0.0	0.0	0.0	1.6	0.0	0.0	0.5	1.1	3.2	21.2	1.1	1.6	68.8	1.1
Store	0.0	0.0	0.5	0.0	1.4	0.0	0.0	0.0	0.9	0.5	1.4	3.3	2.8	0.0	89.3

FIGURE 6: The Confusion Matrix on the 15-class scene category dataset (%). Average classification rates for individual classes are listed along the diagonal. The entry in the ith row and jth column is the percentage of images from class i that are misidentified as class j. Here, average classification accuracy is 84.09%.

FIGURE 7: Some misclassified instances in the 15-class scene category dataset. The images in the first row are from bedroom, but they are misclassified as living room (the misclassified rate is about 16.40 percent), and the images in the second row are from class living room, but they are misclassified as bedroom (the misclassified rate is about 21.20 percent).

Here $\overline{\mathbf{K}_l}$ is a normalized variant of $\mathbf{K}_l$, "$|\,|$" is the absolute operation, and m_l is the largest absolute value in $\mathbf{K}_l$.

To avoid overfitting in the learning procedure, a regularization term $\|\boldsymbol{\beta}\|^2$ can be added to (5). Then, we can find weights $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_{L-1}\}$ by solving the optimal problem:

$$\begin{aligned} \min_{\beta} \quad & \|\mathbf{K} - \mathbf{C}\|_F^2 + \lambda \|\boldsymbol{\beta}\|^2 \\ \text{s.t.} \quad & \sum_{l=0}^{L-1} \beta_l = 1, \quad \beta_l \geq 0, \end{aligned} \tag{7}$$

where λ is a tradeoff parameter to prevent overfitting problem. In our experiments, the parameter λ is set to $0.1 \times \mathrm{tr}(\mathbf{K}_0 \mathbf{K}_0^T)$, and,

$$\mathbf{K}_0 = \frac{1}{L} \sum_{l=0}^{L-1} \overline{\mathbf{K}_l}. \tag{8}$$

More intuitively, we may reformulate the first term in (7) as following:

$$\begin{aligned} & \|\mathbf{K} - \mathbf{C}\|_F^2 \\ & = \left\| \sum_{l=0}^{L-1} \beta_l \overline{\mathbf{K}_l} - \mathbf{C} \right\|_F^2 \\ & = \mathrm{tr}\left(\left(\sum_{l=0}^{L-1} \beta_l \overline{\mathbf{K}_l} \right)^T \left(\sum_{l=0}^{L-1} \beta_l \overline{\mathbf{K}_l} \right) - 2\mathbf{C}^T \sum_{l=0}^{L-1} \beta_l \overline{\mathbf{K}_l} + \mathbf{C}^T \mathbf{C} \right) \\ & = \boldsymbol{\beta}^T \mathbf{A} \boldsymbol{\beta} - 2\mathbf{b}^T \boldsymbol{\beta} + d, \end{aligned} \tag{9}$$

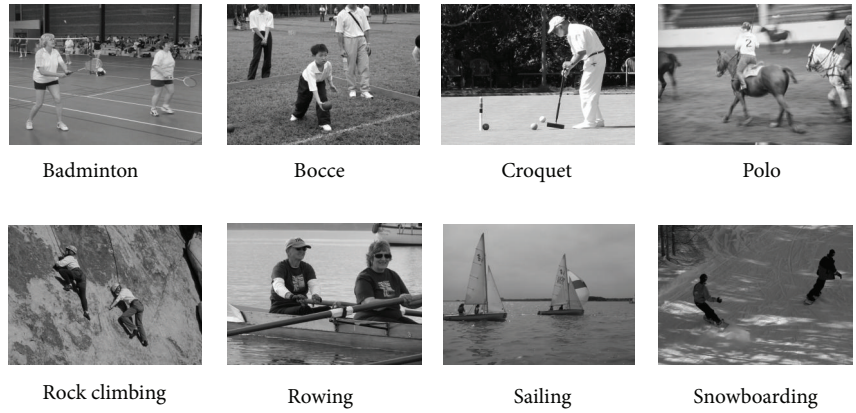


Figure 8: Images from 8 sports event dataset (one image per category).

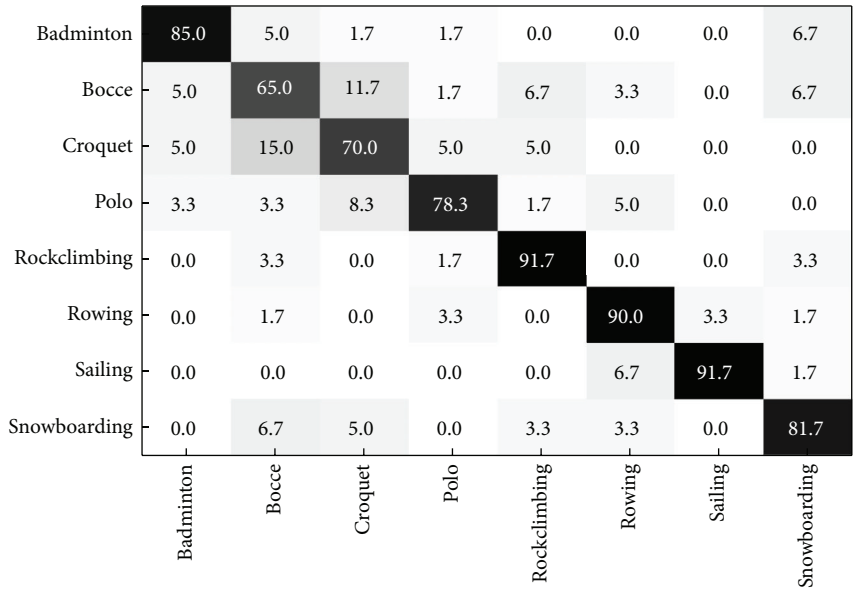


Figure 9: The Confusion Matrix on the 8 class event dataset (%). In this run, the average classification accuracy is 81.67%.

where

$$A_{i,j} = \mathrm{tr}\left(\left(\overline{\mathbf{K}_l}\right)^T \overline{\mathbf{K}_l}\right),$$
$$b_j = \mathrm{tr}\left(\mathbf{C}^T \overline{\mathbf{K}_l}\right), \tag{10}$$
$$d = \mathrm{tr}\left(\mathbf{C}^T \mathbf{C}\right).$$

By combining (7) and (9) together, the optimal problem in (7) can be reformulated as the quadratic programming, and the optimal weights $\boldsymbol{\beta} = \{\beta_0, \beta_1, \ldots, \beta_{L-1}\}$ can be obtained as following:

$$\min_{\beta} \quad \boldsymbol{\beta}^T (\mathbf{A} + \lambda \mathbf{E}) \boldsymbol{\beta} - 2\mathbf{b}^T \boldsymbol{\beta}$$
$$\text{s.t.} \quad \sum_{l=0}^{L-1} \beta_l = 1, \quad \beta_l \geq 0. \tag{11}$$

Here, matrix $\mathbf{E}$ is the identity matrix with the same size of matrix $\mathbf{A}$.

4. Experimental Results

In this section, we implement and evaluate our methods on two widely used datasets: 15-class scene category dataset [13] and the 8-class event dataset [25].

4.1. Experimental Setup. In our setup, following the same experiment procedure of the spatial PACT [17], the images are all preprocessed into gray scale. For each dataset, experiments are repeated five times to obtain reliable results by randomly selecting training and testing images. The average of per-class classification accuracy is recorded for each run, and the final results are shown as the mean and standard deviation of the classification accuracy. For spatial PACT, CENTRIST descriptors are reduced to 40 dimensions by PCA, and the

Table 1: Classification accuracy comparison on the 15-class scene category dataset.

L	Method	Feature type	Classification accuracy (%)
2	SPM [13]	SIFT, 400 cluster centers	81.40 ± 0.50
2	ScSPM [14]	SIFT, 400 cluster centers	80.28 ± 0.93
2	Spatial PACT [17]	CENTRIST, 40 eigenvectors	83.38 ± 0.53
2	Our method	CENTRIST, 40 eigenvectors	83.45 ± 0.47

Table 2: Performance comparison using different kernels on the 15-class scene category dataset.

L	Method	RBF kernel (1)	Linear kernel (2)
2	spatial PACT [17]	83.38 ± 0.53	81.92 ± 0.48
2	Our method	83.45 ± 0.47	82.51 ± 0.70

largest 40 eigenvalues accounted for about 90 percents and 94 percents of the sum of all eigenvalues in these datasets, respectively. In regard to computing CENTRIST descriptors and PCA eigenvectors, we remove the two bins with CT = 0, 255 and normalize the CENTRIST descriptors and PCA eigenvectors such that they have zero mean and unit norm. Also, average value and standard deviation of pixels in each image block are appended to spatial PACT for including some statistics information. In this paper, $L = 3$ is used in spatial PACT for all datasets, which has achieved the best performance in [17]. Thus, the dimension of the spatial PACT feature vector of $L = 3$ levels becomes $(40 + 2) \times (25 + 5 + 1) = 1302$. The dimension of PACT vector in pyramid level $l = 2$ is $(40 + 2) \times 25 = 1050$, level $l = 1$ is $(40 + 2) \times 5 = 210$, and level $l = 0$ is $(40 + 2) \times 1 = 42$, respectively. Lastly, multiclass classification task is done using an excellent toolbox LibSVM [26] which is trained by one-versus-one rule and parameters $(C, \gamma) = (8, 2^{-7})$ are recommended for RBF kernel [17].

4.2. The 15-Class Scene Category Dataset. The 15-class scene category dataset contains totally 4485 images with 15 categories (including coast, forest, kitchen, street, and office). Images are about 300 × 250 in average size, with 210 to 410 images in each category. This dataset contains a wide range of scene categories in both indoor and outdoor environments. In addition, this is one of the most complete scene category dataset used in the literature so far. Figure 5 lists several representative example images from this dataset.

Following the same experimental procedure of the spatial PACT [17], the first 100 images in each category are used to obtain a projection matrix by PCA, 100 images per category are chosen randomly for training and the remaining images for testing. The comparison results are shown in Table 1. From this table, we can see that our adaptive multilevel machine method achieves the highest accuracy and outperforms SPM about 2 percents, ScSPM, about 3 percents, respectively. To further illustrate the performance of our adaptive multilevel kernel machine scheme, Table 2 also shows the performance comparison using different kernels, respectively. We can find that linear kernel does not perform as well as RBF kernel, but the difference is not significant (about 1 percent). In addition, we can find that the linear kernel has more improvement in classification accuracy than RBF kernel. To summary, all these results show that fusing different levels PACT can effectively improve classification performance.

Moreover, Figure 6 shows a confusion matrix from one run on this dataset using our approach. It can be observed that our method performs better for categories suburb, forest, street, and office (more than 90%). Meanwhile, we also notice that the categories bedroom and living room have a high percentage being classified wrongly. Not surprisingly, this may result from that they are visually similar to each other. Figure 7 lists some misclassified images.

4.3. The 8-Class Event Dataset. The event dataset [25] contains 1579 images of eight sport event classes: 200 badminton, 137 bocce, 236 croquet, 182 polo, 194 rock climbing, 250 rowing, 190 sailing, and 190 snowboarding, respectively. The images in this dataset are almost with small resolution variations and are high-resolution ones. Figure 8 lists some images from this dataset. In [25], this dataset is used to classify these events by integrating scene and object categorizations. In our work, we use this dataset for image scene classification purposes only.

Also, following the common benchmarking procedures on this dataset, we select 70 images per category for training and 60 for testing, randomly. The first 50 images in each class are used to produce the eigenvectors by PCA operation. The performance comparison results are shown in Table 3. From Table 3, we can see that our method achieves the best result on this dataset and outperforms spatial PACT by nearly 1.5 percent in classification accuracy, which shows the effectiveness of our adaptive multilevel kernel machine approach. Similar to the 15-class scene category dataset, we also list confusion matrix on this dataset, as shown in Figure 9. It can be observed that confusion occurs between bocce and croquet and also between polo and croquet, which is consistent with the result of spatial PACT. It is because that these two pairs of events share very similar scene or background.

5. Conclusions

In this paper, we propose an adaptive multilevel kernel machine method for scene classification. Unlike the original spatial PACT, which simply concatenates all pyramid levels of compact PACT histograms together and discards the difference between different pyramid levels, our proposed scheme computes a set of base kernels at each pyramid level of PACT and thus considers the difference existing in various levels. Since PACT features in each pyramid level play

Table 3: The performance comparison (%) on the 8-class event dataset.

L	Method	Feature type	RBF kernel (1)	Linear kernel (2)
2	Spatial PACT [17]	CENTRIST, 40 eigenvectors	79.08 ± 1.15	76.96 ± 1.16
2	Our method	CENTRIST, 40 eigenvectors	80.50 ± 1.49	78.38 ± 1.28

different roles in representing scene, the optimal kernel is learnt to be a linear combination of given base kernels; that is, it can be interpreted as finding appropriate weights for best fusing all pyramid levels PACT feature representations. To seek a set of optimal weights, a promising adaptive weight learning method is employed. Experiments on two popular benchmark datasets are present to demonstrate that our adaptive multilevel kernel machine method outperforms the original spatial PACT on image scene classification.

Acknowledgment

This research work was supported by the Grants from the National Natural Science Foundation of China (Nos. 90820010, 61101207).

References

[1] N. Snavely, R. Garg, S. M. Seitz, and R. Szeliski, "Finding paths through the world's photos," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 11–21, 2008.

[2] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.

[3] P. Quelhas, F. Monay, J. M. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. Van Gool, "Modeling scenes with local descriptors and latent aspects," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 883–890, October 2005.

[4] A. Shabou and H. LeBorgne, "Locality-constrained and spatially regularized coding for scene categorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3618–3625, 2012.

[5] J. H. Zhang, J. W. Zhang, S. Y. Chen, Y. Hu, and H. J. Guan, "Constructing dynamic category hierarchies for novel visual category discovery," in *Proceedings of the Intelligent Robots and Systems (IROS '12)*, pp. 2122–2127, 2012.

[6] D. Gokalp and S. Aksoy, "Scene classification using bag-of-regions representations," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.

[7] S. Y. Chen, J. H. Zhang, Y. F. Li, and J. W. Zhang, "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction," *IEEE Trans. Industrial Informatics*, vol. 8, no. 1, pp. 118–127, 2012.

[8] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.

[9] T. Hofmann, "Unsupervised learning by probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 42, no. 1-2, pp. 177–196, 2001.

[10] R. Fergus, P. Perona, and A. Zisserman, "A sparse object category model for efficient learning and exhaustive recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 380–387, June 2005.

[11] J. Wu and J. M. Rehg, "Beyond the Euclidean distance: creating effective visual codebooks using the histogram intersection kernel," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 630–637, October 2009.

[12] D. Xu and S. F. Chang, "Video event recognition using kernel methods with multilevel temporal alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1985–1997, 2008.

[13] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 2169–2178, June 2006.

[14] J. C. Yang, K. Yu, Y. H. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1794–1801, June 2009.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[16] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, June 2008.

[17] J. Wu and J. M. Rehg, "CENTRIST: a visual descriptor for scene categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1489–1501, 2011.

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[19] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proceedings of the European Conference on Computer Vision*, pp. 151–158, 1994.

[20] K. Grauman and T. Darrell, "The pyramid match kernel: discriminative classification with sets of image features," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 1458–1465, October 2005.

[21] B. Scholkopf and A. Smola, *Learning with Kernels*, MIT Press, 2002.

[22] M. Varma and D. Ray, "Learning the discriminative power-invariance trade-off," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, October 2007.

[23] J. Hu and P. Guo, "Multi-level kernel machine for scene image classification," in *Proceedings of the International Conference on Computational Intelligence and Security*, pp. 1169–1173, 2011.

[24] J. He, S. F. Chang, and L. Xie, "Fast kernel learning for spatial pyramid matching," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–7, June 2008.

[25] L.-J. Li and L. Fei-Fei, "What, where and who? Classifying events by scene and object recognition," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV 07)*, October 2007.

[26] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 471963, 7 pages
http://dx.doi.org/10.1155/2013/471963

Research Article

Normality of Ethernet Traffic at Large Time Scales

Zhiping Lu,[1] Ming Li,[1] and Wei Zhao[2]

[1] *School of Finance and Statistics, East China Normal University, No. 500 Dong-Chuan Road, Shanghai 200241, China*
[2] *Department of Computer and Information Science, University of Macau, Padre Tomas Pereira Avenue, Taipa, Macau*

Correspondence should be addressed to Zhiping Lu; camping9@gmail.com

Received 19 January 2013; Accepted 4 February 2013

Academic Editor: Carlo Cattani

We contribute the quantitative descriptions of the large time scales for the Ethernet traffic to be Gaussian. We focus on the normality property of the accumulated traffic data under different time scales. The investigation is carried out graphically by the quantile-quantile (QQ) plots and numerically by statistical tests. The present results indicate that the larger the time scale, the more normal the Ethernet traffic.

1. Introduction

The experimental research of the internet traffic (traffic for short), including the Ethernet one, exhibits that fractional Gaussian noise (fGn) may be a model in the sense of unifractal see, for example, [1–3]. This implies that traffic is Gaussian [4]. However, non-Gaussian models, such as stable processes, were also reported; see, for example, [5–8]. Therefore, the normality of traffic is an issue worth investigating.

Research described in [9, 10] revealed a scaling phenomenon of traffic. Taking into account the scales of traffic, we say that whether a traffic trace is Gaussian or not relies on time scales. Paxson and Floyd [10] and Feldmann et al. [9] claimed that traffic is Gaussian at time scales larger than 1 second. That property was qualitatively further confirmed by [11]. Note that real-traffic data used in [9, 10] were recorded in 1980s and 1990s, which are publicly accessible [12]. Thus, one second, as the critical time point, corresponds to the data in [12] and the infrastructure of the internet then.

Though the research exhibits that the statistics of traffic remain the same from the internet last century to the current years [13], the quantity of the critical time point, say one second, may be vague due to the development of high-speed networking. Therefore, when using the same data as those used in [1, 3, 9, 10], we use the concept of packet count, that is, the number of packets within an interval, to represent the number of bytes of packets within an interval.

Let $x(t(i))$ be a sample record of traffic time series, where $t(i)$ $(i = 0, 1, \ldots)$ is the series of time stamps, indicating the time stamp of the ith packet. The series $x(t(i))$ therefore represents the packet size of the ith packet at time $t(i)$. In this research, instead of using $x(t(i))$, we use $x(i)$ representing the packet size of the ith packet. On an interval-by-interval basis, therefore, the accumulated traffic, denoted by $y(n)$, is given by

$$y(n) = \sum_{i=nT}^{(n+1)T} x(i), \tag{1}$$

where T is the interval width, which also has the similar meaning of time scales. Thus, $y(n)$ stands for the accumulated bytes of arrival traffic in the nth interval. The statistics of $y(n)$ may considerably differ when T is small (small time scale) or large (large time scale) [1, 9, 10].

This research utilizes four real-traffic traces, listed in Table 1, which were measured on an Ethernet at the Bellcore Morristown Research and Engineering facility in 1989 [12]. (the originally statistical properties described in the early literature, e.g., [1, 3], turn to be ubiquitous in today's traffic, according to the research stated in [13]. Thus, the traffic trace, BC-Aug89, which was measured in 1989, keeps its value in the description of traffic pattern today).

Figure 1 illustrates four series of real-traffic trace BC-Aug89. Note that the statistics of $x(t(i))$ is consistent with that of $x(i)$, but we may obtain the time scale represented by T in (1), which is irrelevant of the networking speed. Let the interval width be $T = 1024$. Then, Figure 2 indicates $y(n)$ of BC-Aug89 for $T = 1024$.

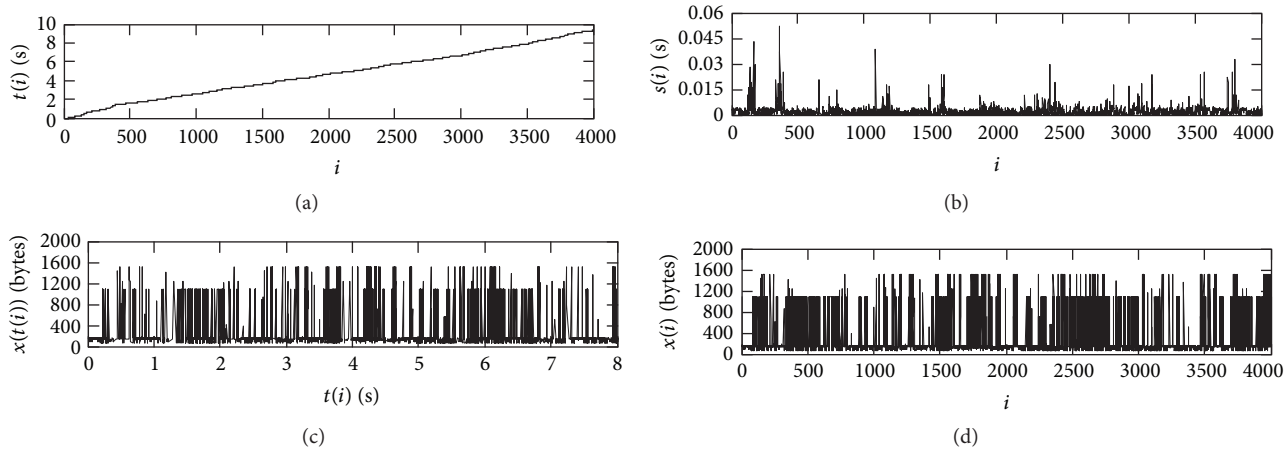


Figure 1: Illustrations of real-traffic trace BC-Aug89. (a) Timestamp series $t(i)$. (b) Interarrival times $s(i)$. (c) Traffic in packet size $x(t(i))$. (d) Traffic in packet size $x(i)$.

Table 1: Four traffic series.

Series name	Starting time	Duration	Series length
pAug.TL	11:25 AM, 29 Aug 89	52 minutes	1 million
pOct.TL	11:00 AM, 05 Oct 89	29 minutes	1 million
OctExt.TL	11:46 PM, 03 Oct 89	34.111 h	1 million
OctExt4.TL	2:37 PM, 10 Oct 89	21.095 h	1 million

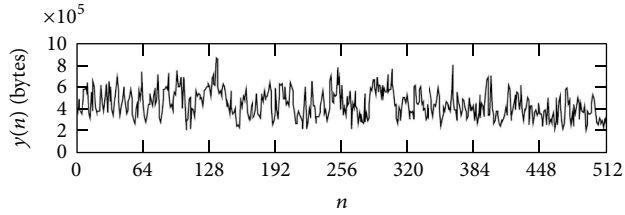


Figure 2: Accumulated traffic of BC-Aug89 with the interval width $T = 1024$.

The paper aims at presenting the quantitatively minimum interval range for the accumulated Ethernet traffic traces to be Gaussian based on the accumulated bytes of the packets within an interval.

The remainder of this paper is organized as follows. In Section 2 we introduce briefly the commonly used normality tests and the idea of the QQ plot. The graphical and numerical results are presented in Section 3, and the discussion of the investigation results is followed in Section 4. Section 5 concludes the paper.

2. Statistical Investigation for Accumulated Traffic

In this section, we discuss the normality tests for the following null and alternative hypotheses:

H_0: the data are sampled from a normal distribution;

H_1: the data are not sampled from a normal distribution.

Many statistical tests have been proposed to find out whether a sample is drawn from a normal distribution or not [14], including the Shapiro-Wilk test, D'Agostino's K^2 test, the Jarque-Bera test, the Anderson-Darling test, the Cramér-Von Mises criterion, the Lilliefors test, the Pearson's χ^2 test, and the Shapiro-Francia test.

The absence of exact solutions for the sampling distributions generated a large number of simulation studies exploring the power of these statistics. A convincing evidence from these studies is that convergence of the sampling distributions to asymptotic results was very slow. The paper [15] concludes that the Shapiro-Wilk test has the best power for a given significance, followed closely by Anderson-Darling test when comparing the Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors, and Anderson-Darling tests. On the other hand, some publications recommend the Jarque-Bera test [16, 17]. But it is not without weakness. It has low power for distributions with short tails. Therefore, we mainly consider three normality test methods listed in the following.

2.1. Shapiro-Wilk Test. The Shapiro-Wilk test tests the null hypothesis that a sample $y(1), \ldots, y(n)$ came from a normally distributed population [18]. The test statistic is

$$W = \frac{\left(\sum_{i=1}^{n} a_i y_{(i)}\right)^2}{\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2}, \tag{2}$$

where $y_{(i)}$ is the ith order statistic; $\overline{y}_i$ is the sample mean; a_i is given by

$$\left(a_1, \ldots, a_n\right) = \frac{m^T V^{-1}}{\left(m^T V^{-1} V^{-1} m\right)^{1/2}}, \tag{3}$$

where $m = (m_1, \ldots, m_n)$; and $m_1, \ldots, m_n$ are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution, and V is the covariance matrix of those order statistics. It is worth mentioning that the Shapiro-Wilk test is restricted for the sample size greater than 3 and less than 5000.

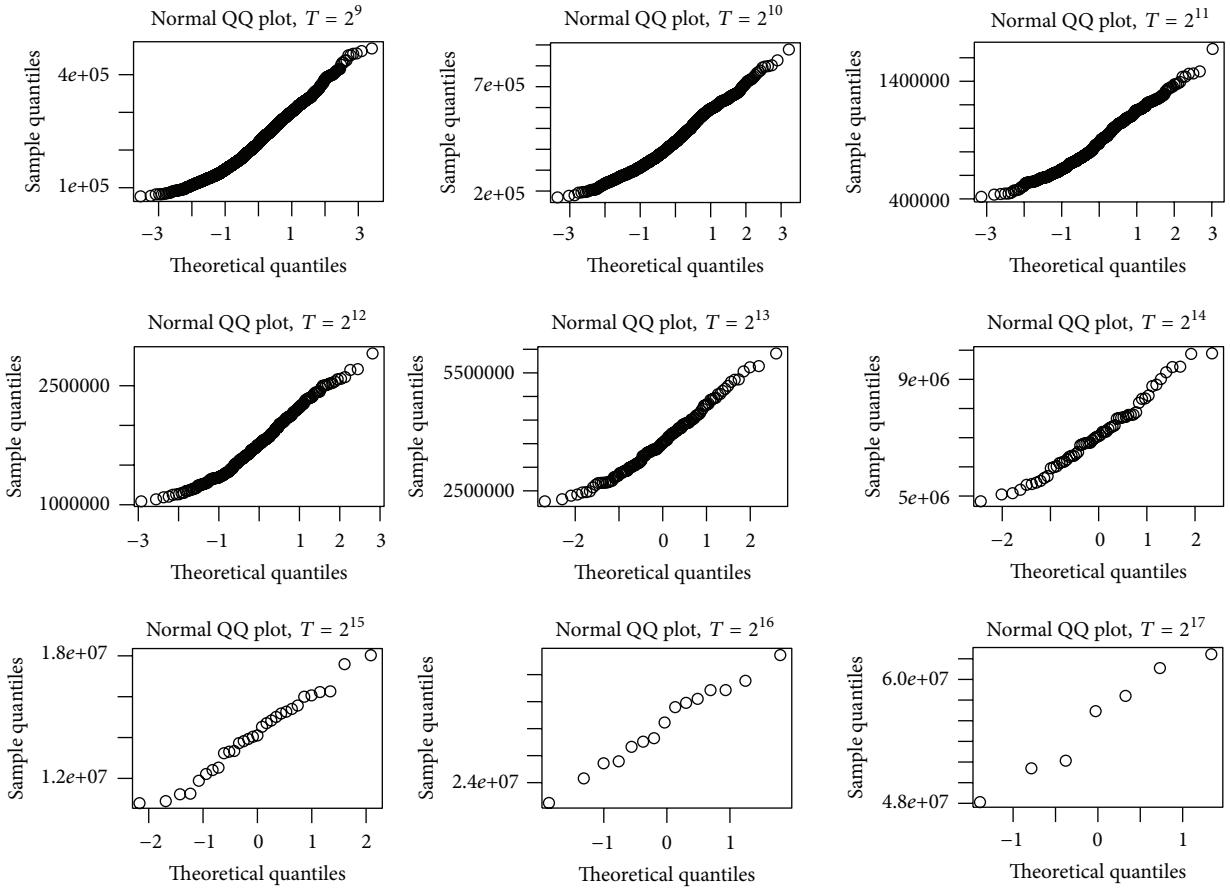


Figure 3: QQ-plot of the pAug.TL data under different time scales.

2.2. Anderson-Darling Test. The Anderson-Darling test is a statistical test of whether a given sample of data is drawn from a given probability distribution [19, 20]. In its basic form, the test assumes that there are no parameters to be estimated in the distribution being tested, in which case the test and its set of critical values are distribution free. When applied to testing if a normal distribution adequately describes a set of data, it is one of the most powerful statistical tools for detecting most departures from normality [21, 22], whereas the sample size needs to be greater than 7.

2.3. Jarque-Bera Test. The Jarque-Bera test is a goodness-of-fit test of whether sample data have the skewness and kurtosis matching a normal distribution [23, 24]. The test statistic JB is defined as

$$\mathrm{JB} = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right), \tag{4}$$

where

$$S = \frac{(1/n)\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^3}{\left((1/n)\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2\right)^{3/2}},$$
$$K = \frac{(1/n)\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^4}{\left((1/n)\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2\right)^{2}}. \tag{5}$$

If the data comes from a normal distribution, the JB statistic asymptotically has a $\chi^2(2)$ distribution, so the statistic can be used to test the hypothesis that the data is from a normal distribution. For small samples, the chi-squared approximation is overly sensitive, often rejecting the null hypothesis when it is in fact true. Thus, JB test only applies to large sample size, at least 7 according to the finite sample study.

Besides statistical tests, we have another informal but powerful tool to assess the normality property of the series, that is, the normal probability plot. This graphical tool is often called the quantile-quantile plot (QQ plot) of the standardized data against the standard normal distribution. The correlation between the sample data and normal quantiles measures how well the data is modeled by a normal distribution. For normal data, the points plotted in the QQ plot should fall approximately on a straight line, indicating high positive correlation.

3. Graphical and Statistical Results

In this section, we present the graphical and numerical results for all the Ethernet traffic series, that is, pAug.TL, pOct.TL, OctExt.TL, and OctExt4.TL data. Figures 3, 4, 5, and 6 are

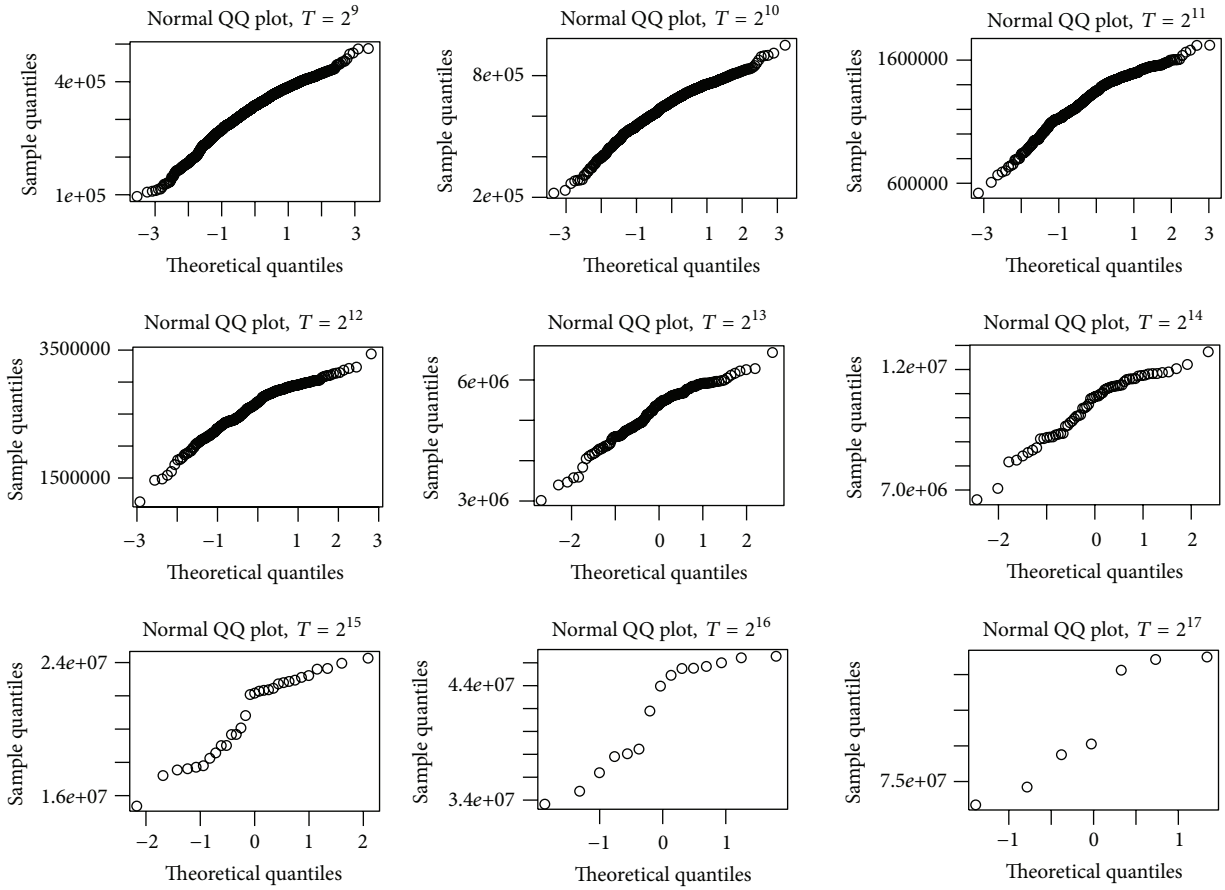


Figure 4: QQ-Plot of the pOct.TL data under different time scales.

the QQ-plot of the four accumulated traffic series under 9 different time scales.

In order to obtain a more complete inference for the series' normality and to be more objective, we finally choose to take advantage of three popular normality tests, that is, the Shapiro-Wilk test, Anderson-Darling test, and Jarque-Bera test to verify the normality property in the application. Based on the software R, we mainly utilize the functions of the packages "fBasics" and "nortest" to realize the statistical tests. The P-value of each test under the time scales $T = 2^n$, $n = 9, \ldots, 17$ are presented in Tables 2, 3, 4, and 5. In particular, since the Anderson-Darling test requires the sample size greater than 7, there is no testing result for the time scale $T = 2^{17}$.

Table 2: Normality test result for pAug.TL series.

Test	Shapiro-Wilk	Anderson-Darling	Jarque-Bera
$T = 512$	$3.222e-16$	$<2.2e-16$	$<2.2e-16$
$T = 1024$	$8.747e-10$	$2.945e-13$	$<2.2e-16$
$T = 2048$	$4.434e-06$	$3.977e-07$	0.004
$T = 4096$	0.0004069	0.001111	0.018
$T = 8192$	0.05681	0.1498	0.086
$T = 16384$	0.2819	0.5117	0.31
$T = 32768$	0.7258	0.8828	0.715
$T = 65536$	0.9528	0.8555	0.716
$T = 131072$	0.6796	N/A	0.524

4. Discussions

Graphically, from Figures 3, 4, 5 and 6, we have some findings listed below.

(i) Comparatively, the pAug.TL series asks for the relatively smallest time scale to be Gaussian among four series.

(ii) The pAug.TL and pOct.TL data seem more likely to be normal than the other two series at each corresponding time scale.

(iii) It is not difficult to observe that the OctExt.TL and OctExt4.TL series exhibit the similar normality behaviors. However, only at quite large time scale, the theoretical normal quantile and the empirical quantile have the high positive correlation.

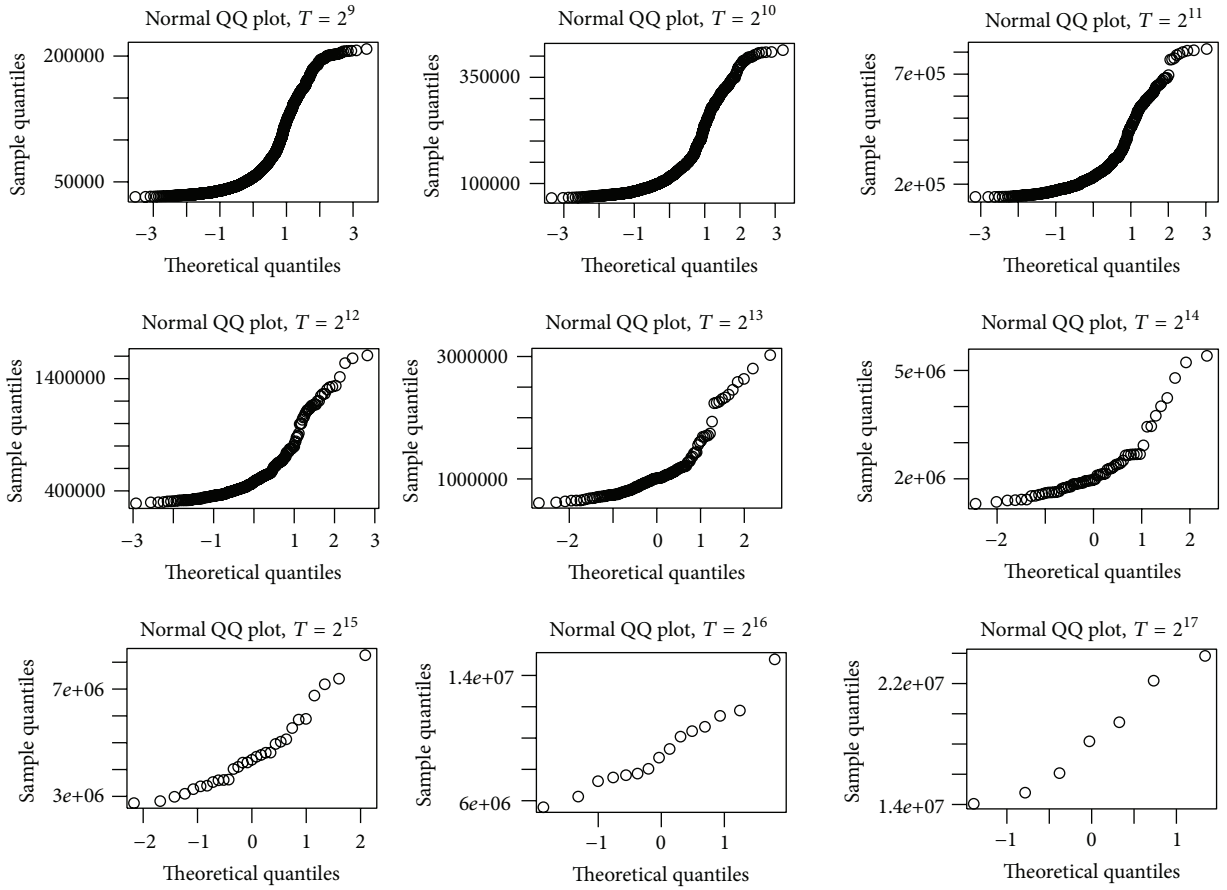


Figure 5: QQ-plot of the OctExt.TL data under different time scales.

Table 3: Normality test result for pOct.TL series.

Test	Shapiro-Wilk	Anderson-Darling	Jarque-Bera
$T = 512$	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 1024$	$4.112e-14$	$<2.2e-16$	$<2.2e-16$
$T = 2048$	$3.803e-11$	$2.125e-14$	$<2.2e-16$
$T = 4096$	$2.293e-07$	$2.801e-08$	$<2.2e-16$
$T = 8192$	0.0006598	0.0003278	0.015
$T = 16384$	0.006528	0.003969	0.051
$T = 32768$	0.01494	0.004604	0.118
$T = 65536$	0.02109	0.01588	0.139
$T = 131072$	0.1887	N/A	0.286

Table 4: Normality test result for OctExt.TL series.

Test	Shapiro-Wilk	Anderson-Darling	Jarque-Bera
$T = 512$	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 1024$	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 2048$	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 4096$	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 8192$	$6.15e-12$	$<2.2e-16$	$<2.2e-16$
$T = 16384$	$7.654e-08$	$4.916e-10$	$<2.2e-16$
$T = 32768$	0.02127	0.03261	0.048
$T = 65536$	0.5325	0.5585	0.208
$T = 131072$	0.6657	N/A	0.518

(iv) The OctExt4.TL series seems to be even more strict on the time scale. It requires minimum time scale about 65536 to be Gaussian.

Numerically, as could be expected, the testing results given in Tables 2, 3, 4, and 5 provide the evidence that the larger the time scale, the more normal the accumulated traffic series $y(n)$. Specifically,

(i) it is straightforward to see that the normality behavior of pAug.TL data "surpasses" the others according to the P values of the tests; that is, given the significance level $\alpha = 1\%$, the null hypothesis of normality could not be rejected when the time scale is greater than 8192;

(ii) whereas, the pOct.TL and OctExt.TL series possess the comparable normality performance who need the time scale to be at least 32768 in order not to be rejected by the null hypothesis given the significance level $\alpha = 1\%$.

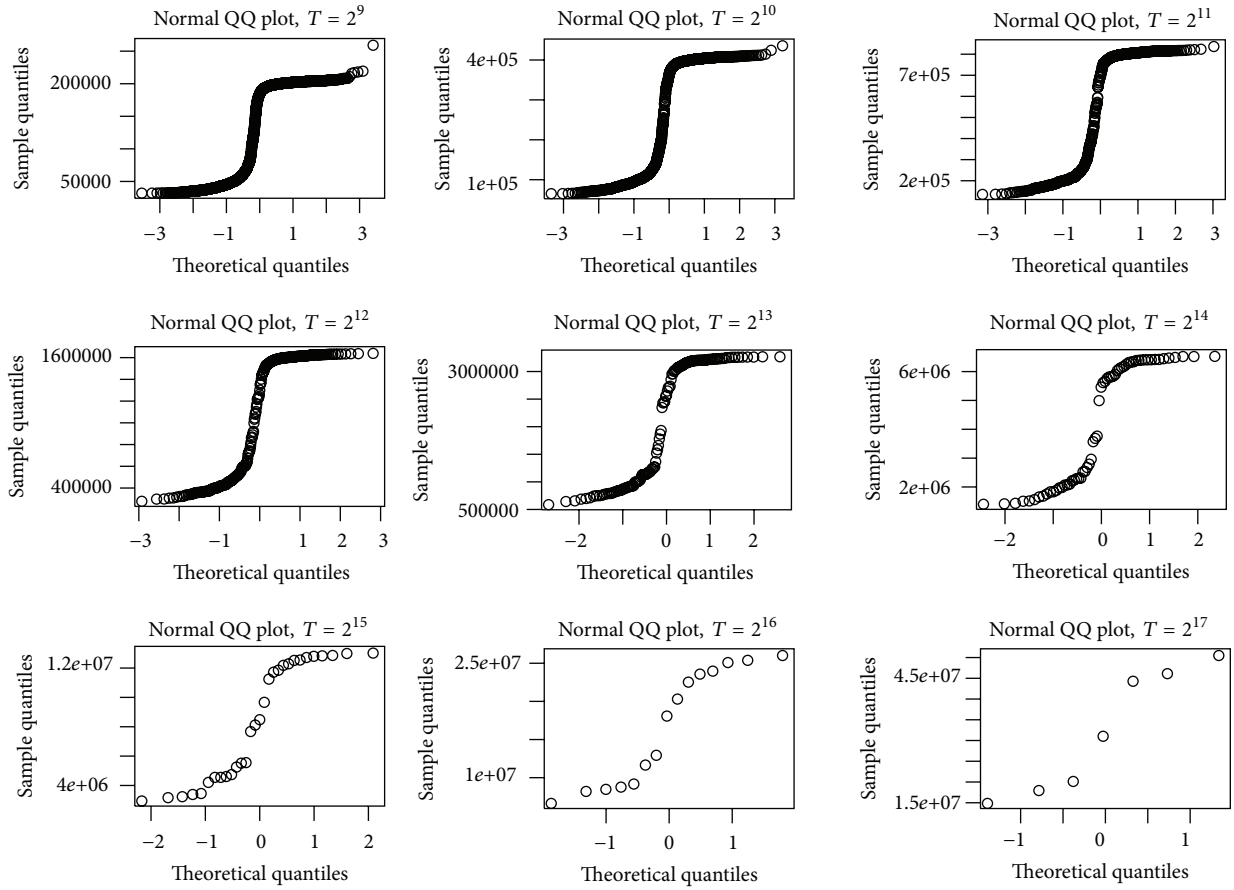


Figure 6: QQ-Plot of the OctExt4.TL data under different time scales.

Table 5: Normality test result for OctExt4.TL series.

Test	Shapiro-Wilk	Anderson-Darling	Jarque-Bera
$T = 512$	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 1024$	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 2048$	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 4096$	$<2.2e-16$	$<2.2e-16$	$<2.2e-16$
$T = 8192$	$6.398e-12$	$<2.2e-16$	0.005
$T = 16384$	$9.832e-08$	$8.577e-13$	0.021
$T = 32768$	0.0002948	$5.561e-05$	0.064
$T = 65536$	0.02198	0.02203	0.121
$T = 131072$	0.227	N/A	0.24

(iii) for the OctExt4.TL series, in order not to reject the null, the time scale should be greater than 65536 given the significance level $\alpha = 1\%$.

The previous discussions are for the Ethernet traffic, but the methods may also be a reference for other types of time series, such as those in [25–28].

5. Conclusions

We have discussed the normality performance of the Ethernet traffic data under different time scales using several normality tests (Shapiro-Wilk test, Anderson-Darling test, and Jarque-Bera test). The graphical results by QQ-plot are consistent with the numerical results, which also provides the evidence for the quantitative results of the large time scales for the normality of the Ethernet traffic traces investigated.

Acknowledgments

This work was in part supported by the 973 plan under the project grant number 2011CB302800, the National Natural Science Foundation of China under the project grant numbers 11101158, 61272402, 61070214, 60873264, and "the Fundamental Research Funds for the Central Universities". We appreciate W. Willinger, W. Leland, and D. Wilson with Bellcore, Morristown, who provided us with their data in this research.

References

[1] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[2] D. McDysan, *QoS & Traffic Management in IP & ATM Networks*, McGraw-Hill, New York, NY, USA, 2000.

[3] P. Abry and D. Veitch, "Wavelet analysis of long-range dependent traffic," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 2–15, 1998.

[4] W. Willinger and V. Paxson, "Where mathematics meets the internet," *Notices of the American Mathematical Society*, vol. 45, no. 8, pp. 961–970, 1998.

[5] Ph. Barbe and W. P. McCormick, "Heavy-traffic approximations for fractionally integrated random walks in the domain of attraction of a non-Gaussian stable distribution," *Stochastic Processes and Their Applications*, vol. 122, no. 4, pp. 1276–1303, 2012.

[6] R. G. Garroppo, S. Giordano, M. Pagano, and G. Procissi, "Testing α-stable processes in capturing the queuing behavior of broadband teletraffic," *Signal Processing*, vol. 82, no. 12, pp. 1861–1872, 2002.

[7] A. Karasaridis and D. Hatzinakos, "Network heavy traffic modeling using α-stable self-similar processes," *IEEE Transactions on Communications*, vol. 49, no. 7, pp. 1203–1214, 2001.

[8] G. Terdik and T. Gyires, "Lévy flights and fractal modeling of internet traffic," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 120–129, 2009.

[9] A. Feldmann, A. C. Gilbert, W. Willinger, and T. G. Kurtz, "The changing nature of network traffic: scaling phenomena," *ACM SIGCOMM Comput Communication Review*, vol. 28, no. 2, pp. 5–29, 1998.

[10] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, 1995.

[11] A. Scherrer, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry, "Non-Gaussian and long memory statistical characterizations for Internet traffic with anomalies," *IEEE Transactions on Dependable and Secure Computing*, vol. 4, no. 1, pp. 56–70, 2007.

[12] http://www.sigcomm.org/ITA/.

[13] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, and K. Cho, "Seven years and one day: sketching the evolution of internet traffic," in *Proceedings of the 28th Conference on Computer Communications (INFOCOM '09)*, pp. 711–719, Rio de Janeiro, Brazil, April 2009.

[14] H. C. Thode, Jr., *Testing for Normality*, vol. 164 of *Statistics: Textbooks and Monographs*, Marcel Dekker, New York, NY, USA, 2002.

[15] N. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[16] D. N. Gujarati, *Basic Econometrics*, McGraw-Hill, New York, NY, USA, 4th edition, 2002.

[17] G. G. Judge, R. C. Hill, W. E. Griffiths, H. Lütkepohl, and T. C. Lee, *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1988.

[18] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality: complete samples," *Biometrika*, vol. 52, no. 3-4, pp. 591–611, 1965.

[19] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes," *Annals of Mathematical Statistics*, vol. 23, pp. 193–212, 1952.

[20] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *Journal of the American Statistical Association*, vol. 49, pp. 765–769, 1954.

[21] M. A. Stephens, "EDF statistics for goodness of fit and some comparisons," *Journal of the American Statistical Association*, vol. 69, pp. 730–737, 1974.

[22] M. A. Stephens, "Tests based on EDF statistics," in *Goodness-of-Fit Techniques*, R. B. d'Agostino and M. A. Stephens, Eds., pp. 97–193, Marcel Dekker, New York, NY, USA, 1986.

[23] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics Letters*, vol. 6, no. 3, pp. 255–259, 1980.

[24] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence," *Economics Letters*, vol. 7, no. 4, pp. 313–318, 1981.

[25] C. Cattani, G. Pierro, and G. Altieri, "Entropy and multifractality for the myeloma multiple TET 2 gene," *Mathematical Problems in Engineering*, vol. 2012, Article ID 193761, 14 pages, 2012.

[26] C. Cattani, "On the existence of wavelet symmetries in archaea DNA," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 673934, 21 pages, 2012.

[27] C. Toma, "Advanced signal processing and command synthesis for memory-limited complex systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 927821, 13 pages, 2012.

[28] E. G. Bakhoum and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 203609, 7 pages
http://dx.doi.org/10.1155/2013/203609

Research Article

Marching Cubes Algorithm for Fast 3D Modeling of Human Face by Incremental Data Fusion

Xiangsheng Huang, Xinghao Chen, Tao Tang, and Ziling Huang

Institute of Automation, Chinese Academy of Sciences, Beijng 100090, China

Correspondence should be addressed to Xiangsheng Huang; xiangsheng.huang@ia.ac.cn

Received 22 December 2012; Accepted 20 January 2013

Academic Editor: Sheng-Yong Chen

We present a 3D reconstruction system to realize fast 3D modeling using a vision sensor. The system can automatically detect the face region and obtain the depth data as well as color image data once a person appears in front of the sensor. When the user rotates his head around, the system will track the pose and integrate the new data incrementally to obtain a complete model of the personal head quickly. In the system, iterative closest point (ICP) algorithm is first used to track the pose of the head, and then a volumetric integration method is used to fuse all the data obtained. Third, ray casting algorithm extracts the final vertices of the model, and finally marching cubes algorithm generates the polygonal mesh of the reconstructed face model for displaying. During the process, we also make improvements to speed up the system for human face reconstruction. The system is very convenient for real-world applications, since it can run very quickly and be easily operated.

1. Introduction

3D reconstruction has always been an interesting topic since Microsoft Kinect camera came to use. The depth data of human face can be easily acquired using depth camera. However, it is still difficult to get a perfect whole face model. One probable method to solve this problem is to obtain depth maps from different cameras in different directions simultaneously [1–5]. Another solution is to get data with only one camera in different time. The next step is to do reconstruction with all the data from different directions, finally generating a mesh model of human face. A lot of algorithms for reconstruction have been proposed recently but hardly get perfect results. Moreover, many of them are not so conveniently or easily used in practice.

In this paper, we present a system that realizes fast human face tracking and 3D reconstruction with only one Kinect sensor collecting depth data (Figure 1). The depth data is acquired from the Kinect sensor and is converted to vertex map and normal map, as there is a corresponding relation between the 2D depth map coordinate system and the 3D camera coordinate system. The ICP algorithm is used to track the relative pose between the current data map and the previous one, and then the transformation matrix frame-to-frame can be estimated. Thereby the pose of current camera in the global coordinate can be obtained, which is actually the human face coordinate system. By doing the volumetric integrating using truncated signed distance function, the depth data of all the frames can be integrated into a volume, and the zero-crossing surface can be extracted for the next ICP tracking. The reconstructed 3D model is rendered using the marching cubes algorithm in the reconstructing process, so the user can adjust his face pose to get a better human face model by filling the hole. This system allows users to reconstruct human face model with only one depth camera and incrementally get higher quality 3D human face model by rotating the head.

2. Related Works

The research on real-time 3D reconstruction is a hot topic in computer vision. An accurate and robust 3D surface registration method for In-hand modeling was proposed by [6]. A complete 3D model of an object can be obtained by simply turning around and being scanned by a camera, using iterative closest points (ICPs) algorithm [1] for coarse and

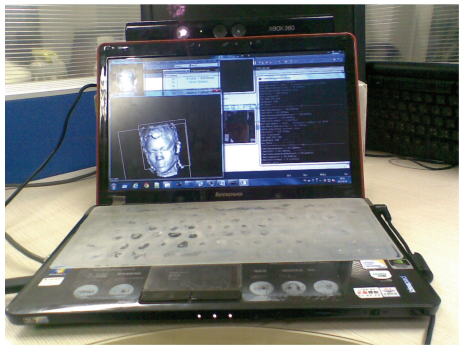


Figure 1: User sits in front of a fixed Kinect sensor and the reconstruction can be done.

fine registration. The authors also proposed a method for detecting registration failure based on both geometric and texture consistences [6]. With the user performing slight head rotation while keeping the facial expression unchanged, the system proposed by Weise et al. [7] aggregated multiple scans into one 3D model of the face. A method for automatically registering multiple 3D data sets without any knowledge of initial pose was proposed by [8]. Jaeggli et al. [9] presented a system which produces complete 3D model using a high-speed acquisition equipment and a registration module. They used pairwise registration as well as multiview refinement to get better results. Azernikov and Fischer [10] proposed volume warping method for surface reconstruction.

KinectFusion project [11–13] presented a system that uses only one moving depth camera to create detailed 3D reconstruction of a complex and arbitrary indoor scenes in real time with GPU. It also enables advanced augmented reality and multitouch on any indoor scene with arbitrary surface geometries. Zollhöfer et al. [14] proposed an algorithm for computing a personalized avatar using a single color image and corresponding depth image. It obtains a high-quality 3D reconstruction model of the face that has one-to-one corresponding geometry and texture with the generic face model.

3. Method and Implementation

The flow chart of our 3D face reconstruction system is shown in Figure 2. The whole system mainly consists of six stages: face detection and segmentation, depth map conversion, face tracking, volumetric integration, recasting, and marching cubes. Each stage will be described in the following sections.

3.1. Face Detection and Segmentation. The Kinect sensor acquires the 640 ∗ 480 color and depth image at 30 Hz. Firstly the face region is detected using a Haar classifier [15]. To get more stable results, we use the frontal-face and profile-face classifiers to detect the face twice.

A more sophisticated extraction is used to ensure that only the face data in consideration is carried out after getting the region of the face. We search the depth image in a window around the central point of the face region to get a valid depth value. Then the depth image will be traversed within the face region and every depth value will be compared with the central depth value. If the depth value changes more than the specific threshold, it will be given an invalid value to distinguish nonface region from face region.

In order to run the algorithm fast, the face region is only detected at the beginning of the algorithm or after resetting. Once a valid bounding rectangle of face is obtained, the face detection phase is omitted and only segmentation task is executed.

3.2. Fast Face Depth Map Conversion. Firstly, a bilateral filter is applied to the raw depth map in order to obtain noise-reduced face depth map and maintain the depth boundaries simultaneously. The filter can be described as follows [12], (raw map is denoted by R_i, and the filtered depth map is denoted by D_i):

$$D_i(\vec{u}) = \frac{1}{W_p} \sum_{\vec{q} \in w} N_\sigma\left(\left\|\vec{u} - \vec{q}\right\|_2\right) N_\sigma\left(\left\|R_i(\vec{u}) - R_i(\vec{q})\right\|_2\right) R_i(\vec{q}), \tag{1}$$

where $\vec{u} = (u, v)^T$ is the depth image pixel and $\vec{q} \in w$ (w is a window to reduce computation complexity), W_p and $N_\sigma(t) = \exp(-t^2\sigma^{-2})$ are normalizing constants.

In a real-time system, the time complexity is an important factor. In the filter previously mentioned, the exponent arithmetic is computationally expensive. Since the distance between the depth image pixels is an integer, we can use a *lookup* table to speed up the computation. Only the pixels within a certain window are considered, so the size of lookup table is not large.

Given the camera intrinsic parameters (f_x, f_y, c_x, c_y, which, respectively, stand for the focal length in x and y axes, the focal point coordinate in x and y axes), the depth map can be converted into the vertex map (denoted by V_i), and corresponding normal vectors for each vertex can be easily computed using neighboring points.

Assuming that (u, v) is a pixel in the raw depth map, and $z = D_i(u, v)$ is the depth of filtered depth map, then the coordinate of V_i can be computed as follows:

$$\begin{aligned} V_i(x) &= \frac{z(u - c_x)}{f_x}, \\ V_i(y) &= \frac{z\left(u - c_y\right)}{f_y}, \\ V_i(z) &= z. \end{aligned} \tag{2}$$

With the vertex data obtained, the normal vectors of each vertex are computed with the following equation [12]:

$$N_i = \left(V_i(u+1, v) - V_i(u, v)\right) \times \left(V_i(u, v+1) - V_i(u, v)\right), \tag{3}$$

and then normalized to the unit length using $N_i/\|N_i\|$.

Assuming that the image being processed has $m \times n$ pixels, and the window of bilateral filter has a size of $w \times w$, the

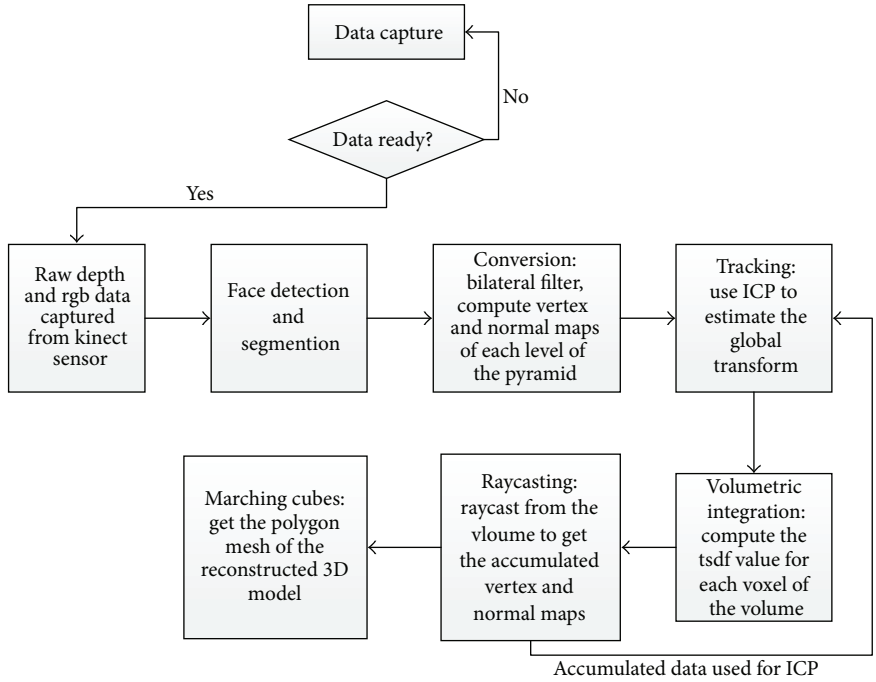


Figure 2: Overall system workflow.

computational complexity of filtering is $O(m \times n \times w^2)$. The Kinect sensor acquires a 640×640 depth image. In the face segmentation stage, a smaller image that contains the face region is extracted, which has a size of 196×196 in our experiment. Besides, $w = 7$ and a lookup table is used to replace the exponent arithmetic. A complexity of $O(m \times n)$ is needed to generate the normal vectors. So the computational complexity of this stage is quite low.

3.3. Face Tracking. The global model of the face is stored as a statue, which means that the vertex map and the normal map do not move and rotate in the coordinate. In the meantime, the sensor is fixed in the front of the user. However, because the user's head rotates (and moves) in the modeling process, the view point of the depth data changes from frame to frame. In order to integrate the history model and the live data, the transformation matrix (a 3D matrix with 6 degrees of freedom) between them is required; otherwise the result will be chaos if we put the data directly. Fortunately, because of the preextraction, the live data do not contain environment and other references. So we come out with the solution that we suppose that the camera is moving around to scan a static head of the user. In this stage, the face depth data is processed frame-to-frame to track the face motion. The Iterative Closest Point (ICP) algorithm [1] is used here to compute the transformation matrix [12].

In ICP algorithm, each iteration could be divided into three stages: first, corresponding point pairs are chosen within the two frames; then a point-to-plane Euclid distance error metric is used to calculate an optimal solution which minimizes the error metric; and finally, the previous frame is transformed using the matrix obtained in the previous stage, preparing for the next iteration. Solution becomes more accurate after each cycle, and an optimal transform matrix can be obtained after finishing ICP algorithm.

In the first stage, the corresponding points between the current frame and the previous frame should be found firstly. In this paper, points in the current frame are projected to the previous one to obtain the corresponding points. Given the previous global camera pose T_{i-1} and the estimated current pose T_i (which is initialized with T_{i-1} before the iteration and is updated with an incremental transform calculated per iteration of ICP), we then transform the current vertex map into the global coordinate, using $V_i^g = T_i V_i(\vec{u})$. (Here, the human face coordinate system is regarded as the global system. Since our target is to obtain a complete face model, so that the data is required in face coordinate system.) The pixel corresponding to the global position V_i^g are required here, so we transform V_i^g into the previous camera coordinate system to get $V_{i-1}^c = T_{i-1}^{-1} V_i^g$ and then project it into the image plane to get the pixel p:

$$p(u) = \frac{V_{i-1}^c(x) \times f_x}{V_{i-1}^c(z)} + c_x,$$
$$p(v) = \frac{V_{i-1}^c(y) \times f_y}{V_{i-1}^c(z)} + c_y. \tag{4}$$

We look up the previous global vertex map V_{i-1}^g and global normal map N_{i-1}^g in the pixel p; obviously $V_{i-1}(\vec{p})$ and $V_i(\vec{u})$ are the correspondences.

Also thresholds of Euclidean distance and angles should be set between them, to reject outliers.

Given these sets of corresponding oriented points, the next step is to minimize the point-to-plane error metric E, to get the transformation matrix T [16]. Here, the metric E stands for the sum of squared distances between the current points and the tangent plane of the corresponding points:

$$E = \sum_{\vec{u}} \left\| \left(TV_i(\vec{u}) - V_{i-1}^{g}(\vec{p}) \right) N_{i-1}^{g}(\vec{p}) \right\|_2. \tag{5}$$

As there is only an incremental transformation between frames, the rigid transform matrix can be written as follows:

$$T = \left[R \mid \vec{t} \right] = \begin{pmatrix} 1 & \alpha & -\gamma & t_x \\ -\alpha & 1 & \beta & t_y \\ \gamma & -\beta & 1 & t_z \end{pmatrix}. \tag{6}$$

In each cycle, an incremental transformation T_{inc}^{z} (z represents the current iteration number) that minimizes the error metric can be estimated. The desired global transformation matrix can be simply updated by $T_g^z = T_{\text{inc}}^z \cdot T_g^{z-1}$.

We update the current global frame vertex estimates using the global transformation T_g^{z-1} computed in the previous iteration, $V_i^g(\vec{u}) = T_g^{z-1} V_i(\vec{u})$. The increment transformation can also be written as follow:

$$\begin{gathered} \vec{x} = \left(\beta, \gamma, \alpha, t_x, t_y, t_z \right) \in \mathfrak{R}^6, \\ T_g^z V_i(\vec{u}) = R^z V_i^g(\vec{u}) + \vec{t}^{\,z} = G(\vec{u})\, \vec{x} + V_i^g(\vec{u}). \end{gathered} \tag{7}$$

Assuming that $V_i^g(\vec{u}) = (x, y, z)^T$, G can be represented as follows:

$$G(\vec{u}) = \begin{pmatrix} 0 & -z & y & 1 & 0 & 0 \\ z & 0 & -x & 0 & 1 & 0 \\ -y & x & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{8}$$

By solving the following expression:

$$\begin{gathered} \min_{x \in \mathfrak{R}^6} \sum_{\vec{u}} \| E \|_2 c, \\ E = N_{i-1}^{g}(\vec{p})^T \cdot \left(G(\vec{u})\, \vec{x} + V_i^g(\vec{u}) - V_{i-1}^{g}(\vec{p}) \right). \end{gathered} \tag{9}$$

An optimal transformation matrix needed can be computed.

Then with the equations previously mentioned, the expression can be expressed as follows:

$$\sum_{\vec{u}} \left(A^T A \right) \vec{x} = \sum_{\vec{u}} A^T b, \tag{10}$$

where $A^T = G^T(\vec{u}) N_{i-1}^{g}(\vec{p}) \in \mathfrak{R}^{6\times 1}$ and $b = N_{i-1}^{g}(\vec{p})^T (V_{i-1}^{g}(\vec{p}) - V_i^g(\vec{u})) \in \mathfrak{R}^{1\times 1}$, and we can easily compute the vector $\vec{x}$ using a Cholesky decomposition.

After the incremental matrix is obtained, we enter the next iteration, where the operations previously mentioned are used again. And after all the iterations, we can get the final camera pose $T_g \leftarrow T_g^{z\ \max}$.

A projecting strategy is used so that it takes $O(1)$ computational complexity for a point to find its corresponding point. To get the transformation matrix between two frames, approximately a computational complexity of $O(m \times n)$ is acquired.

3.4. Small Size Face Volumetric Integration. Using the transformation matrix obtained in the last step, the depth map data can be converted to point cloud in global coordinate. Data of this form is hard to integrate to form a face model. Here we convert them into a volumetric representation [17]. A 3D volume of specific resolution that corresponds to the physical space is divided into a 3D grid of voxels. Truncated signed distance function (TSDF) is used here to convert the 3D vertices into voxels of the volume [12].

In the volumetric integration phase, each voxel in the volume is traversed and the corresponding TSDF value is updated using weighted average strategy. For a voxel (x, y, z) in the volume, we first convert it into the global 3D position v_g:

$$\begin{aligned} v_g(x) &= (x + 0.5f) \times \text{cell\_size}.x, \\ v_g(y) &= (y + 0.5f) \times \text{cell\_size}.y, \\ v_g(z) &= (z + 0.5f) \times \text{cell\_size}.z, \end{aligned} \tag{11}$$

where cell_size represents the size of the cell in the volume:

$$\text{cell\_size}.x = \frac{\text{VOLUME\_SIZE\_}X}{\text{VOLUME\_}X}, \tag{12}$$

where VOLUME_X indicates how many cells are there in x axis of the volume and VOLUME_SIZE_X indicates the corresponding actual length.

Subsequently, the global ordinate v_g is transformed into the camera ordinate v, and the vertex v is projected into the image plane to get the corresponding pixel p.

Assuming that the translation vector of the global camera transformation is denoted as t_i, the distance between voxel (x, y, z) of the volume and the original point of the camera ordinates system can be calculated as $\|v_g - t_i\|$. Since we have got the corresponding pixel p, we can get the actual depth measurement $D_i(p)$. It should be pointed out that $D_i(p)$ is not equal to the vertex map $V_i(p)$, since the former represents the distance between the original point and the specific point, while the latter only represents the z value, so a conversion is necessary to get $D_i(p)$.

The SDF value of the voxel can be computed using $\text{SDF}_i = \|v_g - t_i\| - D_i(p)$. This is normalized to a TSDF using the truncating strategy. The TSDF value of the voxel is updated using a simple running weighted average:

$$\text{tsdf}^{\text{avg}} = \frac{\text{tsdf}_{i-1} \cdot w_{i-1} + \text{tsdf}_i \cdot w_i}{w_{i-1} + w_i}. \tag{13}$$

Actually, in practice, we simply let $w_i = 1$ and can achieve good results.

Given that the goal of our work is fast tracking and reconstruction, time complexity must be considered. In volumetric integration phase, there are VOLUME_X × VOLUME_Y × VOLUME_Z voxels that should be traversed, so the volume size cannot be too large for higher frame rate. In case of face reconstruction, the useful global size is approximately $0.3 \times 0.3 \times 0.3$ m^3. However, in the volumetric integrating algorithm described previously, the $z = 0$ voxel of the volume

lies in the original position of the camera ordinate system. In other words, if the distance between the face and the Kinect sensor is 0.7 m, VOLUME_SIZE_Z cannot be less than 0.7 to ensure valid integration. And at the same time, all the $0 < z < 0.7$ voxels are not used. Consequently, to ensure enough volume resolution, VOLUME_Z cannot be too small, which could result in higher time complexity and do too much useless work.

Thus we introduce an offset to the volume. We move the volume along the z axis a distance of offset as the $z = 0$ plane can get close to but cannot reach the mesh of the face. Then the conversion from voxel into global position should be modified as

$$v_g(z) = (z + 0.5f) \times \text{cell\_size}.z + \text{offset}. \quad (14)$$

Some little modification should also be made in the other part of the algorithm.

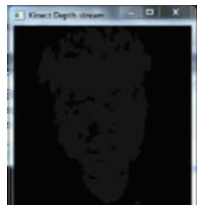
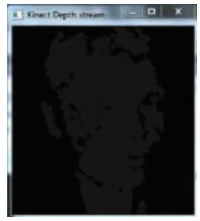
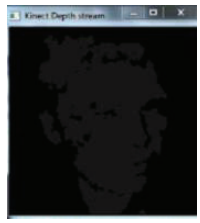
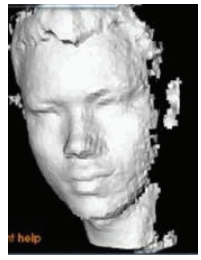
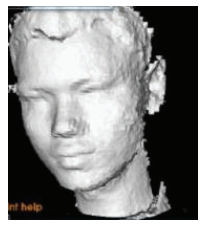
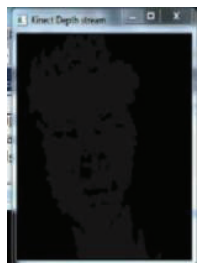
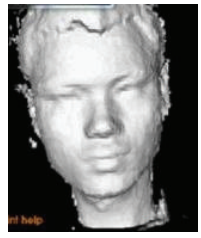
In this stage, a computational complexity of $O(n^3)$ (n is the length of the volume) is needed to update all the voxels in the volume. In our experiment, we use $64 \times 64 \times 64$ ($n = 64$ volume and $0.3 \times 0.3 \times 0.3$ volume size along with an offset of 0.6 m and get rather good fast reconstruction results.

3.5. Ray Casting. The ray casting algorithm [18] applied here is to generate views of implicit surface for rendering and tracking. For each pixel (x, y) of the output image, a single ray is emitted from the original point of the camera coordinate system and goes through the point (x, y) of the image plane. With the direction of the ray, we can extract the surface position by traversing along the ray. And the surface intersection point can be easily obtained using linear interpolation. Then we can easily obtain the normal map with TSDF. There are two contributions of the ray casting algorithm, one is the ability to view the implicit surface of the reconstructed 3D model and the other is to generate higher quality data for ICP camera tracking. When rendered to screen, the noise, shadows, and holes will be much less than the raw depth data.

We need to traverse all the voxels in the volume to extract the zero-crossing surface. Therefore a computational complexity of $O(n^3)$ (n is the length of the volume) is acquired.

3.6. Marching Cubes. In our work, we use marching cubes algorithm [19, 20] to obtain the mesh of the reconstructed 3D model. Each voxel of the volume is traversed and an index is created to a precalculated array of 256 possible polygon configurations within the cube, by treating each of the 8 scalar values as a bit in an 8-bit integer. If the scalar's value is higher than the iso value (it is inside the surface) then the appropriate bit is set to one, while if it is lower (outside), it is set to zero. The final value after all 8 scalars are checked is the actual index to the polygon indices array. Finally each vertex of the generated polygons is placed on the appropriate position along the cube's edge by linearly interpolating the two scalar values that are connected by that edge.

Table 1: Results of human face model reconstruction.

Raw depth	Reconstructed mesh	Time
		Shortly after the start of the reconstruction
		Begin to scan the right side of the face
		The right side of the face has almost been reconstructed
		Begin to scan the left side of the face
		The left side of the face has almost been reconstructed
		Reconstruction done

In marching cubes algorithm, the computational complexity that is approximately $O(n^3)$ (n is the length of the volume) is acquired.

4. Results

We test our 3D reconstruction system on a computer with 3.2 GHz CPU and 4 GB memory. We set the volume resolution to $64 \times 64 \times 64$ and the volumetric size to $0.3 \times 0.3 \times 0.3\ m^3$ with an offset distance of 0.4 m. Note that running the reconstruction algorithm for one new frame only costs about 180ms, which is quite acceptable in practice.

The results of our 3D reconstruction system are shown in Table 1. As shown in the table, the 3D face model keeps being refined, while the user's head rotates in order to let the Kinect scan the whole face.

We can find that the reconstruction result is very good, and is much smoother than the raw depth data. And the reconstruction speed is also very acceptable.

5. Conclusions

After depth cameras like Kinect sensor appear, users can easily obtain the depth data of an object. 3D reconstruction, especially human face reconstruction, has always been a challenging problem. In this paper, we represent a novel way to create a 3D face reconstruction model. Just sitting in front of the Kinect camera and rotating his head, the user can get a perfect human face model. We use a volumetric integrating strategy to fuse all the data, so the reconstructed face model becomes more and more clear.

We contribute the method to fast human face 3D reconstruction. Our efforts to speed up the system are threefold. First, we decrease the frequency of detecting face, by only detecting when the shift of the face exceeds a specific threshold. Second, we use a lookup table to replace the computationally expensive exponent arithmetic and try hard to reduce repeated computation. Third, we introduce some variances to the volumetric integration algorithm to use less voxels while keeping the good resolution. Using the methods previously mentioned, we get a well-performed face 3D reconstruction system.

We will focus on larger object such as full body 3D reconstruction, and add color information to the model to make the visualization better in the future work.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (NSFC) under the Project 61175034/F030410.

References

[1] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[2] C. Shengyong, W. Yuehui, and C. Carlo, "Key issues in modeling of complex 3D structures from video sequences," *Mathematical Problems in Engineering*, vol. 2012, Article ID 856523, 17 pages, 2012.

[3] S. Y. Chen and Y. F. Li, "Vision sensor planning for 3-D model acquisition," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 35, no. 5, pp. 894–904, 2005.

[4] C. Carlo, C. Shengyong, and A. Gani, "Information and modeling in complexity," *Mathematical Problems in Engineering*, vol. 2012, Article ID 868413, 4 pages, 2012.

[5] S. Y. Chen, Y. F. Li, Q. Guan, and G. Xiao, "Real-time three-dimensional surface measurement by color encoded light projection," *Applied Physics Letters*, vol. 89, no. 11, Article ID 111108, 2006.

[6] T. Weise, B. Leibe, and L. Van Gool, "Accurate and robust registration for in-hand modeling," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, June 2008.

[7] T. Weise, S. Bouaziz, and H. Li, "Realtime performance-based facial animation," in *Proceedings of the 38th Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH '11)*, Vancouver, Canada, August 2011.

[8] F. Huber and M. Hebert, "Fully automatic registration of multiple 3D data sets," in *Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (CVBVS '01)*, Kauai, Hawaii, USA, December 2001.

[9] T. Jaeggli, T. Koninckx, and L. V. Gool, "Online 3d acquisition and model integration," in *Proceedings of IEEE International Workshop on Projector-Camera Systems (PROCAMS '03)*, Nice, France, 2003.

[10] S. Azernikov and A. Fischer, "A new volume warping method for surface reconstruction," *Virtual and Physical Prototyping*, vol. 1, no. 2, pp. 65–71, 2006.

[11] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli et al., "Kinectfusion: real-time 3D reconstruction and interaction using a moving depth camera," in *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*, pp. 559–568, ACM, New York, NY, USA, 2011.

[12] R. A. Newcombe, S. Izadi, O. Hilliges et al., *Kinect Fusion: Real-Time Dense Surface Mapping and Tracking*, IEEE ISMAR, 2011.

[13] L. Yong-Wan, L. Hyuk-Zae, Y. Na-Eun et al., "3-D reconstruction using the kinect sensor and its application to a visualization system," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '12)*, pp. 3361–3366, 2012.

[14] M. Zollhöfer, M. Martinek, G. Greiner, M. Stamminger, and J. Süßmuth, "Automatic reconstruction of personalized avatars from 3D face scans," *Computer Animation and Virtual Worlds*, vol. 22, no. 2-3, pp. 195–202, 2011.

[15] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," Tech. Rep., Microprocessor Research Lab, 2002.

[16] K. Low, "Linear least-squares optimization for point-to-plane ICP surface registration," Tech. Rep. TR04-004, University of North Carolina, 2004.

[17] B. Curless and M. Levoy, "Volumetric method for building complex models from range images," in *Proceedings of the 1996 Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, pp. 303–312, August 1996.

[18] S. D. Roth, "Ray casting for modeling solids," *Computer Graphicsand Image Processing*, vol. 18, no. 2, pp. 109–144, 1982.

[19] W. E. Lorensen and H. E. Cline, "Marching cubes: a high resolution 3D surface construction algorithm," *ACM Transactions on Graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[20] T. S. Newman and H. Yi, "A survey of the marching cubes algorithm," *Computers and Graphics*, vol. 30, no. 5, pp. 854–879, 2006.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 624957, 11 pages
http://dx.doi.org/10.1155/2013/624957

Research Article

Construction of Bivariate Nonseparable Compactly Supported Orthogonal Wavelets

Jinsong Leng,[1] Tingzhu Huang,[1] and Carlo Cattani[2]

[1] *School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China*
[2] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

Correspondence should be addressed to Jinsong Leng; jslengjs@gmail.com

Received 7 January 2013; Accepted 24 January 2013

Academic Editor: Sheng-Yong Chen

A method for constructing bivariate nonseparable compactly supported orthogonal scaling functions, and the corresponding wavelets, using the dilation matrix $M := 2^n\mathbb{I} = 2^n \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, $(d = \det M = 2^{2n} \geq 4, n \in \mathbb{N})$ is given. The accuracy and smoothness of the scaling functions are studied, thus showing that they have the same accuracy order as the univariate Daubechies low-pass filter $m_0(\omega)$, to be used in our method. There follows that the wavelets can be made arbitrarily smooth by properly choosing the accuracy parameter r.

1. Introduction

Wavelet analysis (see [1, 2]) has become a powerful tool in neural networks, harmonic analysis, numerical analysis, and signal processing, especially in the area of image compression [3]. The wavelet transform is a simple and practical mathematical tool that cuts up data or functions into different frequency components and then studies each component with a resolution matched to its scale. The main feature of the wavelet transform is to hierarchically decompose general functions, as a signal or a process, into a set of approximation functions with different scales [1, 2, 4, 5].

Wavelet theory is closely related to subband coding, and it provides a functional space structure for subband coding, often leading to a better understanding of signals.

Univariate wavelets have found successful applications in signal processing [5–7]. However, in order to apply wavelet methods to data processing and to analyze multidimensional signals, such as image processing, we have to construct multivariate wavelets. The most commonly used method to define multivariate wavelets is the tensor product of univariate wavelets [4, 8–10]. This construction leads to a separable wavelet function having, as a major drawback, the horizontal and vertical (obliged) directions, while usually in imaging process the information contents spread along all directions, in a nonisotropic way. Much efforts have been spent for constructing multivariate nonseparable compactly supported orthogonal wavelets (see, e.g., [11–26]); however, the generalization to higher dimensions is not a simple task, so that a general simple methodology is still missing. However, it is possible to characterize the construction of bivariate scaling functions by the two scale difference equation, which in 1 dimension is

$$\varphi(x) = \sum_{k\in\mathbb{Z}} c_k \varphi(2x-k), \quad (x \in \mathbb{R}) \tag{1}$$

and in 2 dimensions

$$\varphi(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \varphi(M\mathbf{x}-\mathbf{k}), \quad \left(\mathbf{x} \in \mathbb{R}^2\right) \tag{2}$$

depends on the coefficients of a suitable second-order dilation matrix M. Usually [9, 11–13] the determinant of this matrix is 2. For example, Ayache [11, 12] proposed two methods for constructing bivariate nonseparable compactly supported orthonormal wavelets. The wavelets constructed using one of the methods are called semiseparable wavelets. Belogay and Wang [13] constructed bivariate nonseparable compactly supported orthonormal wavelets using the dilation matrix $M = \left[\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right]$. Belogay and Wang gave the sufficient condition for the low-pass filters constructed by their method to generate nonseparable compactly supported orthogonal

accurate scaling function with arbitrarily high smoothness. However, for proving the accuracy these authors used the matrix $M^2 = \pm 2\mathbb{I}$. Moreover, their method has been modified by Lai [18] to construct nonseparable wavelets using the dilation matrix $\left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right]$ with determinant equal to 4. He and Lai [16] constructed many examples of nonseparable orthogonal wavelets. Lai and Roach [19] provided a method to construct bivariate nonseparable compactly supported wavelets by using the method of symmetry. Stanhill and Zeevi [24, 25] provided other methods for constructing bivariate nonseparable orthogonal wavelets.

By combining the Belogay and Wang method, the Lai method, and the Ayache method, we present another method for constructing bivariate nonseparable orthogonal wavelets using the dilation matrix $M = 2\mathbb{I} = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right]$. Our method holds true also for dilation matrix of the form

$$M = 2^n \mathbb{I} = 2^n \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (n \in \mathbb{N}) \tag{3}$$

being $\mathbb{I} := \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ the identity matrix, so that

$$d = \det M = 2^{2n} \quad (n \in \mathbb{N}). \tag{4}$$

It follows that the two-scale difference equation simplifies into

$$\varphi(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^2} \varphi\left(2^n \mathbf{x} - \mathbf{k}\right) \quad \left(\mathbf{x} \in \mathbb{R}^2,\ n \in \mathbb{N}\right) \tag{5}$$

which is more alike the corresponding equation in 1 dimension. The minor drawback is that for the multivariate scaling function there exist $d - 1$ corresponding wavelets (see, e.g., [9]), so that for a matrix with determinant 4 there exist, together with a scaling function, 3 wavelet functions (instead of 1 wavelet when $d = 2$). However, due to the special form of the dilation matrix, we can take into account the results of [11–13] and show that the scaling function constructed by our method has the same accuracy as the univariate Daubechies low-pass filter to be used in our method. Following [13] we will also show that the smoothness of our bivariate scaling function can be manipulated by properly choosing the accuracy parameter r.

The paper is organized as follows. In Section 2, we introduce some notations and briefly recall some basic elements of the theory of bivariate multiresolution analysis and bivariate orthogonal wavelets. In Section 3, we provide a method for constructing a class of bivariate nonseparable compactly supported orthogonal filter banks. Then we study the accuracy and the smoothness of the scaling function constructed by our method. Finally, we give an example of the bivariate nonseparable compactly supported scaling function.

2. Preliminaries

Let $\mathbb{Z}$ and $\mathbb{R}$ be the sets of all integers and real numbers, respectively. Denote

$$L^2\left(\mathbb{R}^2\right) := \left\{ f(\mathbf{x}) : \int_{\mathbb{R}^2} \left|f(\mathbf{x})\right|^2 d\mathbf{x} < \infty, \mathbf{x} \in \mathbb{R}^2 \right\},$$
$$\ell^2\left(\mathbb{Z}^2\right) = \left\{ \{C_{\mathbf{n}}\}_{\mathbf{n} \in \mathbb{Z}^2} : \sum_{\mathbf{n} \in \mathbb{Z}^2} \left|C_{\mathbf{n}}\right|^2 < \infty \right\}. \tag{6}$$

The Fourier transform of $f(\mathbf{x})$ is defined by

$$\widehat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} f(\mathbf{x}) e^{-i\boldsymbol{\omega}\cdot\mathbf{x}} d\mathbf{x}, \tag{7}$$

where $\boldsymbol{\omega}$, $\mathbf{x} \in \mathbb{R}^2$, $i = \sqrt{-1}$, and $\boldsymbol{\omega} \cdot \mathbf{x} := \omega_1 x_1 + \omega_2 x_2$ is the dot product.

For $f(\mathbf{x})$, $g(\mathbf{x}) \in L^2(\mathbb{R}^2)$, their inner product is defined by

$$\left\langle f(\mathbf{x}), g(\mathbf{x}) \right\rangle := \int_{\mathbb{R}^2} f(\mathbf{x})\, g(\mathbf{x})\, d\mathbf{x}, \tag{8}$$

which is explicitly

$$\left\langle f(x_1, x_2), g(x_1, x_2) \right\rangle := \int_{\mathbb{R}^2} f(x_1, x_2)\, g(x_1, x_2)\, dx_1 dx_2. \tag{9}$$

A function $f(\mathbf{x}) \in L^2(\mathbb{R}^2)$ is called orthogonal if the set of its translate $\{f(\mathbf{x} - \mathbf{k})\}_{\mathbf{k} \in \mathbb{Z}^2}$ is orthogonal; that is,

$$\left\langle f(\mathbf{x} - \mathbf{k}), f(\mathbf{x} - \mathbf{h}) \right\rangle = \delta_{\mathbf{kh}}, \tag{10}$$

where $\mathbf{k}$, $\mathbf{h} \in \mathbb{Z}^2$ and $\delta_{\mathbf{kh}}$ is the Kronecker notation defined by $\delta_{\mathbf{kh}} = 1$ if $\mathbf{k} = \mathbf{h}$ and $\delta_{\mathbf{kh}} = 0$ if $\mathbf{k} \neq \mathbf{h}$, which is explicitly

$$\left\langle f(x_1 - k_1, x_2 - k_2), f(x_1 - h_1, x_2 - h_2) \right\rangle = \delta_{k_1 h_1} \delta_{k_2 h_2}. \tag{11}$$

The whole construction of wavelets is based on the multiresolution analysis (MRA) and its corresponding axioms, which can be easily extended to the bivariate multiresolution analysis as follows. Let $\varphi(\mathbf{x}) \in L^2(\mathbb{R}^2)$ satisfy the following refinement equation:

$$\varphi(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^2} p_{\mathbf{k}} \varphi(2\mathbf{x} - \mathbf{k}), \tag{12}$$

where $\{p_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^2}$ is a real constant sequence.

We assume that $p_{\mathbf{k}} \neq 0$ for only finitely many $\mathbf{k} \in \mathbb{Z}^2$ (ensuring that $\varphi(\mathbf{x})$ has compact support).

Define a closed subspace $V_j \subset L^2(\mathbb{R}^2)$ by

$$V_j = \operatorname{clos}_{L^2(\mathbb{R}^2)}\left(\operatorname{span}\left\{2^j \varphi\left(2^j \mathbf{x} - \mathbf{k}\right) : \mathbf{k} \in \mathbb{Z}^2\right\}\right), \quad j \in \mathbb{Z}. \tag{13}$$

Definition 1. One says that $\varphi(\mathbf{x})$ in (12) generates a bivariate multiresolution analysis (MRA) $\{V_j\}_{j \in \mathbb{Z}}$ of $L^2(\mathbb{R}^2)$ if the sequence $\{V_j\}_{j \in \mathbb{Z}}$ defined by (13) satisfies

(i) $\cdots \subset V_{-1} \subset V_0 \subset V_1 \cdots$;

(ii) $\bigcup_{j \in \mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}^2)$ and $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$, where 0 is the zero function;

(iii) $f(\mathbf{x}) \in V_j \Leftrightarrow f(2\mathbf{x}) \in V_{j+1}$, $j \in \mathbb{Z}$;

(iv) $\{\varphi(\mathbf{x} - \mathbf{k})\}_{\mathbf{k} \in \mathbb{Z}^2}$ is an orthogonal basis of V_0.

The function $\varphi(\mathbf{x})$ is called an orthogonal scaling function of the bivariate multiresolution analysis $\{V_j\}_{j \in \mathbb{Z}}$.

By taking the Fourier transform for both sides of (12), we have

$$\widehat{\varphi}(\xi,\eta) = P_0(z_1,z_2)\,\widehat{\varphi}\left(\frac{\xi}{2},\frac{\eta}{2}\right), \quad \xi,\eta\in\mathbb{R}, \tag{14}$$

where

$$P_0(z_1,z_2) = \sum_{(k,l)\in\mathbb{Z}^2} p_{k,l} z_1^k z_2^l, \quad z_1 := e^{i\xi/2},\ z_2 := e^{i\eta/2}, \tag{15}$$

is a Laurent polynomial satisfying $P_0(1,1) = 1$.

The orthogonality condition (10) implies that

$$\sum_{k_1,k_2\in\{0,1\}\times\{0,1\}} \left|P_0\left((-1)^{k_1}z_1,(-1)^{k_2}z_2\right)\right|^2 = 1. \tag{16}$$

The function $P_0(z_1,z_2)$ is called a low-pass filter (refinement mask). By (14) and $P_0(1,1) = 1$, we get

$$\widehat{\varphi}(\xi,\eta) = \prod_{k=1}^{\infty} P_0\left(e^{i\xi/2^k}, e^{i\eta/2^k}\right), \tag{17}$$

so that the scaling function can be defined by a suitable choice of the low-pass filter.

Let $C(\xi,\eta) = P_0(z_1,z_2)$; the filter $P_0(z_1,z_2)$ satisfies Cohen's criterion [26] if there exists a compact fundamental domain Ω of the lattice $2\pi\mathbb{Z}^2$ with the property

$$C\left(\frac{\xi}{2^j},\frac{\eta}{2^j}\right) \neq 0, \quad \forall j\geq 1,\ \forall(\xi,\eta)\in\Omega. \tag{18}$$

Let us note that (16) is only a necessary condition for $\varphi(\mathbf{x})$ to be orthogonal. The condition (16) becomes also a sufficient condition if the filter $P_0(z_1,z_2)$ satisfies Cohen's criterion (18).

Let us now assume that there exist three closed subspaces $W_{0,i}\subset L^2(\mathbb{R}^2)$, $i = 1,2,3$, such that

$$V_1 = V_0\oplus W_{0,1}\oplus W_{0,2}\oplus W_{0,3}, \tag{19}$$

where $\oplus$ denotes orthogonal direct sum of spaces. We say that the compactly supported functions $\psi_i(\mathbf{x})$, $i = 1,2,3$ are orthogonal wavelets if $\{\psi_i(\mathbf{x}-\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^2}$ form orthogonal bases of $W_{0,i}$, $i = 1,2,3$; that is,

$$W_{0,i} = \operatorname{clos}_{L^2(\mathbb{R}^2)}\left(\operatorname{span}\left\{\psi_i(\mathbf{x}-\mathbf{k}) : \mathbf{k}\in\mathbb{Z}^2\right\}\right), \quad i = 1,2,3,$$
$$\left\langle\psi_i(\mathbf{x}-\mathbf{k}),\psi_j(\mathbf{x}-\mathbf{l})\right\rangle = \delta_{i,j}\delta_{\mathbf{k},l}, \tag{20}$$

where $i,j = 1,2,3$ and $\mathbf{k}, l\in\mathbb{Z}^2$.

If there exist three Laurent polynomials $P_i(z_1,z_2)$, $i = 1,2,3$, such that

$$Q = \begin{bmatrix} P_0(z_1,z_2) & P_0(-z_1,z_2) & P_0(z_1,-z_2) & P_0(-z_1,-z_2)\\ P_1(z_1,z_2) & P_1(-z_1,z_2) & P_1(z_1,-z_2) & P_1(-z_1,-z_2)\\ P_2(z_1,z_2) & P_2(-z_1,z_2) & P_2(z_1,-z_2) & P_2(-z_1,-z_2)\\ P_3(z_1,z_2) & P_3(-z_1,z_2) & P_3(z_1,-z_2) & P_3(-z_1,-z_2)\end{bmatrix} \tag{21}$$

is unitary, that is, $\det Q = 1$, then the Fourier transforms of the functions $\psi_i(\mathbf{x})$, $i = 1,2,3$, can be given by

$$\widehat{\psi}_i(\xi,\eta) = P_i\left(e^{i\xi/2},e^{i\eta/2}\right)\widehat{\varphi}\left(\frac{\xi}{2},\frac{\eta}{2}\right), \tag{22}$$

and the three Laurent polynomials $P_i(z_1,z_2)$, $i = 1,2,3$, are called the high-pass filters (wavelet masks). The whole set of functions $\{P_i(z_1,z_2)\}_{0\leq i\leq 3}$ is called a filter bank.

By (17) and (22), the construction of bivariate compactly supported orthogonal scaling function and the corresponding wavelets is reduced to the construction of filter bank $\{P_i(z_1,z_2)\}_{0\leq i\leq 3}$ which satisfies (16), (18), and (21).

3. Construction of Bivariate Nonseparable Orthogonal Wavelets

3.1. Construction of the Filter Banks. According to Section 2, for constructing bivariate orthogonal scaling function and the corresponding wavelets, we mainly need to define the bivariate orthogonal filter bank $\{P_i(z_1,z_2)\}_{0\leq i\leq 3}$ satisfying (16), (18), and (21).

Let us shortly recall the Belogay and Wang method, the Lai method, and the Ayache method for constructing bivariate nonseparable compactly supported orthogonal wavelets and their corresponding filter banks.

Belogay and Wang [13] constructed the bivariate nonseparable orthogonal scaling functions using the dilation matrix $\left[\begin{smallmatrix}0&2\\1&0\end{smallmatrix}\right]$ with special coefficients: the coefficients are aligned along two adjacent rows. The corresponding low-pass filters have the form

$$P_0(z_1,z_2) = A(z_1) + z_2B(z_1), \tag{23}$$

where $A(z)$, $B(z)$ are two univariate polynomials of $z = e^{i\xi}$ satisfying some conditions (see [13]). The high-pass filter coefficients $d_{\mathbf{n}}$ are given by

$$d_{\mathbf{n}} = (-1)^{n_1}c_{\mathbf{e}-\mathbf{n}}, \tag{24}$$

where $\mathbf{n} = (n_1,n_2)\in\mathbb{Z}^2$, $\mathbf{e} = (1,0)$, and $c_{\mathbf{n}}$ are low-pass filter coefficients.

Lai [18] modified the Belogay and Wang method to construct nonseparable wavelets using the dilation matrix $\left[\begin{smallmatrix}2&0\\0&2\end{smallmatrix}\right]$. According to the Lai method, their filter bank has the form

$$\begin{aligned}
P_0(z_1,z_2) &= \left[A(z_1)+z_2B(z_1)\right]\left[A(z_2)+z_1^2B(z_2)\right],\\
P_1(z_1,z_2) &= z_2\left[A(z_1)+z_2B(z_1)\right]\left[A\left(-z_2^{-1}\right)+z_1^2B\left(-z_2^{-1}\right)\right],\\
P_2(z_1,z_2) &= z_1\left[A\left(-z_1^{-1}\right)+z_2^{-1}B\left(-z_1^{-1}\right)\right]\\
&\quad\times\left[A(z_2)+z_1^2B(z_2)\right],\\
P_3(z_1,z_2) &= z_1z_2\left[A\left(-z_1^{-1}\right)+z_2^{-1}B\left(-z_1^{-1}\right)\right]\\
&\quad\times\left[A\left(-z_2^{-1}\right)+z_1^2B\left(-z_2^{-1}\right)\right],
\end{aligned} \tag{25}$$

where $A(z)$, $B(z)$ are two univariate polynomials of $z = e^{i\xi}$ satisfying some conditions, $z_1 = e^{i\xi}$, and $z_2 = e^{i\eta}$ (see [18]).

Ayache [11, 12] proposed two methods for constructing bivariate compactly supported orthogonal wavelets using the dilation matrix $\left[\begin{smallmatrix}2&0\\0&2\end{smallmatrix}\right]$. Let $\alpha(x)$, $\beta(x)$ be two univariate trigonometric polynomials, such that

$$\alpha(0)=1, \qquad |\alpha(x)|^2+|\beta(x)|^2=1. \tag{26}$$

Then according to one of the two methods, the nonseparable filter banks have the form

$$\begin{aligned}
P_0(z_1,z_2)&=u_0(z_1)\left[\alpha(2\eta)\,v_0(z_2)+\beta(2\eta)\,v_1(z_2)\,z_1^{-2}\right],\\
P_1(z_1,z_2)&=u_0(z_1)\left[\overline{\beta(2\eta)}v_0(z_2)-\overline{\alpha(2\eta)}v_1(z_2)\,z_1^{-2}\right],\\
P_2(z_1,z_2)&=u_1(z_1)\left[\alpha(2\eta)\,v_0(z_2)+\beta(2\eta)\,v_1(z_2)\,z_1^{-2}\right],\\
P_3(z_1,z_2)&=u_1(z_1)\left[\overline{\beta(2\eta)}v_0(z_2)-\overline{\alpha(2\eta)}v_1(z_2)\,z_1^{-2}\right],
\end{aligned} \tag{27}$$

where $\{u_k(z)\}_{0\le k\le 1}$, $\{v_k(z)\}_{0\le k\le 1}$ are two given univariate filter banks, $z_1=e^{i\xi}$, and $z_2=e^{i\eta}$.

By combining the previous three methods, we propose, in the following, a method based on a suitable choice of the univariate filter bank, together with the basic ideas of Belogay-Wang approach. In particular, we use the univariate Daubechies filter bank to define the bivariate nonseparable filter bank as follows.

Let r be a nonnegative integer, and let $\{m_i(z)\}_{0\le i\le 1}$ be the univariate Daubechies filter bank [2, 27] with accuracy $r+1$; that is,

$$m_0(z)=\left(\frac{1+z}{2}\right)^{r+1}H_{r+1}(z), \qquad m_1(z)=e^{-i\xi}m_0\left(-z^{-1}\right), \tag{28}$$

where

$$\left|H_{r+1}(z)\right|^2:=\sum_{k=0}^{r}\binom{r+k}{k}\left(\sin^2\frac{\xi}{2}\right)^k \tag{29}$$

and $z=e^{i\xi}$. Let $A(z), B(z)$ be two univariate polynomials of $z=e^{i\xi}$ with real coefficients, and let $A(1)\neq 0$, $B(0)\neq 0$. We define four bivariate polynomials of $z_1=e^{i\xi}$, $z_2=e^{i\eta}$ which are the forms of the filter banks of this paper, as follows:

$$P_0(z_1,z_2):=m_0(z_2)\left[A(z_1)+z_2^2B(z_1)\right], \tag{30}$$

$$P_1(z_1,z_2):=z_1m_0(z_2)\left[B\left(-z_1^{-1}\right)+z_2^2A\left(-z_1^{-1}\right)\right], \tag{31}$$

$$P_2(z_1,z_2):=m_1(z_2)\left[A(z_1)+z_2^2B(z_1)\right], \tag{32}$$

$$P_3(z_1,z_2):=z_1m_1(z_2)\left[B\left(-z_1^{-1}\right)+z_2^2A\left(-z_1^{-1}\right)\right]. \tag{33}$$

Let r be a nonnegative integer, and let ν be an odd integer with $\nu\ge\deg A(z)$. Now, we define the polynomials $A(z)$ and $B(z)$ of $z=e^{i\xi}$ as follows:

$$\begin{aligned}
A(z)&:=z^{\nu}\left(\frac{1+z^{-1}}{2}\right)^{r}L\left(z^{-1}\right)S\left(z^{-2}\right),\\
B(z)&:=\left(\frac{1+z}{2}\right)^{r}\left(\frac{1-z}{2}\right)^{2r}L(-z)\,Q\left(z^2\right),
\end{aligned} \tag{34}$$

where $L(z)$, $S(z)$, and $Q(z)$ are given by

$$\left|S\left(z^2\right)\right|^2=1-\left(\frac{1-z^2}{4}\right)^{r}\left(\frac{1-z^{-2}}{4}\right)^{r}\left|Q\left(z^2\right)\right|^2, \tag{35}$$

$$\begin{aligned}
|L(z)|^2&=\sum_{k=0}^{r-1}\binom{r+k-1}{k}\left(\frac{1-u}{2}\right)^k+(1-u)^ruR\left(u^2\right),\\
u&:=\frac{1}{2}\left(z+z^{-1}\right),
\end{aligned} \tag{36}$$

$$\begin{aligned}
S(1)=L(1)=1, \qquad Q(1)=(-1)^rL(-1),\\
L(0)\,Q(0)\neq 0,
\end{aligned} \tag{37}$$

where $R(z)$, $Q(z)$ are two arbitrary polynomials and $Q(z)$ satisfies (37).

The following condition (to be used in the following) is given for the minimal degree of $A(z)$ and $B(z)$:

$$\begin{aligned}
Q(z)&=\text{constant}=(-1)^rL(-1),\\
R(z)&=\text{constant}=0, \qquad \nu=4r-1.
\end{aligned} \tag{38}$$

Remark 2. Obviously, the polynomial $L(z)$ defined by (36) satisfies

$$\left|\frac{1+z}{2}\right|^{2r}|L(z)|^2+\left|\frac{1-z}{2}\right|^{2r}|L(-z)|^2=1. \tag{39}$$

This equation was solved by Daubechies in [27]. If (38) holds, then $L(z)=H_r(z)$.

Lemma 3. *The polynomial $P_0(z_1,z_2)$ given by* (30)*, where $m_0(z)$ is the univariate Daubechies filer defined by* (28)*, satisfies* (16) *if and only if the polynomials $A(z)$, $B(z)$ in* (30) *satisfy*

$$|A(z)|^2+|A(-z)|^2+|B(z)|^2+|B(-z)|^2=1, \tag{40}$$

$$A\left(z^{-1}\right)B(z)+A\left(-z^{-1}\right)B(-z)=0. \tag{41}$$

Proof. Note that

$$\begin{aligned}
\left|P_0(z_1,z_2)\right|^2&=\left|m_0(z_2)\right|^2\\
&\quad\times\left[\left|A(z_1)\right|^2+\left|B(z_1)\right|^2+z_2^{-2}A(z_1)\,B\left(z_1^{-1}\right)\right.\\
&\quad\left.+z_2^2A\left(z_1^{-1}\right)B(z_1)\right],\\
&\left|m_0(z)\right|^2+\left|m_0(-z)\right|^2=1.
\end{aligned} \tag{42}$$

If (40) and (41) hold, we have

$$\begin{aligned}
&\sum_{k_1,k_2\in\{0,1\}\times\{0,1\}} \left|P_0\left((-1)^{k_1}z_1,(-1)^{k_2}z_2\right)\right|^2 \\
&= \left|m_0(z_2)\right|^2\left[\left|A(z_1)\right|^2+\left|B(z_1)\right|^2\right. \\
&\qquad +\left|A(-z_1)\right|^2+\left|B(-z_1)\right|^2+z_2^{-2} \\
&\qquad \times\left(A(z_1)B\left(z_1^{-1}\right)+A(-z_1)B\left(-z_1^{-1}\right)\right) \\
&\qquad \left.+z_2^2\left(A\left(z_1^{-1}\right)B(z_1)+A\left(-z_1^{-1}\right)B(-z_1)\right)\right] \\
&\quad +\left|m_0(-z_2)\right|^2\left[\left|A(z_1)\right|^2+\left|B(z_1)\right|^2\left|A(-z_1)\right|^2\right. \\
&\qquad +\left|B(-z_1)\right|^2+z_2^{-2} \\
&\qquad \times\left(A(z_1)B\left(z_1^{-1}\right)+A(-z_1)B\left(-z_1^{-1}\right)\right) \\
&\qquad +z_2^2\left(A\left(z_1^{-1}\right)B(z_1)\right. \\
&\qquad \left.\left.+A\left(-z_1^{-1}\right)B(-z_1)\right)\right] \\
&= \left|m_0(z_2)\right|^2+\left|m_0(-z_2)\right|^2=1.
\end{aligned} \tag{43}$$

Hence (16) holds.

Conversely, if (16) holds, we have

$$\begin{aligned}
&\sum_{k_1,k_2\in\{0,1\}\times\{0,1\}} \left|P_0\left((-1)^{k_1}z_1,(-1)^{k_2}z_2\right)\right|^2 \\
&= \left[\left|m_0(z_2)\right|^2+\left|m_0(-z_2)\right|^2\right] \\
&\quad \times\left[\left|A(z_1)\right|^2+\left|B(z_1)\right|^2+z_2^{-2}A(z_1)B\left(z_1^{-1}\right)\right. \\
&\qquad \left.+z_2^2A\left(z_1^{-1}\right)B\left(z_1^{-1}\right)\right]+\left[\left|m_0(z_2)\right|^2+\left|m_0(-z_2)\right|^2\right] \\
&\quad \times\left[\left|A(-z_1)\right|^2+\left|B(-z_1)\right|^2+z_2^{-2}A(-z_1)B\left(-z_1^{-1}\right)\right. \\
&\qquad \left.+z_2^2A\left(-z_1^{-1}\right)B\left(-z_1^{-1}\right)\right] \\
&= \left|A(z_1)\right|^2+\left|B(z_1)\right|^2+\left|A(-z_1)\right|^2+\left|B(-z_1)\right|^2 \\
&\quad +z_2^{-2}\left[A(z_1)B\left(z_1^{-1}\right)+A(-z_1)B\left(-z_1^{-1}\right)\right] \\
&\quad \times z_2^2\left[A\left(z_1^{-1}\right)B(z_1)+A\left(-z_1^{-1}\right)B(-z_1)\right]=1,
\end{aligned} \tag{44}$$

for all $|z_1|=1$ and $|z_2|=1$. Hence (40) and (41) hold. □

Remark 4. Condition (41) means that $A(z^{-1})B(z)$ contains no even powers of z. Therefore, deg $A(z)$ must be odd if $B(0)\neq 0$.

Lemma 5. *The following conditions are equivalent:*

(i) *the univariate real polynomials $A(z)$ and $B(z)$ satisfy condition* (41) *and $B(0)\neq 0$;*

(ii) *there exist an odd integer number $\nu\geq \deg A(z)$ and real polynomials $s(z)$, $q(z)$, and $l(z)$, such that $z^{\nu}A(z^{-1})=s(z^2)l(z)$, $B(z)=q(z^2)l(-z)$, and $gcd(l(z),l(-z))=1$.*

Proof. See the proof of Lemmas 3.2 and 3.3 of [13]. □

Lemma 6. *Let the polynomials $A(z)$ and $B(z)$ be given by* (34)–(38) *for $z=e^{i\xi}$; then*

$$A\left(e^{i\xi}\right)+e^{2i\eta}B\left(e^{i\xi}\right)=0 \tag{45}$$

if and only if $\xi\equiv\pi(mod 2\pi)$.

Proof. See the proof of Lemma 4.1 and Lemma 4.7 of [13]. □

By collecting the previous results and definition we can easily show the main result of this paper that is as follows.

Theorem 7 (existence). *Let r and ν be, respectively, a nonnegative integer and an odd integer number with $\nu\geq \deg A(z)$; let the polynomials $A(z)$ and $B(z)$ be given by* (34)–(37)*; let the filter $m_0(z)$ be given by* (28)*; let the polynomials $P_0(z_1,z_2)$, $P_1(z_1,z_2)$, $P_2(z_1,z_2)$, and $P_3(z_1,z_2)$ be defined by* (30)–(33)*. Then $\{P_k(z_1,z_2)\}_{0\leq k\leq 3}$ is a bivariate nonseparable filter bank satisfying* (16) *and* (21)*. If, in addition, the polynomials $A(z)$ and $B(z)$ satisfy condition* (38)*, then $P_0(z_1,z_2)$ satisfies Cohen's criterion* (18)*, and the filter bank $\{P_k(z_1,z_2)\}_{0\leq k\leq 3}$ can be used, according to* (17)*,* (22)*, to define bivariate nonseparable compactly supported orthogonal wavelets.*

Proof. According to Lemmas 3 and 5 by letting

$$\begin{aligned}
s\left(z^2\right)&=S\left(z^2\right),\qquad q\left(z^2\right)=\left(\frac{1-z^2}{2}\right)^r Q\left(z^2\right), \\
l(z)&=\left(\frac{1+z}{2}\right)^r L(z),
\end{aligned} \tag{46}$$

in order to prove that $P_0(z_1,z_2)$ satisfies (16), we need only to show that $A(z)$, $B(z)$ satisfy (40). Taking into account (34) and (35), we have

$$\begin{aligned}
&|A(z)|^2+|B(-z)|^2 \\
&= \left|\frac{1+z}{2}\right|^{2r}|L(z)|^2\left|S\left(z^2\right)\right|^2 \\
&\quad +\left|\frac{1-z}{2}\right|^{2r}\left|\frac{1+z}{2}\right|^{4r}|L(z)|^2\left|Q\left(z^2\right)\right|^2 \\
&= \left|\frac{1+z}{2}\right|^{2r}|L(z)|^2\left(\left|S\left(z^2\right)\right|^2+\left|\frac{1-z}{2}\right|^{2r}\left|\frac{1+z}{2}\right|^{2r}\left|Q\left(z^2\right)\right|^2\right) \\
&= \left|\frac{1+z}{2}\right|^{2r}|L(z)|^2.
\end{aligned} \tag{47}$$

Hence, (40) follows from (39).

By a simple calculation, we can also get that $\{P_k(z_1,z_2)\}_{0\leq k\leq 3}$ satisfies (21). In fact, let us assume that

$$P_0(z_1,z_2)=u(z_1)v(z_2), \tag{48}$$

where $u(z)$ and $v(z)$ are univariate low-pass filters. Then we have

$$P_0(1, z_2) = v(z_2) = m_0(z_2). \tag{49}$$

This leads to the contradiction that

$$u(z_1) = A(z_1) + z_2^2 B(z_1). \tag{50}$$

Hence $P_0(z_1, z_2)$ is nonseparable.

Finally, suppose that $A(z)$ and $B(z)$ satisfy (34)–(38). Let $\Omega = [-\pi, \pi] \times [-\pi, \pi]$, then Ω is a compact fundamental domain of the lattice $2\pi\mathbb{Z}^2$. When $j \geq 1$ and $(\xi, \eta) \in \Omega$, it is $(\frac{\xi}{2^j}, \frac{\eta}{2^j}) \in [-\frac{\pi}{2}, \frac{\pi}{2}]^2$, so that we have

$$A\left(e^{i\xi/2^j}\right) + e^{2i\eta/2^j} B\left(e^{i\xi/2^j}\right) \neq 0, \qquad m_0\left(e^{i\eta/2^j}\right) \neq 0, \tag{51}$$

for all $j \geq 1$ and $(\xi, \eta) \in \Omega$ by using Lemma 6 and the result of [27]. Hence $P_0(z_1, z_2)$ satisfies Cohen's criterion (18). □

3.2. Accuracy. We will use the same method as given in [13] to study the accuracy of our scaling function. Let recall its definition and how the accuracy of the scaling function implies some restrictions on the low-pass filter.

According to the result of [14], an orthogonal scaling function $\varphi(\mathbf{x})$ with the dilation matrix $\left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right]$ has accuracy $r+1$ if and only if its low-pass filter $P_0(z_1, z_2)$ satisfies the following accuracy condition:

$$\begin{aligned} &\frac{\partial^{i+j}}{\partial z_1^i \partial z_2^j} P_0(z_1, z_2)\,|_{z_1=-1, z_2=1} = 0, \\ &\frac{\partial^{i+j}}{\partial z_1^i \partial z_2^j} P_0(z_1, z_2)\,|_{z_1=1, z_2=-1} = 0, \\ &\frac{\partial^{i+j}}{\partial z_1^i \partial z_2^j} P_0(z_1, z_2)\,|_{z_1=-1, z_2=-1} = 0, \end{aligned} \tag{52}$$

for all $i, j \geq 0$ with $i + j \leq r$. Therefore, when $P_0(z_1, z_2)$ satisfies (52), we say that it has accuracy $r + 1$.

Lemma 8. *Let the polynomials $m_0(z)$ and $P_0(z_1, z_2)$ be given by* (28) *and* (30)*, respectively; then $P_0(z_1, z_2)$ satisfies* (52) *if and only if the polynomials $A(z), B(z)$ in* (30) *defined by* (34) *have the form*

$$A(z) = \left(\frac{1+z}{2}\right)^r A_0(z), \qquad B(z) = \left(\frac{1+z}{2}\right)^r B_0(z), \tag{53}$$

where $A_0(z)$, $B_0(z)$ are polynomials with real coefficients satisfying

$$A_0(-1) + B_0(-1) = 0. \tag{54}$$

Proof. It is easy to check that

$$\begin{aligned} \frac{\partial^i P_0(z_1, z_2)}{\partial z_1^i} &= m_0(z_2)\left[A^{(i)}(z_1) + z_2^2 B^{(i)}(z_1)\right], \\ \frac{\partial^{i+1} P_0(z_1, z_2)}{\partial z_1^i \partial z_2} &= m_0'(z_2)\left[A^{(i)}(z_1) + z_2^2 B^{(i)}(z_1)\right] \\ &\quad + 2z_2 m_0(z_2) B^{(i)}(z_1), \\ &\vdots \\ \frac{\partial^{i+j} P_0(z_1, z_2)}{\partial z_1^i \partial z_2^j} &= m_0^{(j)}(z_2)\left[A^{(i)}(z_1) + z_2^2 B^{(i)}(z_1)\right] \\ &\quad + 2jz_2 m_0^{(j-1)}(z_2) B^{(i)}(z_1) \\ &\quad + j(j-1) m_0^{(j-2)}(z_2) B^{(i)}(z_1), \end{aligned} \tag{55}$$

for all $i \geq 0$ and $j \geq 2$.

If $A(z)$, $B(z)$ satisfy (53) and (54), it is not difficult to show that $P_0(z_1, z_2)$ satisfies also (52). On the other hand, if $P_0(z_1, z_2)$ satisfies (52) and noting that $m_0(1) = 1$, we have

$$\begin{aligned} &A^{(i)}(-1) + B^{(i)}(-1) = 0, \quad \text{for } 0 \leq i \leq r, \\ &A^{(i)}(-1) = 0, \quad B^{(i)}(-1) = 0, \quad \text{for } 0 \leq i < r. \end{aligned} \tag{56}$$

So that $z = -1$ is a root of multiplicity r of $A(z), B(z)$. Hence $A(z), B(z)$ have the form of (53), and $A^{(r)}(-1) + B^{(r)}(-1) = 0$ yields (54). □

Lemma 9. *Let the polynomials $m_0(z)$ and $P_0(z_1, z_2)$ be given by* (28) *and* (30)*, respectively, and assume also that $A(z)$ and $B(z)$ in* (30) *are polynomials with $A(1) \neq 0$ and $B(0) \neq 0$. Then $P_0(z_1, z_2)$ satisfies* (41) *with accuracy $r + 1$ if and only if there exist an odd integer number $\nu \geq \deg A(z)$ and real polynomials $S(z)$, $Q(z)$, and $L(z)$, such that $A(z)$ and $B(z)$ satisfy* (34) *and* (37).

Proof. If there exist an odd integer $\nu \geq \deg A(z)$ and real polynomials $S(z), Q(z)$, and $L(z)$ such that $A(z)$ and $B(z)$ satisfy (34), and (37), then $P_0(z_1, z_2)$ satisfies (41) and has accuracy $r + 1$ by Lemmas 5 and 8 by letting

$$\begin{aligned} s(z^2) &= S(z^2), \qquad q(z^2) = \left(\frac{1-z^2}{2}\right)^r Q(z^2), \\ l(z^2) &= \left(\frac{1+z}{2}\right)^r L(z), \\ A_0(z) &= z^{\nu-r} L(z^{-1}) S(z^{-2}), \\ B_0(z) &= L(-z)\left(\frac{1-z}{2}\right)^{2r} Q(z^2). \end{aligned} \tag{57}$$

Now, suppose that $P_0(z_1, z_2)$ satisfies (41) with accuracy $r + 1$. By Lemma 8, $z = -1$ is a root of multiplicity $2r$ of $G(z) := z^\nu A(z^{-1}) B(z)$. Since $G(-z) = G(z)$ by (41), $z = 1$ is another

root of multiplicity $2r$ of $G(z)$. But $A(1) \neq 0$, so $z = 1$ is a root of multiplicity $2r$ of $B(z)$. Hence, we have

$$A\left(z^{-1}\right) = \left(\frac{1+z}{2}\right)^r A_1\left(z^{-1}\right),$$
$$B(z) = \left(\frac{1+z}{2}\right)^r \left(\frac{1-z}{2}\right)^{2r} B_1(z). \tag{58}$$

By substituting (58) in (41), we have

$$A_1\left(z^{-1}\right) B_1(z) + A_1\left(-z^{-1}\right) B_1(-z) = 0. \tag{59}$$

Then, by applying Lemma 5 to $A_1(z)$ and $B_1(z)$, we get that there exist an odd integer $\nu \geq \deg A(z)$ and real polynomials $S(z)$, $Q(z)$, and $L(z)$, such that (34) hold.

Since $B(1) = 0$, and therefore $A(1) = 1$, we can normalize $S(z)$ and $L(z)$, so that $S(1) = L(1) = 1$. And $Q(1) = (-1)^r L(-1)$ follows from condition (54). Hence $A(z)$ and $B(z)$ satisfy (34), and (37). □

Theorem 10 (accuracy). *Let the polynomials $A(z)$, $B(z)$, and $P_0(z_1, z_2)$ be given by* (34)–(38) *and* (30) *with $m_0(z_2)$ according to* (28). *Then $P_0(z_1, z_2)$ is a low-pass filter which can generate bivariate nonseparable compactly supported orthogonal scaling function with accuracy $r + 1$.*

Proof. This theorem follows from Lemmas 8, 9, and Theorem 7. □

Remark 11. (i) According to the proof of Theorem 2.1 of [13], if $A(z)$, $B(z)$ satisfy (40), the polynomials $S(z)$ and $L(z)$ in Lemma 9 must satisfy (35) and (36).

(ii) The existence of $S(z)$ and $L(z)$ satisfying (35)–(38) is ensured by Corollary 4.2 of [13].

3.3. Smoothness. In this section, we measure the smoothness of the scaling function $\varphi(\mathbf{x})$ generated by $P_0(z_1, z_2)$ by its Hölder exponent (see [13, 23]). Let $s = n + t$, where n is a nonnegative integer and $0 \leq t < 1$. For nonnegative integers α_1, α_2, and $\alpha = \alpha_1 + \alpha_2$, $D^\alpha \varphi(\mathbf{x})$ denotes the partial derivative of $\varphi(\mathbf{x})$; that is,

$$D^\alpha \varphi(\mathbf{x}) := \frac{\partial^\alpha \varphi(\mathbf{x})}{\partial^{\alpha_1} x_1 \partial^{\alpha_2} x_2}, \quad \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2. \tag{60}$$

If there exists a constant c, such that

$$\left|D^\alpha \varphi(\mathbf{x}) - D^\alpha \varphi(\mathbf{y})\right| \leq c|\mathbf{x} - \mathbf{y}|^t, \quad \forall \alpha \leq n \text{ and } \mathbf{x}, \mathbf{y} \in \mathbb{R}^2, \tag{61}$$

we say that $\varphi(\mathbf{x}) \in C^s(\mathbb{R}^2)$. It is well known that if the Fourier transform $\widehat{\varphi}(\omega)$ of $\varphi(\mathbf{x})$ satisfies

$$\left|\widehat{\varphi}(\omega)\right| \leq c\left(1 + |\xi|\right)^{-s-1-\varepsilon}\left(1 + |\eta|\right)^{-s-1-\varepsilon}, \quad \omega = (\xi, \eta) \in \mathbb{R}^2, \tag{62}$$

for some constants $c > 0$, $\varepsilon > 0$, then $\varphi(\mathbf{x}) \in C^s(\mathbb{R}^2)$ (see [13, 23]).

Table 1: The coefficients of the polynomials $A(z)$ and $B(z)$ of the example given in Section 3.4

k	$a_k,\ b_k$
0	0.007903570868, 0.036969146934
1	−0.013689386305, −0.064032440902
2	−0.138084015734, −0.056780853066
3	0.121182418797, 0.137970734671
4	0.549926093940, 0.002654265329
5	0.346172096397, −0.083844146934
6	0.080254350926, 0.017157440802
7	0.046334871111, 0.009905853066

Let us assume that the polynomials $A(z)$ and $B(z)$ satisfying (34)–(38) are

$$A(z) = \left(\frac{1+z}{2}\right)^r A_0(z),$$
$$B(z) = \left(\frac{1+z}{2}\right)^r B_0(z), \tag{63}$$

and let

$$l(z) = \left|A_0(z)\right| + \left|B_0(z)\right|. \tag{64}$$

Define a polynomial $T_r(y)$ as follows:

$$T_r(y) = \sum_{k=0}^{r-1} \binom{r+k-1}{k} y^k, \quad \text{for } y \in \mathbb{R}. \tag{65}$$

Lemma 12. *Let $l(z)$ be given by* (64). *Then there exist constants $c > 0$ and $0 < \varepsilon < 1$, such that*

$$\prod_{k=1}^{\infty} l\left(z^{2^{-k}}\right) \leq c|\xi|^{\mu_r + 1 - \varepsilon}, \tag{66}$$

where $z = e^{i\xi}$ and $\mu_r := (1/2)\log_2 T_r(3/4)$. Furthermore,it is

$$r - \mu_r > \left(1 - \frac{1}{2}\log_2 3\right) r + \frac{1}{2}\log_2 3,$$
$$\lim_{r \to \infty} \frac{r - \mu_r - 2}{r} = 1 - \frac{1}{2}\log_2 3 \approx 0.2075. \tag{67}$$

Proof. See the proof of Lemma 4.1, 4.3–4.5 of [13]. □

Theorem 13 (smoothness). *Let the polynomials $m_0(z)$, $A(z)$, and $B(z)$ be given by* (28) *and* (34)–(38), *respectively. If $\varphi(\mathbf{x})$ is the bivariate scaling function generated by the low-pass filter $P_0(z_1, z_2)$ according to* (30) *and $r \geq 5$, then $\varphi(\mathbf{x}) \in C^{r-\mu_r-2}(\mathbb{R}^2)$.*

Proof. We only need to prove that the Fourier transform $\widehat{\varphi}(\omega)$ of $\varphi(\mathbf{x})$ satisfies (62). In fact, by applying Lemmas 7.1.1–7.1.8 of [2], we have

$$\left|\prod_{k=1}^{\infty} m_0\left(e^{i\eta/2^k}\right)\right| \leq c\left(1 + |\eta|\right)^{-r+\mu_r} \leq c\left(1 + |\eta|\right)^{-r+\mu_r+1-\varepsilon}, \tag{68}$$

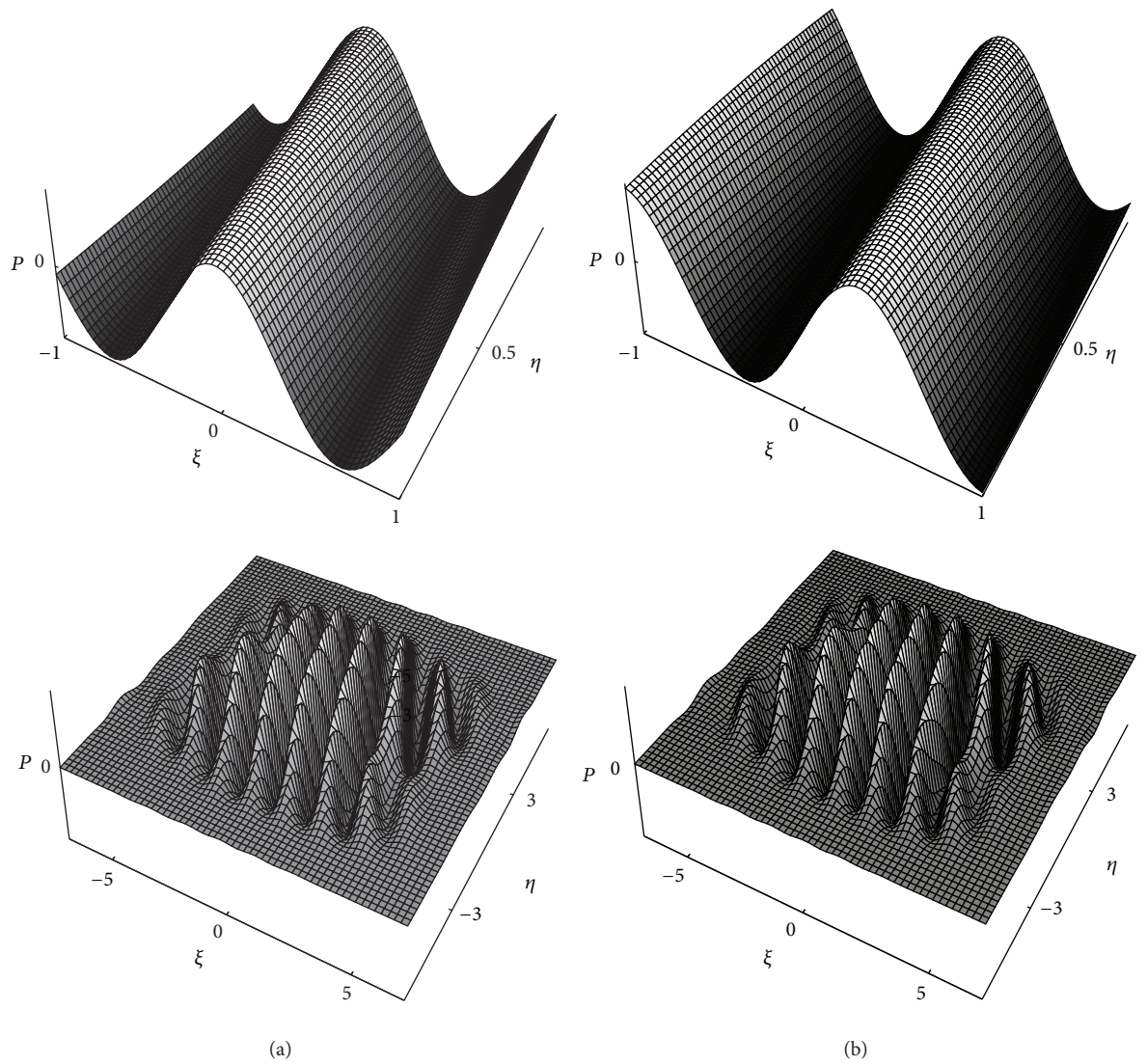


Figure 1: Plot of the real (a) and imaginary (b) parts of the bivariate compactly supported low-pass filter (74).

for the same ε as Lemma 12, and

$$\left|\prod_{k=1}^{\infty}\left(\frac{1+e^{i\xi/2^k}}{2}\right)^r\right| \le c(1+|\xi|)^{-r}. \tag{69}$$

Thus, according to (63), (64), (68), and (69) and Lemma 12, we get

$$\begin{aligned}&\left|\prod_{k=1}^{\infty}P\left(e^{i\xi/2^k},e^{i\eta/2^k}\right)\right|\\ &\le\left|\prod_{k=1}^{\infty}m_0\left(\frac{i\eta}{e^{2^k}}\right)\right|\left|\prod_{k=1}^{\infty}\left(\frac{1+e^{\frac{i\xi}{2^k}}}{2}\right)^r\right|\left|\prod_{k=1}^{\infty}l\left(e^{i\xi/2^k}\right)\right|\\ &\le c(1+|\xi|)^{-r+\mu_r+1-\varepsilon}(1+|\eta|)^{-r+\mu_r+1-\varepsilon}.\end{aligned} \tag{70}$$

Hence by using (17) it is shown that (62) holds for $s = r-\mu_r-2$. □

Remark 14. (i) The wavelets have the same smoothness as the corresponding scaling function. So we only need to study the smoothness of the scaling function.

(ii) As the remark of paper [13], (67) mean that the smoothness of the bivariate orthogonal wavelets improves asymptotically by 0.2075 when r is incremented by 1. Hence, (67) guarantees the existence of bivariate orthogonal wavelets of any desired smoothness. In particular, if $r > 4$, then $\varphi(\mathbf{x}) \in C^0(\mathbb{R}^2)$, if $r > 8$, then $\varphi(\mathbf{x}) \in C^1(\mathbb{R}^2)$, and if $r > 12$, then $\varphi(\mathbf{x}) \in C^2(\mathbb{R}^2)$.

According to the results of this section, we can summarize the following procedure for constructing the filter banks.

Step 1. Choose a nonnegative integer r, and let $\nu = 4r-1$.

Table 2: The coefficients of the low-pass filter $P_0(z_1, z_2)$ of the example given in Section 3.4

k	$p_{k,0}$, $p_{k,1}$, $p_{k,2}$, $p_{k,3}$
0	0.002629285291, 0.006377324227, 0.015933220981, 0.028763021603
1	−0.004554055712, −0.011045849577, −0.027597148268, −0.049819014795
2	−0.045936485868, −0.111418819867, −0.082391050025, −0.027173124419
3	0.040313822269, 0.097781064805, 0.101627868651, 0.094966452370
4	0.182944217753, 0.443730695949, 0.253781634345, −0.072104378716
5	0.115161262724, 0.279323325343, 0.131304280272, −0.114390178100
6	0.026698259299, 0.064756554347, 0.042614945757, 0.003008971526
7	0.015414247193, 0.037387214084, 0.024603750404, 0.001737230521
k	$p_{k,4}$, $p_{k,5}$, $p_{k,6}$, $p_{k,7}$
0	0.016325987784, −0.004712828742, −0.003158691008, 0.001302285961
1	−0.028277440326, 0.008162858829, 0.005471013312, −0.002255625450
2	−0.014314162669, 0.002801853046, 0.004851428418, −0.002000178904
3	0.053095656592, −0.014358762363, −0.011788395329, 0.004860197371
4	−0.045765749098, 0.019013502031, −0.000226783811, 0.000093499925
5	−0.068135421758, 0.023514243110, 0.007163750722, −0.002953518393
6	0.001033287041, 0.000510619607, −0.001465953599, 0.000604393018
7	0.000596568551, 0.000294806368, −0.000846368705, 0.000348946471

Step 2. Use some spectral factorization method to make a list of all polynomials $S(z)$ and $L(z)$ that satisfy (35)–(38). Substitute $S(z)$, $L(z)$, and Q in (34) to get $A(z)$ and $B(z)$.

Step 3. Construct the low-pass filter $P_0(z_1, z_2)$ by substituting $A(z)$, $B(z)$, and univariate Daubechies low-pass filter $m_0(z_2)$ in (30).

Step 4. Construct the high-pass filters $P_i(z_1, z_2), i = 1, 2, 3$ by substituting $A(z)$, $B(z)$, and univariate Daubechies filters $m_0(z_2)$ and $m_1(z_2)$ in (31)–(33).

3.4. Example. Finally, we give an example of bivariate nonseparable compactly supported orthogonal low-pass filter by using our method.

Let $r = 2$. The following polynomial $m_0(z)$ is the univariate Daubechies low-pass filter with accuracy 3 (see [27]):

$$m_0(z) = \sum_{k=0}^{5} p_k z^k, \tag{71}$$

where

$$\begin{aligned} p_0 &= 0.332670552950, \quad & p_1 &= 0.806891509311, \\ p_2 &= 0.459877502118, \quad & p_3 &= -0.135011020010, \\ p_4 &= -0.085441273882, \quad & p_5 &= 0.035226291886. \end{aligned} \tag{72}$$

If (38) holds, then there are four polynomials $S(z)$ satisfying (35), (37) and two polynomials $L(z)$ satisfying (36), (37). Hence, one can get 8 kinds of $A(z)$ and $B(z)$ satisfying (34)–(38) (see [13]). One of them is given as follows:

$$A(z) = \sum_{k=0}^{7} a_k z^k, \qquad B(z) = \sum_{k=0}^{7} b_k z^k, \tag{73}$$

where the coefficients $a_k, b_k, k = 0, 1, \ldots, 7$ are displayed in Table 1.

Then, by using (30), we can get a bivariate nonseparable compactly supported orthogonal low-pass filter (Figure 1) given as follows:

$$P_0(z_1, z_2) = \sum_{k,l=0}^{7} p_{k,l} z_1^k z_2^l, \tag{74}$$

where the coefficients $p_{k,l}$ are displayed in Table 2.

For instance, by explicit computation and the approximation of coefficients the first few terms of the filter look like

$$\begin{aligned} &0.002629 - 0.004553e^{(i/2)\,\xi} - 0.045936e^{i\,\xi} \\ &+ 0.040313e^{(3i/2)\,\xi} + 0.182944e^{2i\xi} + 0.115161e^{(5i/2)\,\xi} \\ &+ 0.026698e^{3i\,\xi} + 0.015413e^{(7i/2)\,\xi} \\ &- 0.011045e^{(i/2)\,\xi+(i/2)\,\eta} - 0.111418e^{i\,\xi+(i/2)\,\eta} \\ &+ 0.09778e^{(3i/2)\,\xi+(i/2)\,\eta} + 0.44373e^{2\,i\,\xi+(i/2)\,\eta} \cdots. \end{aligned} \tag{75}$$

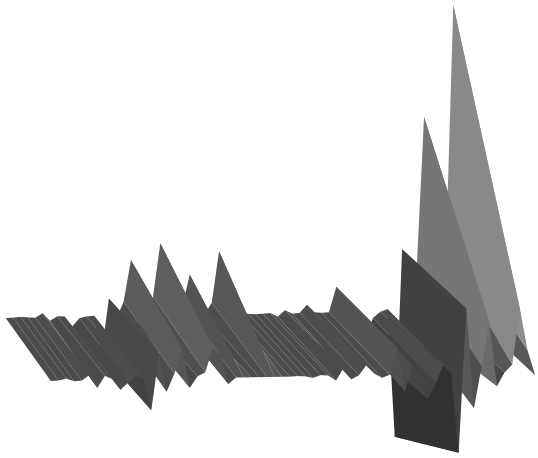


Figure 2: The scaling function generated by the low-pass filter (74).

By the inverse Fourier transform we finally obtain the scaling function, so that we have a sequence in the following form:

$$\begin{aligned}
&-0.00921\delta(-3+x)\,\delta(-3+y)\\
&-0.001424\delta(-2+x)\,\delta(-3+y)\\
&+0.030481\delta(-1+x)\,\delta(-3+y)\\
&-0.019846\delta(x)\,\delta(-3+y)\\
&-0.010634\delta(-7+2x)\,\delta(-3+y)\\
&+0.090022\delta(-5+2x)\,\delta(-3+y)\\
&-0.148136\delta(-3+2x)\,\delta(-3+y)\\
&+0.06875\delta(-1+2x)\,\delta(-3+y)+\cdots.
\end{aligned} \tag{76}$$

The scaling function generated by the low-pass filter (Figure 2) has accuracy 3. By (31)–(33), one can easily obtain the corresponding high-pass filters. We omit it here.

4. Conclusion

We gave a method for constructing bivariate nonseparable compactly supported orthogonal wavelets by combining three methods using the dilation matrix $\left[\begin{smallmatrix}2&0\\0&2\end{smallmatrix}\right]$. We studied the accuracy of the scaling functions and the smoothness of the wavelets constructed by this method. We proved that the scaling functions have the same accuracy order as the univariate Daubechies low-pass filter used in this method and the wavelets can be made arbitrarily smooth by choosing the accuracy order.

Acknowledgments

The authors wish to thank the anonymous reviewers for their valuable comments and suggestions that have improved the presentation of this paper. This work was supported by the National Natural Science Foundation of China (11271001, 61170311), 973 Program (2013CB329404), and Sichuan Province Science and Technology Research Project (12ZC1802). a

References

[1] C. K. Chui, *An Introduction to Wavelets*, vol. 1 of *Wavelet Analysis and Its Applications*, Academic Press, Boston, Mass, USA, 1992.

[2] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, Pa, USA, 1992.

[3] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.

[4] C. Cattani and J. Rushchitsky, *Wavelet and Wave Analysis as Applied to Materials with Micro or Nanostructure*, vol. 74 of *Series on Advances in Mathematics for Applied Sciences*, World Scientific, Singapore, 2007.

[5] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, Amsterdam, The Netherlands, 3rd edition, 2008.

[6] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, Mass, USA, 1996.

[7] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall, New York, NY, USA, 1995.

[8] C. Cattani and L. M. S. Ruiz, "Discrete differential operators in multidimensional Haar wavelet spaces," *International Journal of Mathematics and Mathematical Sciences*, vol. 2004, no. 44, pp. 2347–2355, 2004.

[9] A. Cohen and I. Daubechies, "Nonseparable bidimensional wavelet bases," *Revista Matemática Iberoamericana*, vol. 9, no. 1, pp. 51–137, 1993.

[10] P. Wojtaszczyk, *A Mathematical Introduction to Wavelets*, vol. 37 of *London Mathematical Society Student Texts*, Cambridge University Press, Cambridge, UK, 1997.

[11] A. Ayache, "Construction of non-separable dyadic compactly supported orthonormal wavelet bases for $L^2(R^2)$ of arbitrarily high regularity," *Revista Matemática Iberoamericana*, vol. 15, no. 1, pp. 37–58, 1999.

[12] A. Ayache, "Some methods for constructing nonseparable, orthonormal, compactly supported wavelet bases," *Applied and Computational Harmonic Analysis*, vol. 10, no. 1, pp. 99–111, 2001.

[13] E. Belogay and Y. Wang, "Arbitrarily smooth orthogonal nonseparable wavelets in R^2," *SIAM Journal on Mathematical Analysis*, vol. 30, no. 3, pp. 678–697, 1999.

[14] C. Cabrelli, C. Heil, and U. Molter, "Accuracy of lattice translates of several multidimensional refinable functions," *Journal of Approximation Theory*, vol. 95, no. 1, pp. 5–52, 1998.

[15] B. Han, "Symmetric multivariate orthogonal refinable functions," *Applied and Computational Harmonic Analysis*, vol. 17, no. 3, pp. 277–292, 2004.

[16] W. He and M. J. Lai, "Examples of bivariate nonseparable compactly supported orthonormal continuous wavelets," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 949–953, 2000.

[17] A. Karoui, "A note on the design of nonseparable orthonormal wavelet bases of $L^2(R^3)$," *Applied Mathematics Letters*, vol. 18, no. 3, pp. 293–298, 2005.

[18] M. J. Lai, "Methods for constructing nonseparable compactly supported orthonormal wavelets," in *Wavelet Analysis: Twenty*

Year's Development, D. X. Zhou, Ed., vol. 1, pp. 231–251, World Scientific, Singapore, 2002.

[19] M. J. Lai and D. W. Roach, "Nonseparable symmetric wavelets with short support," in *7th Wavelet Applications in Signal and Image Processing*, vol. 3813 of *Proceedings of SPIE*, pp. 132–146, 1999.

[20] J. S. Leng, T. Z. Huang, and Y. D. Fu, "Construction of bivariate nonseparable compactly supported biorthogonal wavelets," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 3625–3629, Kunming, China, July 2008.

[21] J. S. Leng, T. Z. Huang, Y. D. Fu, and C. H. Lai, "A method for constructing trivariate nonseparable compactly supported orthogonal wavelets," *International Journal of Computer Mathematics*, vol. 86, no. 7, pp. 1264–1273, 2009.

[22] J. S. Leng, T. Z. Huang, Y. F. Jing, and W. Jiang, "A study on conjugate quadrature filters," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, Article ID 231754, 2011.

[23] R. L. Long, *Multi-Dimensional Wavelet Analysis*, Springer, Beijing,China, 1995.

[24] D. Stanhill and Y. Y. Zeevi, "Two dimensional orthogonal wavelets with vanishing moments," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2579–2590, 1996.

[25] D. Stanhill and Y. Y. Zeevi, "Two dimensional orthogonal filter banks and wavelets with linear phase," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 183–190, 1998.

[26] L. F. Villemoes, "Continuity of nonseparable quincunx wavelets," *Applied and Computational Harmonic Analysis*, vol. 1, no. 2, pp. 180–187, 1994.

[27] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 182638, 6 pages
http://dx.doi.org/10.1155/2013/182638

Research Article

Dynamically Weighted Clique Evolution Model in Clique Networks

Zhang-Wei Li,[1] Xu-Hua Yang,[1] Feng-Ling Jiang,[1] Guang Chen,[1] Guo-Qing Weng,[1] and Mei Zhu[2]

[1] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*
[2] *College of Information Engineering, Zhejiang University of Science & Technology, Hangzhou 310023, China*

Correspondence should be addressed to Xu-Hua Yang; xhyang@zjut.edu.cn

Received 2 February 2013; Accepted 26 February 2013

Academic Editor: Shengyong Chen

This paper proposes a weighted clique evolution model based on clique (maximal complete subgraph) growth and edge-weight driven for complex networks. The model simulates the scheme of real-world networks that the evolution of networks is likely to be driven by the flow, such as traffic or information flow needs, as well as considers that real-world networks commonly consist of communities. At each time step of a network's evolution progress, an edge is randomly selected according to a preferential scheme. Then a new clique which contains the edge is added into the network while the weight of the edge is adjusted to simulate the flow change brought by the new clique addition. We give the theoretical analysis based on the mean field theory, as well as some numerical simulation for this model. The result shows that the model can generate networks with scale-free distributions, such as edge weight distribution and node strength distribution, which can be found in many real-world networks. It indicates that the evolution rule of the model may attribute to the formation of real-world networks.

1. Introduction

With the rapid development of information and other sciences, today connections between people have been widely extended and communications within systems which can be abstracted into networks can be found all round us. Nowadays, human's living and producing activities depend much wider and deeper on those complex network systems than ever before [1–3]. How to ensure those systems to work in a secure and efficient way is a big deal [4, 5]. Beginning from the studies on a small-world [6], scale-free [7], and navigable [8] phenomenon in common networks, a significant concern and attention have been attracted from different fields to establish and study a new interdisciplinary subject, the network science.

Just like the three phenomena mentioned above, real-world networks have been found that they commonly evolve by simple mechanisms and gradually form [7] into the complex systems observed by us. Revealing those simple rules is one important aspect of network science research, as those rules can help us better understand and improve the properties and behaviors of networks. In many of those researches, simple unweighted network models were widely used for analysis [7, 9, 10]. However, sometimes it is not sufficient, as simply using an edge to represent the connections between nodes in network will lose some features we need to be concerned, such as the importance of nodes, the traffic flow on edges. As a result, many weighted evolution network models [11–15] were proposed. Among them, two famous and widely accepted examples are the BBV model [12] and the DM model [13]. The two models have successfully reproduced the scale-free phenomena while they proposed that a network's evolution should be driven by traffic flow needs, which is similar to the situation in many real-world networks, like road network [16], and mobile agent network [17].

Researches have also revealed that communities (communities are known as kinds of subgraphs of networks that the connections within themselves are tight, while connections between them and other parts of networks are sparse) exist widely in complex networks [18–21]. For instance, social

networks [3, 22] include communities such as families and clubs and bus transport networks [23] consist of communities representing bus routes. Usually, these networks evolve with units of communities and communities may overlap with each other. For example, the bus transport network expands by adding communities which are comprised of stations in one bus route. Obviously, there exist overlapping communities in the network because an arbitrary station may be included in several bus routes [24–26].

On the basis, in this paper we propose a new weighted network evolution model based on edge weight driven (where edge weight can be viewed as to represent the flow or other meaningful things of a connection) and clique [19] growth (where cliques are maximal complete sub graphs in a network, used in this paper to mimic communities in real-world networks). Through theoretical analysis and numerical simulation of the model, it is proven that our model can reproduce scale-free distributions [7] while it involves addition of cliques. The remainder of the paper is organized as follows. In section two, we will present our model. In sections three and four, some theoretical analysis and numerical simulation about the model will be introduced. Finally in section five, we will give our discussion and conclusion.

2. Presentation of the Evolution Model of Clique Growth and Edge-Weight Driven

We present our new model in this section. At the beginning time step $t = 0$, the model starts with a clique (a maximal complete subgraph) denoted by c_0, which consists of m_0 vertices. The weight of each edge in the clique is initially set as w_0. Then the network will be expanded continuously as well as change its inner edges' weight on the meanwhile dynamically based on the scheme of clique growth and edge-weight driven. The specific generating progress for each time step of the evolution is as follows.

(1) Preferential selection: randomly select from the network an edge according to a preferential scheme based on edge weight. Specifically, the probability of an edge being selected is proportional to its edge weight. That is, if w_{ij} denotes the weight of an edge connecting nodes i and j, P_{ij} denotes the probability for the edge being selected, then $P_{ij} = w_{ij} / \sum_{m,n} w_{mn}$.

(2) Clique growth: add a new clique consisting of $m_2 + 2$ nodes into the network, while rule that the new clique must contain the edge selected by step (1). For the edges except the selected edge, the weight is set as w_0.

(3) Edge weight adjustment: give the selected edge a weight increment of Δ, which makes the weight of this edge change from w_{ij} to $w_{ij} + \Delta$.

It is worth noting that the increment of Δ is introduced in the above step (3), as in real situations; when a community is added into a network, it will bring some influence on the original network. For example, in bus transport networks, when a bus route is added, the passenger flow of the original network will be changed.

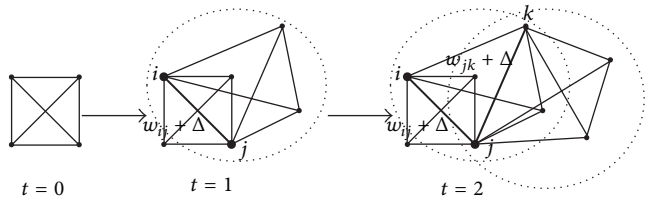


Figure 1: Illustration example of the evolution process.

We give an illustrative example for the evolution process in Figure 1. The parameters are set as $m_0 = 4$, $m_2 = 2$, and $w_0 = 1$. At the time step $t = 0$, the initial network is constructed as a clique which is composed of $m_0 = 4$ vertices. At the time step $t = 1$, a new clique is added into the network (dotted lines). The edge connecting nodes i and j is selected from the network based on the edge weight preferential scheme. Two new nodes are added into the network to constitute the clique. After the addition, we set the weights of edges expecting the selected edge w_0 and change the weight of the selected edge into $w_{ij} + \Delta$. At the time step $t = 2$, another clique is added into the network and the rule the same.

3. Theoretical Analysis of the Model with a Brief Simulation

We give some theoretical analysis of the model based on mean field theory [27] in this section. The analysisis concerned with the edge weight distribution and the node strength distribution, which are two important indices to reflect the structure properties in weighed networks. The edge weight is widely used to depict the specific differences of connections in networks, such as the intimacy between friends, the distance between places, and the traffic flow capacity of a road, and so forth. On the other side, the node strength is used to depict the importance of the nodes which may represent people, routers, and so forth, in networks. Usually, the edge weight and the node strength have kinds of relationship. In many previous researches, the relationship is defined as $s_i = \sum_j w_{ij}$ [11, 12] which is similar to the real-world situation, where s_i denotes the strength of a specific node i. In this paper, we adapt this relationship for further analysis.

As can be seen, when the incensement parameter is set as $\Delta = 0$, the new model will behave similarly like the BA scale-free model [7]. The only difference is that new cliques instead of new nodes are added into networks during their evolution. So in this situation, the theoretical analysis of our model should resemble BAs. But when nonzero Δ is introduced, we need to make some changes in order to obtain correct results.

In each time step, the evolution process of our model begins with a preferential selection of an edge, followed by an addition of a new clique. As a result, when an edge connecting nodes i and j is selected, the weight w_{ij} will obey (1) according to the mean-field theory:

$$\frac{dw_{ij}}{dt} = \frac{w_{ij}}{\sum_{m,n} w_{mn}} \Delta. \quad (1)$$

Given t_{ij} denotes the time step when this edge was added into the network, then the initial condition

$$w_{ij}\left(t_{ij}\right) = w_0 \tag{2}$$

for (1) should be satisfied. According to the model, the total weight being added into the network in each time step is equal to

$$\Gamma = \left[\frac{\left(m_2+2\right)\left(m_2-1\right)}{2} - 1\right] w_0 + \Delta. \tag{3}$$

So

$$\sum_{m,n} w_{mn} \approx \Gamma t \tag{4}$$

when $t \to \infty$.

Solving (1), (2), and (3), we can obtain the specific formula of $w_{ij}(t)$. For example, with the scenario in Figure 1, the parameters are $m_0 = 4$, $m_2 = 2$, and $w_0 = 1$, so we get $\Gamma = 5 + \Delta$, and the formation of $w_{ij}(t)$ can be obtained as

$$w_{ij}(t) = \left(\frac{t}{t_{ij}}\right)^{\Delta/(5+\Delta)}. \tag{5}$$

With solved $w_{ij}(t)$, the edge weight distribution can be further derived. Here, also taking Figure 1 scenario as an example, we can get

$$t_{ij} = t w_{ij}^{-(5+\Delta)/\Delta}. \tag{6}$$

According to the mean field theory, the equations can be transformed into

$$\begin{aligned} P\left(w_{ij} < w\right) &= P\left(t_{ij} > t w^{-(5+\Delta)/\Delta}\right) \\ &= 1 - P\left(t_{ij} \le t w^{-(5+\Delta)/\Delta}\right) \\ &= 1 - w^{-(5+\Delta)/\Delta}, \end{aligned} \tag{7}$$

where $P(w_{ij} < w)$ is the proportion of edges in a network whose weight is less than w and $P(t_{ij} > t)$ ($P(t_{ij} \le t)$) is the proportion of edges which were added into the network after, before, or at the time step t.

Further, we can get the result as

$$p(w,t) = \frac{\partial\left(P\left(w_{ij}(t) < w\right)\right)}{\partial w} = \frac{(5+\Delta)}{\Delta \times w^{-(2\Delta+5)/\Delta}}, \tag{8}$$

$$p(w) = \lim_{t\to\infty} p(w,t) = \frac{1}{\beta} w^{-\Upsilon} \sim w^{-\gamma}, \tag{9}$$

where $p(w)$ is the edge weight distribution, $\beta = \Delta/(5+\Delta)$, and $\gamma = (2\Delta+5)/\Delta$.

Equation (9) shows that, the edge weight distribution of our model is a kind of scale-free distribution. We conduct a numerical simulation for the same parameters in the scenario of Figure 1 to verify the theoretical result. The value of Δ is set as $\Delta = 3$. The network evolution progress stops at $t = 10^4$. We show the resulting edge weight distribution for the generated network in Figure 2. From the theoretical analysis, we know that exponent γ of the scale-free distribution should be equal to $(2\Delta+5)/\Delta \approx 3.67$. Figure 2 shows that simulation result agrees with the theoretical analysis. Although there is some deviation at the head of the distribution, the error should be decreased with the increase of evolution time.

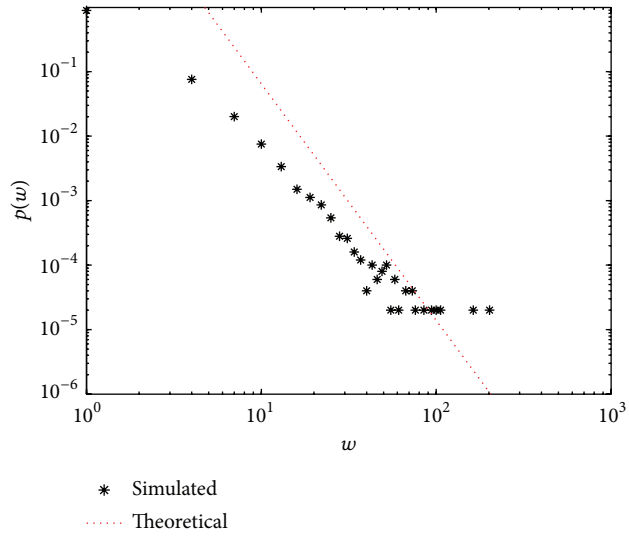


Figure 2: The edge weight distribution of the simulated network.

Similarly, for node i, when an edge connecting this node with other node j is selected in one time step, its strength will increase by m_2 (brought by the new added clique) $+\Delta$ (brought by the weight adjustment). Summing up all the possible increments brought when one of the edges emanated from node i, ds_i/dt can be derived as

$$\begin{aligned} \frac{ds_i}{dt} &= \sum_j \frac{w_{ij}}{\sum_{m,n} w_{mn}} \left(m_2+\Delta\right) \\ &= \sum_j \frac{w_{ij}}{\Gamma t} \left(m_2+\Delta\right) \\ &= \frac{s_i}{\Gamma t} \left(m_2+\Delta\right). \end{aligned} \tag{10}$$

Here we have the initial condition $s_i(t_i) = 1$, where t_i denotes the time when node i is added into network. After solving (10) with the initial condition, we can obtain the specific formula of $s_i(t)$. For Figure 1 scenario, the result should be

$$s_i = \left(\frac{t}{t_i}\right)^{5/(4+2\Delta)}. \tag{11}$$

Similarly with the derivation of the edge weight distribution, we can obtain the node strength distribution denoted as $p(s)$ with the help of (11). The result for Figure 1 scenario is

$$p(s) \sim s^{-\gamma}, \tag{12}$$

where $\gamma = (3\Delta+4)/5$.

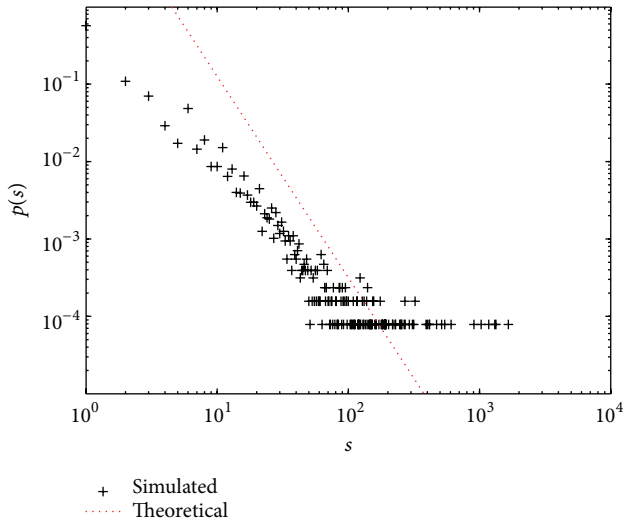


Figure 3: The node strength distribution of the simulated network.

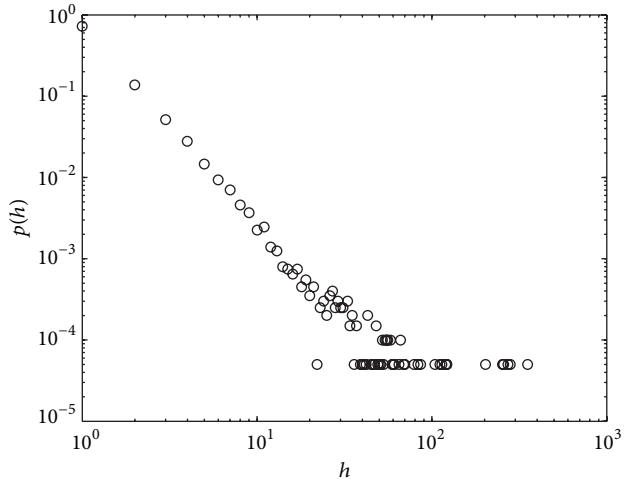


Figure 4: The distribution of the number of cliques that a node belongs to.

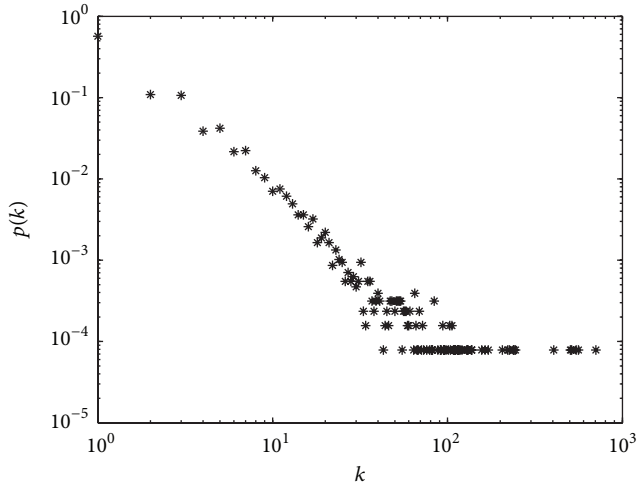


Figure 5: The node degree distribution.

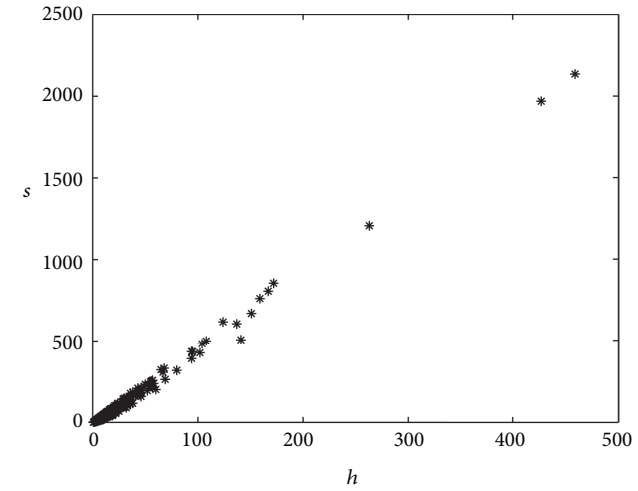


Figure 6: The correlation between node strength (s) and number of cliques nodes belong to (h).

Equation (12) indicates that the node strength distribution is also a kind of scale-free distributions. The node strength distribution for the simulated network is presented in Figure 3. The theoretical value for the scale-free exponent γ equals 2.6. From Figure 3, we can see that the simulated node strength distribution agrees well with the theoretical result.

4. More Detailed Information about the Simulation

In this section, we give more information about the simulation mentioned in the above section. The result is presented from Figure 4 to Figure 8. Figures 4 and 5 indicate that the distributions of the number of cliques that a node belongs to and the node degree distribution are kind of scale-free distributions. The reason for this phenomenon is that node strength should have some linear relationship with the node degree as well as with the number of cliques that a node belongs to. These relationships have been verified with Figures 6 and 7. Figures 8 and 9 indicate that the weight of an edge and the strength of a node in the network increase logarithmically with time. This is in accordance with the theoretical result showed in (5) and (11).

5. Conclusion

In this paper, we have proposed a new weighted evolution model for complex networks based on clique growth and edge weight-driven. The result has proven that our model can generate networks consisting of cliques and also with scale-free distributions, such as edge weight distribution, and node strength distribution.

Using weighted and community network models can help us figure out the actions in real-world network much better. The work of this paper is just a preliminary exploration of

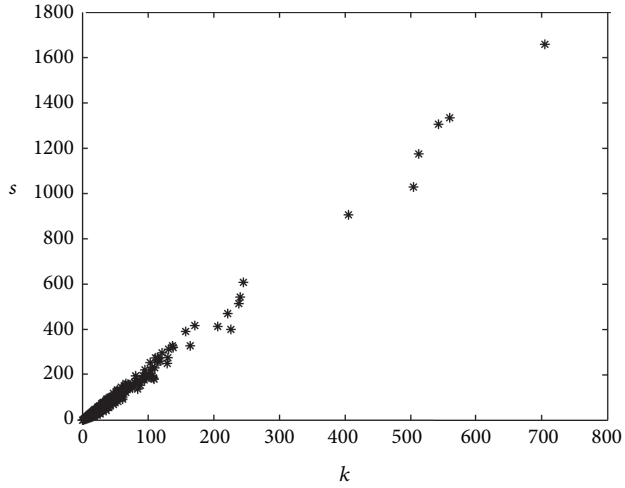


FIGURE 7: The correlation between node strength (s) and node degree (k).

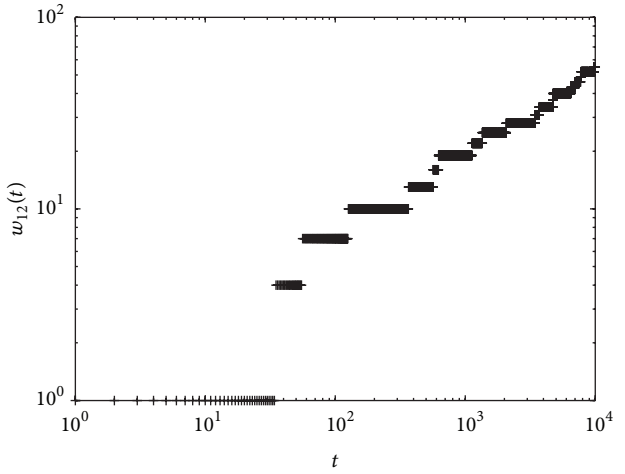


FIGURE 8: The change of the weight of the very first edge (w_{12}) with time (t).

the evolution mechanisms of real-world networks. The real-world networks are much more complicated and diverse although the rules behind them may be simple, but the rules are coupled with each other. For instance, the real-world networks sometimes may grow by either single nodes or communities at a specific time step, while sometimes they may reduce nodes or communities during their evolution as well as take some inner structure adjustment. In this paper, we have only considered some static properties of networks. The fact is that the dynamical properties of networks are as equally important as the static ones. We hope to solve these problems in a good way. this will be our future research content.

Acknowledgments

The work is supported by the National Natural Science Foundation of China under Grant no. 60874080, the Commonweal Application Technique Research Project of Zhejiang Province under Grant no. 2012C23126, and the Natural Science Foundation of Zhejiang Province of China under Grant no. LY12E07005.

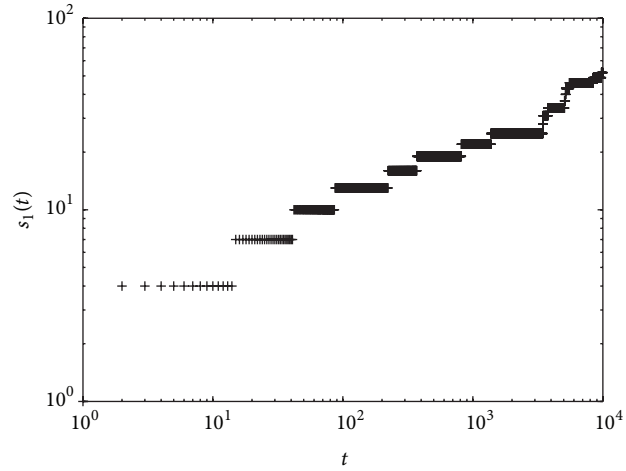


FIGURE 9: The change of the strength of the very first node (s_1) with time (t).

References

[1] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, "Economic networks: the new challenges," *American Association for the Advancement of Science. Science*, vol. 325, no. 5939, pp. 422–425, 2009.

[2] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[3] A. L. Traud, P. J. Mucha, and M. A. Porter, "Social structure of Facebook networks," *Physica A*, vol. 391, no. 16, pp. 4165–4180, 2012.

[4] X.-H. Yang, G. Chen, B. Sun, S.-Y. Chen, and W.-L. Wang, "Bus transport network model with ideal n-depth clique network topology," *Physica A*, vol. 390, no. 23-24, pp. 4660–4672, 2011.

[5] J. Xue, M. Li, W. Zhao, and S. Y. Chen, "Bound maxima as a traffic feature under DDOS flood attacks," *Mathematical Problems in Engineering*, vol. 2012, Article ID 465613, 9 pages, 2012.

[6] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[7] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *American Association for the Advancement of Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[8] J. M. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, no. 6798, p. 845, 2000.

[9] Z.-J. Bao and Y.-J. Cao, "Cascading failures in local-world evolving networks," *Journal of Zhejiang University A*, vol. 9, no. 10, pp. 1336–1340, 2008.

[10] Y.-X. Li, X.-G. Jin, F.-S. Kong, and H.-L. Luo, "Strategic games on a hierarchical network model," *Journal of Zhejiang University A*, vol. 9, no. 2, pp. 271–278, 2008.

[11] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 11, pp. 3747–3752, 2004.

[12] A. Barrat, M. Barthélemy, and A. Vespignani, "Weighted evolving networks: coupling topology and weight dynamics," *Physical Review Letters*, vol. 92, no. 22, Article ID 228701, 4 pages, 2004.

[13] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, "Size-dependent degree distribution of a scale-free growing network," *Physical Review E*, vol. 63, no. 6, part 1, Article ID 062101, 2001.

[14] G. H. Wen, Z. S. Duan, G. R. Chen, and X. M. Geng, "A weighted local-world evolving network model with aging nodes," *Physica A*, vol. 390, no. 21-22, pp. 4012–4026, 2011.

[15] H. H. Jo, R. K. Pan, and K. Kaski, "Emergence of bursts and communities in evolving weighted networks," *PLOS ONE*, vol. 6, no. 8, Article ID e22687, 2011.

[16] F. Bono, E. Gutirrez, and K. Poljansek, "Road traffic: a case study of flow and path-dependency in weighted directed networks," *Physica A*, vol. 389, no. 22, pp. 5287–5297, 2010.

[17] H. X. Yang, W.-X. Wang, Y.-C. Lai, and B. H. Wang, "Traffic-driven epidemic spreading on networks of mobile agents," *EPL*, vol. 98, no. 6, Article ID 68003, 2012.

[18] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Physical Review E*, vol. 69, no. 6, Article ID 066113, 2004.

[19] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[20] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 4, pp. 1118–1123, 2008.

[21] C. Seshadhri, T. G. Kolda, and A. Pinar, "Community structure and scale-free collections of Erdos-Renyi graphs," *Physical Review E*, vol. 85, no. 5, Article ID 056109, 2012.

[22] T. Zhou, M. Medo, G. Cimini, Z. K. Zhang, and Y. C. Zhang, "Emergence of Scale-Free leadership structure in social recommender systems," *PLoS ONE*, vol. 6, no. 7, Article ID e20648, 2011.

[23] X.-H. Yang, G. Chen, B. Sun, S.-Y. Chen, and W.-L. Wang, "Bus transport network model with ideal n-depth clique network topology," *Physica A*, vol. 390, no. 23-24, pp. 4660–4672, 2011.

[24] S. Chen, W. Huang, C. Cattani, and G. Altieri, "Traffic dynamics on complex networks: a survey," *Mathematical Problems in Engineering*, vol. 2012, Article ID 732698, 23 pages, 2012.

[25] M. Li, W. Zhao, and S. Chen, "mBm-based scalings of traffic propagated in internet," *Mathematical Problems in Engineering*, vol. 2011, Article ID 389803, 21 pages, 2011.

[26] W. Huang and S.-Y. Chen, "Epidemic metapopulation model with traffic routing in scale-free networks," *Journal of Statistical Mechanics*, Article ID P12004, 19 pages, 2011.

[27] A. L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A*, vol. 272, no. 1-2, pp. 173–187, 1999.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 163734, 5 pages
http://dx.doi.org/10.1155/2013/163734

Research Article

A Shannon-Runge-Kutta-Gill Method for Convection-Diffusion Equations

Xiaoming Duan,[1] Jinsong Leng,[1] Carlo Cattani,[2] and Caiyun Li[1]

[1] *School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China*
[2] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

Correspondence should be addressed to Jinsong Leng; jslengjs@gmail.com

Received 20 January 2013; Accepted 5 February 2013

Academic Editor: Shengyong Chen

A Shannon-Rugge-Kutta-Gill method for solving convection-diffusion equations is discussed. This approach transforms convection-diffusion equations into one-dimensional equations at collocations points, which we solve by Runge-Kutta-Gill method. A concrete example solved is used to examine the method's feasibility.

1. Introduction

Most of the physics phenomenon are stated in terms of partial differential equations (PDEs). Convection-diffusion equation is a kind of PDE which can be used in many science and technology fields, especially in image and signal procession such as image segmentation and the quickly stability of image processing. The numerical solution of convection-diffusion equations as an important subject has always attracted the attentions of the researchers for a long time.

The standard Galerkin finite-element method can solve the solution of the equations, but it is numerically unstable for small values of the diffusion parameter. So on the basis of this method, in [1] King and Krueger investigated the effect of a stabilized finite-element approximation and drew a conclusion that the stabilized system can provide accurate controllers. A novel multilevel particle methods and two complementary approaches are researched in [2]. In this paper, a new class of particle based on mapping functions is introduced, and particle remeshing is used as a key element in overlapping domains in the particle-AMR method. For the two-dimensional convection-diffusion equation, Gupta et al. proposed a fourth-order nine-point compact finite-difference formulae, which is shown to be computationally efficient and stable and yield highly accurate numerical solutions in [3, 4]. The resulting linear system is solved by classical iterative methods for large values of the Reynolds number in [4]. With the wavelet method, Shi et al. solved the solution of covection-diffusion equations by Haar wavelet method in [5]. In short, convection-diffusion equations are studied by scholars via different methods.

Recently, Wavelet analysis as a new subject has attracted a lot of attention. As a mathematical tool, it has been widely used in numerical analysis, signal processing [6, 7], image processing, and so forth. Many years ago, wavelet methods were used for numerical analysis, particularly the numerical solution of PDEs. Up to now, researchers have utilized the simplest Haar wavelets to solve kinds of PDEs. Chen and Hsiao, in [8], established an operational matrix of integration based on Haar wavelets and used a procedure for applying the matrix to obtain wavelet solution of PDEs. In [9, 10], Cattani solved Poisson's problem and Fredholm type integral equations by Harmonic wavelet method. Other wavelets are also extensively used to solve the kinds of PDEs, in which Shannon wavelet is applied in the numerical solution of some equations, such as [11, 12]. Shannon scaling function and Sinc function combined with other methods (Galerkin, etc.) are used to solve some PDEs [13, 14]. In light of the above description, we are enlightened that Shannon wavelet is a useful tool to obtain the solution of convection-diffusion equations, which combined with Rugge-Kutta-Gill method.

In this paper, the content is assigned as follows. In Section 2, Shannon wavelet is introduced. We elaborate the concrete method solving convection-diffusion equation in

Section 3. In Section 4 the viability of Shannon wavelet collocation is tested by a listed example.

2. Preliminaries

2.1. Shannon Wavelet. Wavelets are classified as families with names, such as Haar wavelet, Meyer wavelet, and Shannon wavelet. Shannon wavelets are the real part of harmonic wavelets. They have a slow decay in the time domain but a very sharp compact support in the frequency (Fourier) domain. This fact, together with the Parseval's identity, is used to compute the inner product and the expansion coefficients of the Shannon wavelets easily. A set of Shannon scaling functions in the subspace V_j is defined as

$$\varphi_{j,k}(t) = 2^{j/2}\frac{\sin\pi\left(2^j t-k\right)}{\pi\left(2^j t-k\right)}, \quad k\in\mathbf{Z} \tag{1}$$

and the mother wavelets are

$$\psi_{j,k}(t) = 2^{j/2}\frac{\sin\pi\left(2^j t-k-1/2\right)-\sin 2\pi\left(2^j t-k-1/2\right)}{\pi\left(2^j t-k-1/2\right)}, \quad k\in\mathbf{Z}. \tag{2}$$

In (1) and (2), the scaling function and mother wavelet for $j = k = 0$ (see Figure 1) are as follows:

$$\varphi(t) = \frac{\sin\pi t}{\pi t} = \frac{e^{\pi i t}-e^{-\pi i t}}{2\pi i t}, \tag{3}$$

$$\psi(t) = \frac{\sin\pi(t-1/2)-\sin 2\pi(t-1/2)}{\pi(t-1/2)}. \tag{4}$$

To some properties of Shannon scaling and wavelet functions, Cattani has detailedly researched in [15–18]. So in this paper, we will not narrate the properties again.

To (3), its Fourier transform $\widehat{\varphi}(\omega) = \chi_{[-1/2,1/2)}$. It is very easy to see that

$$\sum_{n=-\infty}^{\infty}\left|\widehat{\varphi}(\omega+n)\right|^2 = 1. \tag{5}$$

According to this equation, the sequence of function $\{\varphi(x-n)\}_{n=-\infty}^{\infty}$ is orthonormal. A reproducing kernel is generated [19] as follows:

$$K(x-y) = \frac{\sin\pi(x-y)}{\pi(x-y)}. \tag{6}$$

Recomposing (6), we obtain a new reproducing kernel

$$w(x-y) = \frac{\sin(\pi/\Delta)(x-y)}{(\pi/\Delta)(x-y)}, \tag{7}$$

where Δ is the spatial mesh size.

In one-dimensional function $f(x)$, we make the domain $[a,b]$ be discrete and set the grid size

$$\Delta = \frac{b-a}{2^j} \tag{8}$$

So we obtain collocation points

$$x_i = i\Delta, \quad i = 0,1,2,\ldots,2^j, \tag{9}$$

where 2^j is a number of nodes, which used in the discretization and also is the maximum wavelet index number. Now, a basis function $w_j(x-x_n)$ of Shannon wavelet will be constructed by (7)

$$w_j(x-x_n) = \frac{\sin(\pi/\Delta)(x-x_n)}{(\pi/\Delta)(x-x_n)}, \quad n = 0,1,2,\ldots,2^j. \tag{10}$$

It has some properties as follows.

(i) To the random x_k ($k = 0,1,2,\ldots,2^j$), the function $w_j(x-x_n)$ fulfills interpolation property:

$$w_j(x_k-x_n) = \delta_{kn} = \begin{cases}1 & k=n,\\ 0 & k\neq n.\end{cases} \tag{11}$$

(ii) We have noticed that the constructed basis functions are orthogonal to each other as follows:

$$\int_{-\infty}^{\infty} w_j(x-x_k)\,w_j(x-x_n)\,dx = \Delta\delta_{kn}. \tag{12}$$

(iii) If we make the integral with the basis functions and their derived functions, we obtain.

$$\int_{-\infty}^{\infty} w_j(x-x_k)\frac{d^m w_j(x-x_n)}{dx^m}dx = \Delta\frac{d^m w_j(x_k-x_n)}{dx^m}. \tag{13}$$

Both $w_j(x-x_k)$ and its associated wavelet play an important part in signal processing. Unfortunately, when $x\to\infty$, the reduction of $w_j(x-x_k)$ is very slow. So our paper only researches the case which x belongs to finite interval.

2.2. Function Approximation. According to Shannon's sampling theorem, any function $f(x)\in B_\pi^2$ can be denoted as [19]

$$f(x) = \sum_{n\in Z} f(x_n)\,w_j(x-x_n), \tag{14}$$

where the coefficients $f(x_n)$ is the value of the function $f(x)$ at the point x_n. B_π^2 is the Paley-Wiener reproducing kernel Hilbert space which is a subspace of the Hilbert space $L^2(R)$.

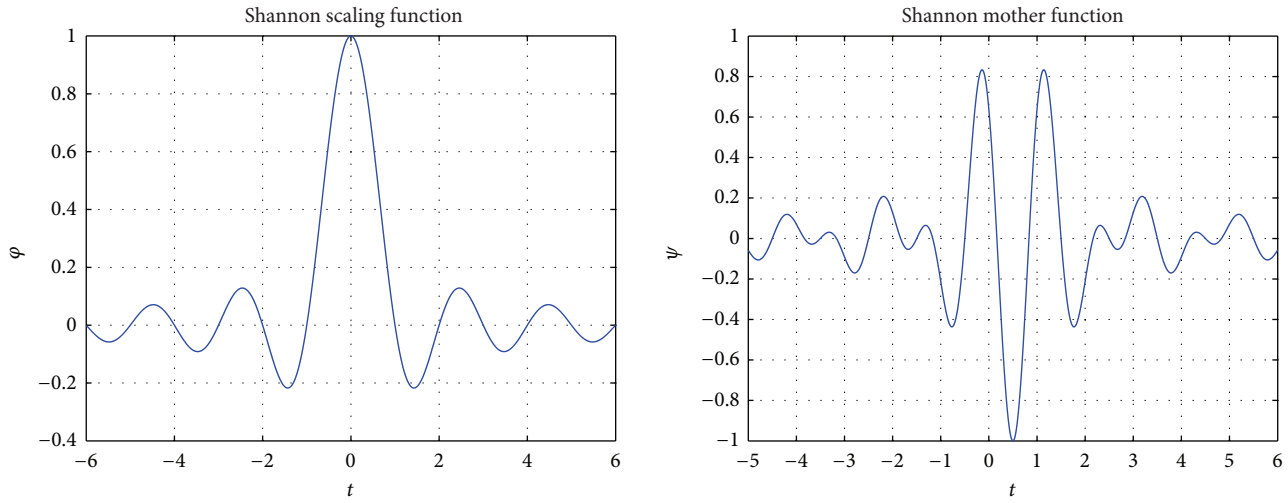


Figure 1: Shannon scaling function φ and mother function ψ.

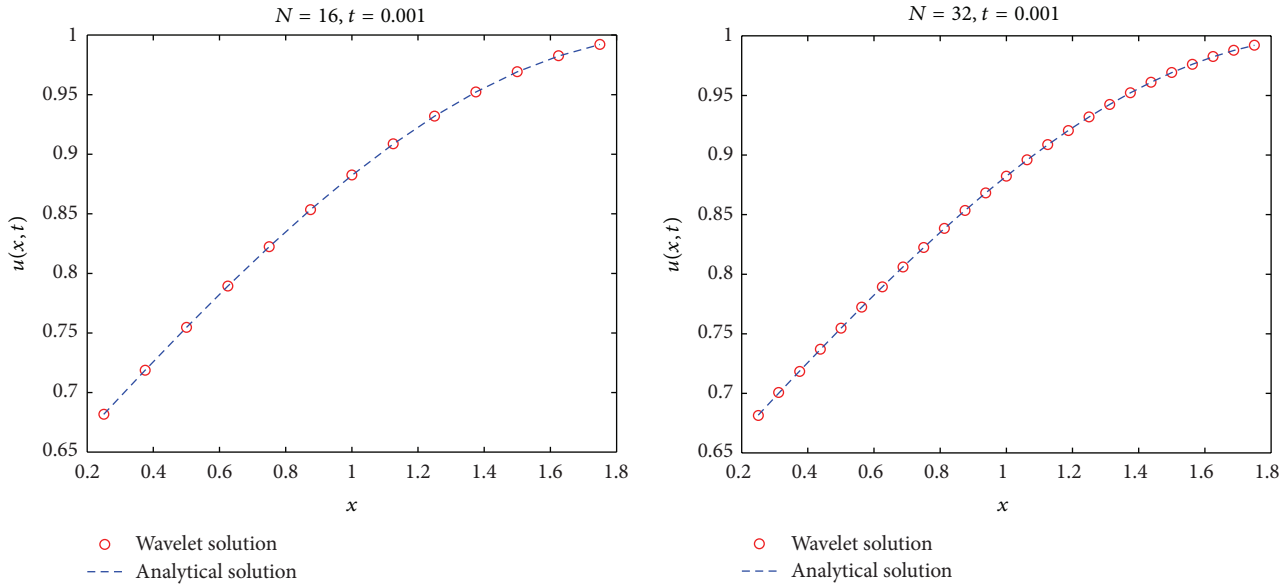


Figure 2: Comparison of the analytical solution and Shannon wavelet solution.

In V_j, $f(x)$ can be approximated by $f_j(x) \in V_j$. So we get

$$f(x) \approx f_j(x) = \sum_{n=0}^{2^j} f_j(x_n)\, w_j(x - x_n). \tag{15}$$

3. Method of Solution of Convection-Diffusion Equation

In this section, let us consider the one-dimensional convection-diffusion equation with constant coefficients:

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = \alpha\frac{\partial^2 u}{\partial x^2} \quad 0 < x < 2,\ 0 < t < T \tag{16}$$

with initial condition and boundary conditions:

$$\begin{aligned} u(x,0) &= f(x), \quad 0 \le x \le b, \\ u(0,t) &= g_0(t), \quad u(2,t) = g_1(t), \quad 0 < t \le T. \end{aligned} \tag{17}$$

Like (8) and (9), we will also divide the interval $[0,2]$ into $N = 2^j$ equal parts of length $\Delta = 2/N$ and denote $x_i = i\Delta$, $i = 0, 1, 2, \ldots, N$. We know that $u(x,t)$ can be approximated by $u_j(x,t) \in V_j$ expanded in terms of the constructed basis function as formula (15) as follows:

$$u(x,t) \approx u_j(x,t) = \sum_{n=0}^{N} u_j(x_n,t)\, w_j(x - x_n). \tag{18}$$

We multiply formula (16) with the constructed basis function $w_j(x - x_k)$, then we obtain

$$\frac{\partial u_j(x,t)}{\partial t} w_j(x - x_k) = -a\frac{\partial u_j(x,t)}{\partial x} w_j(x - x_k) + \alpha\frac{\partial^2 u_j(x,t)}{\partial x^2} w_j(x - x_k). \tag{19}$$

Integrate that formula (19) with respect to x from $-\infty$ to ∞ as follows:

$$\int_{-\infty}^{\infty} \frac{\partial u_j(x,t)}{\partial t} w_j(x - x_k)\,dx = \int_{-\infty}^{\infty} -a\frac{\partial u_j(x,t)}{\partial x} w_j(x - x_k)\,dx + \int_{-\infty}^{\infty} \alpha\frac{\partial^2 u_j(x,t)}{\partial x^2} w_j(x - x_k)\,dx. \tag{20}$$

The left expression of formula (20) is as follows:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{\partial u_j(x,t)}{\partial t} w_j(x - x_k)\,dx \\
&= \int_{-\infty}^{\infty} \frac{\partial \sum_{n=0}^{N} u_j(x_n,t)\, w_j(x - x_n)}{\partial t} w_j(x - x_k)\,dx \\
&= \sum_{n=0}^{N} \frac{\partial u_j(x_n,t)}{\partial t} \int_{-\infty}^{\infty} w_j(x - x_n)\, w_j(x - x_k)\,dx \\
&= \frac{\partial \sum_{n=0}^{N} u_j(x_n,t)\, w_j(x_k - x_n)}{\partial t}\Delta \\
&= \frac{\partial u_j(x_k,t)}{\partial t}\Delta.
\end{aligned} \tag{21}$$

The right expression of formula (20) is as follows:

$$\begin{aligned}
&\int_{-\infty}^{\infty} -a\frac{\partial \sum_{n=0}^{N} u_j(x_n,t)\, w_j(x - x_n)}{\partial x} w_j(x - x_k)\,dx \\
&\quad + \int_{-\infty}^{\infty} \alpha\frac{\partial^2 \sum_{n=0}^{N} u_j(x_n,t)\, w_j(x - x_n)}{\partial x^2} w_j(x - x_k)\,dx \\
&= -a\sum_{n=0} N u_j(x_n,t) \int_{-\infty}^{\infty} \frac{\partial w_j(x - x_n)}{\partial x} w_j(x - x_k)\,dx \\
&\quad + \alpha\sum_{n=0} N u_j(x_n,t) \int_{-\infty}^{\infty} \frac{\partial^2 w_j(x - x_n)}{\partial x^2} w_j(x - x_k)\,dx \\
&= -a\Delta\sum_{n=0} N u_j(x_n,t)\, w_j'(x_k - x_n) \\
&\quad + \alpha\Delta\sum_{n=0} N u_j(x_n,t)\, w_j''(x_k - x_n).
\end{aligned} \tag{22}$$

Via (21) = (22), we obtain the following expression:

$$\frac{\partial u_j(x_k,t)}{\partial t} = \sum_{n=0}^{N} u_j(x_n,t)\left[-a w_j'(x_k - x_n) + \alpha w_j''(x_k - x_n)\right] \tag{23}$$

To any x_k, we will get one equation. So $N + 1$ equations will be obtained. In order to simplify the $N + 1$ equations, we define the matrices U and V as follows:

$$\begin{aligned}
U &= \left[u_j(x_0,t), u_j(x_1,t), u_j(x_2,t), \ldots, u_j(x_N,t)\right]^{\top}, \\
v_{kn} &= -a w_j'(x_k - x_n) + \alpha w_j''(x_k - x_n), \\
V &= (v_{kn})_{(N+1)\times(N+1)}.
\end{aligned} \tag{24}$$

Combining with (24), the formula (23) is evolved into a matrix equation:

$$\frac{\partial U}{\partial t} = VU. \tag{25}$$

Now we will use Runge-Kutta-Gill method to solve formula (25) as follows:

$$\begin{aligned}
U_{i+1} &= U_i + \frac{\Delta t}{6}\left[K_1 + \left(2 - \sqrt{2}\right)K_2 + \left(2 + \sqrt{2}\right)K_3 + K_4\right], \\
K_1 &= VU_i, \\
K_2 &= V\left(U_i + \frac{1}{2}K_1\right), \\
K_3 &= V\left(U_i + \frac{\sqrt{2} - 1}{2}K_1 + \frac{2 - \sqrt{2}}{2}K_2\right), \\
K_4 &= V\left(U_i - \frac{\sqrt{2}}{2}K_2 + \frac{2 + \sqrt{2}}{2}K_3\right),
\end{aligned} \tag{26}$$

where Δt is the time interval. From (9) and (17), the initial value U_0 is obtained, and then we can evaluate the numerical solution at any collocation point within the different parameter t.

4. Test of Example

A concrete convection-diffusion equation has known exact solution will be considered, and we observe how well the Shannon wavelet solution approximates the exact solution. We assume that $\alpha = 0.1$, $a = 0.8$,

$$\begin{aligned}
f(x) &= e^{-(x-2)^2/8}, \\
g_0(x) &= \sqrt{\frac{20}{20 + t}}\, e^{-(5+4t)^2/10(t+20)}, \\
g_1(t) &= \sqrt{\frac{20}{20 + t}}\, e^{-2(5+2t)^2/5(t+20)},
\end{aligned} \tag{27}$$

For which the exact solution is

$$u(x,t) = \sqrt{\frac{20}{20+t}} e^{-(x-2-0.8t)^2/0.4(t+20)}. \quad (28)$$

In the course of the experiment, we got $t_{\max} = 0.01$ and setted $\Delta t = 0.00001$. We got the approximate charts in the case of $N = 32$ and $N = 16$ and obtained the conclusion that the wavelet solution is approximate to the exact solution more precisely (see Figure 2).

5. Conclusion

In this paper, the theory of Shannon wavelet combined with Runge-Kutta-Gill method is used to solve the approximation of convection-diffusion equations. It has been shown that the key idea of shannon wavelet collocation method is to transform convection-diffusion equations into one-dimensional equations at collocations points and to solve the problem via Rugge-Kutta-Gill method.

Acknowledgment

The authors are thankful to the referees for their valuable comments and suggestions that improved the presentation of this paper. This work was supported by the National Natural Science Foundation of China (11271001).

References

[1] B. B. King and D. A. Krueger, "The 1-D convection diffusion equation: galerkin least squares approximations and feedback control," in *Proceedings of the 43rd IEEE Conference on Decision and Control*, pp. 1502–1507, 2004.

[2] M. Bergdorf, G.-H. Cottet, and P. Koumoutsakos, "Multilevel adaptive particle methods for convection-diffusion equations," *Multiscale Modeling and Simulation*, vol. 4, no. 1, pp. 328–357, 2005.

[3] M. M. Gupta, R. P. Manohar, and J. W. Stephenson, "A fourth order, cost effective and stable finite difference scheme for the convection-diffusion equation," in *Numerical Properties and Methodologies in Heat Transfer*, pp. 201–209, Hemisphere, 1983.

[4] M. M. Gupta, R. P. Manohar, and J. W. Stephenson, "A single cell high order scheme for the convection-diffusion equation with variable coefficients," *International Journal for Numerical Methods in Fluids*, vol. 4, no. 7, pp. 641–651, 1984.

[5] Z. Shi, L. Y. Deng, and Q. J. Chen, "Numerical solution of differential equations by using Haar wavelets," in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition*, pp. 2–4, Beijing, China, 2007.

[6] C. Cattani, "Wavelet based approach to fractals and fractal signal denoising," *Transactions on Computational Sciences Journal*, vol. 5730, pp. 143–162, 2009.

[7] J. S. Leng, T. Z. Hung, Y. F. Jing, and W. Jiang, "A study on conjugate quadrature filters," *EURASIP Journal on Advances in Singnal Processing*, vol. 2011, Article ID 231754, 7 pages, 2011.

[8] C. F. Chen and C. H. Hsiao, "Haar wavelet method for solving lumped and distributedparameter systems," *IEE Proceedings—Control Theory and Applications*, vol. 144, pp. 87–94, 1997.

[9] C. Cattani, "Harmonic wavelet solution of Poisson's problem," *Balkan Journal of Geometry and Its Applications*, vol. 13, no. 1, pp. 27–37, 2008.

[10] C. Cattani and A. Kudreyko, "Harmonic wavelet method towards solution of the Fredholm type integral equations of the second kind," *Applied Mathematics and Computation*, vol. 215, no. 12, pp. 4164–4171, 2010.

[11] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, vol. 2010, Article ID 408418, 22 pages, 2010.

[12] C. Cattani, "Second order Shannon wavelet approximation of C^2-functions," *Politehnica University of Bucharest Scientific Bulletin A*, vol. 73, no. 3, pp. 73–84, 2011.

[13] V. G. Koures, "Solving the Coulomb Schrödinger equation in $d = 2+1$ via sinc collocation," *Journal of Computational Physics*, vol. 128, no. 1, pp. 1–5, 1996.

[14] F. Stenger, "A "sinc-Galerkin" method of solution of boundary value problems," *Mathematics of Computation*, vol. 33, no. 145, pp. 85–109, 1979.

[15] C. Cattani, "Connection coefficients of Shannon wavelets," *Mathematical Modelling and Analysis*, vol. 11, no. 2, pp. 117–132, 2006.

[16] C. Cattani, "Shannon wavelets theory," *Mathematical Problems in Engineering*, vol. 2008, Article ID 164808, 24 pages, 2008.

[17] C. Cattani, "Signorini cylindrical waves and Shannon wavelets," *Advances in Numerical Analysis*, vol. 2012, Article ID 731591, 24 pages, 2012.

[18] C. Cattani, "Fractional calculus and Shannon wavelets," *Mathematical Problems in Engineering*, vol. 2012, Article ID 502812, 26 pages, 2012.

[19] G. W. Wei, "Quasi wavelets and quasi interpolating wavelets," *Chemical Physics Letters*, vol. 296, pp. 215–222, 1998.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 364791, 7 pages
http://dx.doi.org/10.1155/2013/364791

Research Article

Dynamics-Based Stranded-Crowd Model for Evacuation in Building Bottlenecks

Lidi Huang,[1,2] Deming Liu,[1] and Yongyi Zhang[2]

[1] *School of Architecture, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China*
[2] *College of Civil and Architecture Engineering, Northeast Petroleum University, Daqing, Heilongjiang 163318, China*

Correspondence should be addressed to Lidi Huang; huangld@yeah.net

Received 7 December 2012; Revised 5 January 2013; Accepted 15 January 2013

Academic Editor: Sheng-Yong Chen

In high-density public buildings, it is difficult to evacuate. So in this paper, we propose a novel quantitative evacuation model to insure people's safety and reduce the risk of crowding. We analyze the mechanism of arch-like clogging phenomena during evacuation and the influencing factors in emergency situations at bottleneck passages; then we design a model based on crowd dynamics and apply the model to a stadium example. The example is used to compare evacuation results of crowd density with different egress widths in stranded zones. The results show this model proposed can guide the safe and dangerous egress widths in performance design and can help evacuation routes to be selected and optimized.

1. Introduction

One of the most important essential requirements for safety design of a building is safe evacuation of people from buildings in emergency conditions. From the early 20th century, there have been a lot of disasters that occurred during emergency evacuation of crowds. For example, 173 people died in the 1943 stampede incident in the subway station in London, UK; in the 1989 Hillsborough English FA Cup Stampede, 96 people died, and 400 were injured [1]; 20 soccer fans were killed during a stampede at a soccer game which became Africa's worst sports-related disaster in Accra, Ghana, in 2001; 21 children died, and 47 people were injured in a grandstand failure at a school in Xuzhou, China, in 2002, and so forth [2]. So it is very significant to study on safe design of large scale buildings and influence factors during the evacuation process, since these incidents often occurred in high-density crowds buildings.

There have been more than one hundred years for the research on human behavior during the evacuation process. With improvement of computer technology, the research has developed to computer simulation [3, 4] from the initial postdisaster investigations. Egress modeling is one of the most important means of egress investigation [5, 6]. Popular models, such as the social force model [7] and the discrete model [8], the latter including lattice gas model and cellular automata model, are able to successfully simulate some typical phenomena observed in pedestrian dynamics. Software based on the above models, EVACENT4, STEPS, building EXODUS, EXITT, Legion, and so forth, have been applied widely to simulate and evaluate very complex factors influencing people evacuation. Recently, some experimental results have also been presented to validate the models of pedestrian flow and evacuation [9–11].

However, it is very complicated to analyze human behavior in emergency conditions. Furthermore, systematic study on panic behavior and quantitative theories of predicting crowd dynamics are few. Computation for especially determination of people evacuation is so complex that their application is limited. Thus, several authors have devoted their work to the research of simple model and the calculation of evacuation time. Zhang et al. [12, 13] proposed an SCM to calculate the number of stranded crowd with different egress widths and modify the evacuation time; Liu et al. [14] developed an evacuation model of crowd streams at a stadium to calculate and analyze the total evacuation time.

Relieving panic and stress of people from building space and construction environment is the duty of architects. However, evacuation design research has not been properly

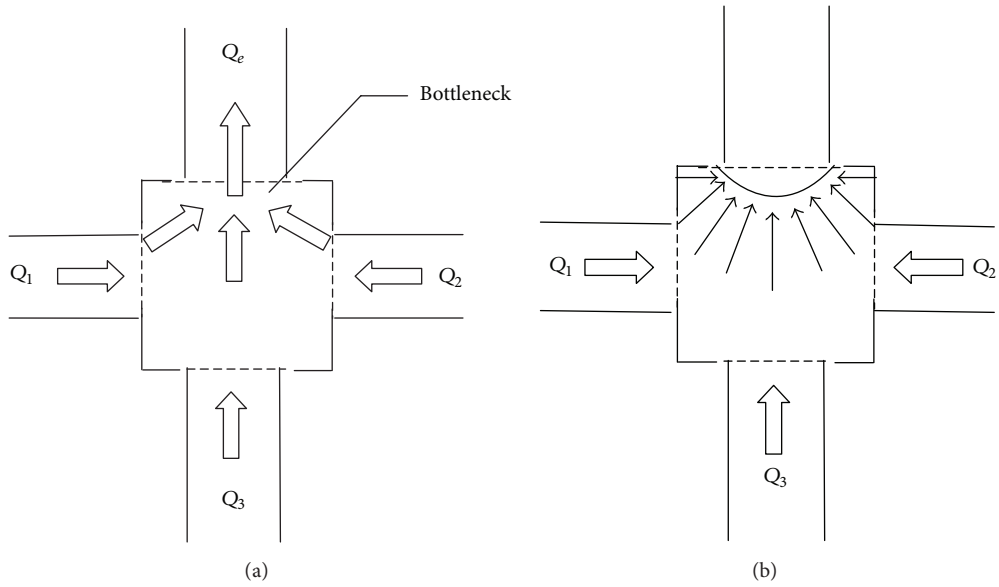


Figure 1: Sketch of typical bottleneck.

established in architectural discipline. Thus, the current study hopes to serve such a need and is inspired by the study of Zhang et al. [12] and Liu et al. [14]. In this paper, a simple continuous-time model based on the dependency of physical parameters (density, velocity, and flow-rate) is set up to discuss the concepts of safe space design in terms of evacuation processes efficiency in high-density crowd situations and panicked behavior. This paper intends to introduce this model and apply the analysis results to architectures, to provide modifications in architectural layouts to optimize evacuation.

A brief description of the organization of this paper is as follows. The main calculation principles in this paper are stated in Section 3. The parameters analysis of Section 3 is described in Section 2. The application is given in Section 4. The results and analysis of the application are considered in Section 5. Finally, a conclusion is drawn to summarize the paper.

2. Preliminaries

2.1. Concept. The evacuation route of pedestrians in a stadium can be simply attributed as leaving from spectators' stands to vertical passages, from passages to doors or stairs, and then to safe areas. Traffic nodes among the evacuation routes, such as changed building space, narrowed channels, incompletely opened exits, merging points of several crowd streams (Figure 1(a)), and stairs, are called "Building Bottlenecks," and where potential hazards would break out in the case of high gathering crowd. In an urgent situation, crowd streams from spacious areas rushing here would hinder the positive streams flowing out, increase crowd density, and then form an irregular succession of arch-like blockings at exit bottlenecks (Figure 1(b)), which instantaneously causes a cease in evacuation flow, and when the arches break, pedestrians will suddenly leave just like avalanche bunches [15, 16]. More seriously, this phenomenon can cause people psychological panic and induce collective human behavior, a crowd stampede, which is one of the most disastrous forms, often leading to fatalities as people are crushed or trampled [17].

2.2. Parameters Analysis. During the last few decades, pedestrians flow and evacuation have attracted researchers' attention. They concluded that egress evacuation capacity is related to egress width, w, m, egress stream flow-rate, f, persons/(ms), crowd velocity, v, m/s, and crowd density, ρ, persons/m^2.

Egress stream flow-rate is the number of persons passing through per unit of distance width of evacuation passage or exit per unit of time, person/(ms), crowd velocity is moving distances per unit of time, m/s, and crowd density is the number of persons per unit area, persons/m^2.

It is essential to understand the main characteristics of their relationship among the above parameters. In the past, researches about these parameters were mainly through two kinds of means: field observation and video record. So far, Predteckenskii and Milinski [18], Fruin [19], Ando et al. [20], Paul [21], Nelson and Maclennan [22], Smith and Petly [23], and Lu [24] et al., have already accumulated a large amount of observation data, from which typical relationship curves between the crowd density and velocity was summarized by Thompson (Figure 2) [25]. In general, in different environments and with different observation methods, each result of fleeing velocity varies slightly because of all kinds of constraints, but the changing trend of these curves is similar; that is, an increase in crowd streams density makes the distance between people decrease, and makes personnel movement velocity slow; on the contrary, a decrease in the density makes personnel movement velocity increase.

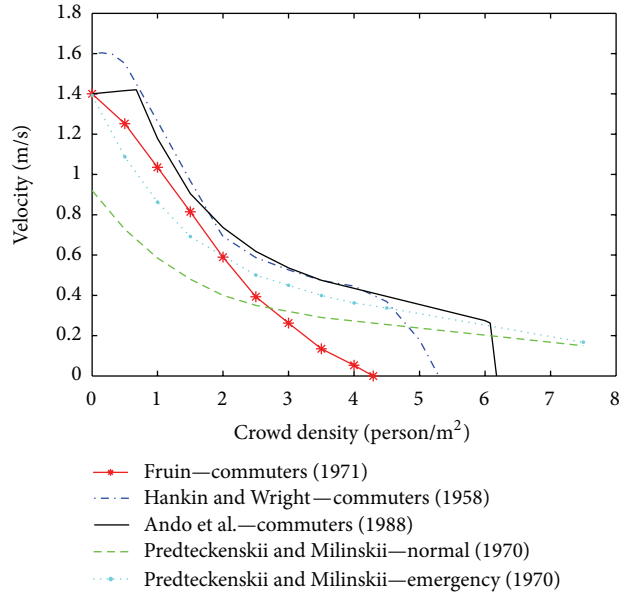


FIGURE 2: The relationship of crowd density and velocity.

Based on influence mechanism of personnel acceleration that is produced by people around in urgency evacuation from buildings, Lu et al. [24] proposed a velocity dynamics equation under crowding environment, which is a function of density:

$$v(\rho) = v_m(\alpha A + \beta B + \gamma),$$
$$A = 1.32 - 0.82\ln(\rho), \quad (1)$$
$$B = 3.0 - 0.76\rho,$$

where v_m is fleeing velocity, m/s, generally taken as 1.669 m/s, the range of α is 0.25~0.44, β is 0.014~0.088, and γ is 0.15~0.26.

Egress stream flow-rate depends on velocity and density, showed in (2):

$$f(\rho) = v(\rho)\rho. \quad (2)$$

Then the number of people evacuating through exit or passage is expressed as

$$Q = f(\rho) wt = v(\rho)\rho wt, \quad (3)$$

where w is the width of the evacuation exit or passage, m and t is evacuation time, s.

Figure 3 shows the essential relationship of three parameters [24]. v_0 is free flow velocity, m/s, $f_{\max}$ is the maximum flow-rate value, person/(ms), and ρ_c and v_c are critical density and velocity, respectively, persons/m^2, m/s. When crowd density is lower than the critical density, flow-rate increases as density increases, pedestrian flow in free flow state, but when crowd density is greater than the critical density, flow-rate decreases as density increases, pedestrian flow in blocked state, where flow-rate and velocity decrease and density increases. ρ_s is the safe density; in this situation, crowd is gathering but slowly flowing, and $\rho_{\max}$ is the maximum density; flow-rate and velocity all equal zero, in a completely jammed situation [26].

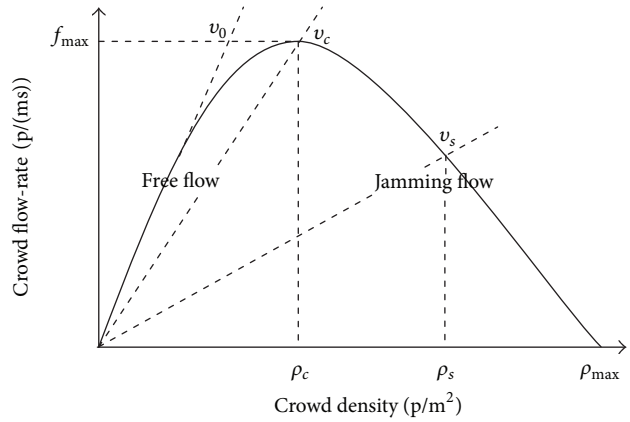


FIGURE 3: The essential relationship of crowd flow-rate, velocity and density.

The Green Guide [26] defines the limit of 4 persons/m^2 as a safe density for gathering crowd, that is, ρ_s = 4 persons/m^2, and proposes $\rho_{\max}$ = 7.4 persons/m^2. Under extreme conditions, density of 15 persons/m^2 has been observed. According to the Chinese physical characteristic, we choose ρ_s = 3.57 persons/m^2 as a safe density for a moving crowd and $\rho_{\max}$ = 8 persons/m^2 as a maximum density [27] in the jammed situation.

3. Calculation Principle

Streams of pedestrians evacuating towards a certain direction and the whole evacuation path inevitably encounter changed building space, such as path nodes, where streams merged into one, narrowed egresses, stairs, doors, steps, and so forth; these bottlenecks possibly become stranded zones due to clogging. Figure 4 shows a sketch of crowd flow. Assuming all leaving crowd will flow through the bottleneck into the next space, we define people who continue to flow in the zone as inflow crowd and out of the zone as outflow crowd; if jamming happens, crowd staying the zone as stranded crowd. Obviously, the balance principle of import and export applied and the number of stranded crowd should be the difference between numbers of inflow and outflow.

3.1. The Number of Inflow Crowd. Assuming several pedestrians' streams from internal passages merge into the bottleneck and then walk towards the next egress, the number of inflow crowd can be

$$Q_i = \sum_{i=1}^{k} \int_0^t v_i(t) w_i \rho_i(t)\, dt, \quad (4)$$

where Q_i is the number of inflow crowd from branch entrances; $v_i(t)$ and $\rho_i(t)$ are outflow velocity and density of branch entrances No. i at time t, m/s and persons/m^2; w_i is the width of entrances No. i, m; k is the number of branch entrances; t is the evacuation time, s.

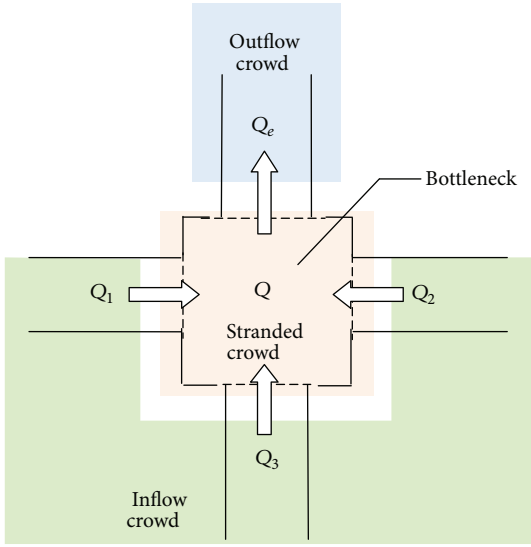


Figure 4: Sketch of crowd flow.

3.2. The Number of Outflow Crowd. During the time t, the evacuation crowd number of flowout of the bottleneck is

$$Q_e = \sum_{i=1}^{k} \int_0^{t_0} v_i(t)\, w_i \rho_i(t)\, dt + \int_{t_0}^{t} v(t)\, w \rho(t)\, dt, \quad (5)$$

where Q_e is the number of outflow crowd; t_0 is the time that the first person reaches the bottleneck exit, s; $v(t)$ is the outflow velocity of the exit at time t, m/s; w is the width of total exit, m; $\rho(t)$ is the density of bottleneck zone in front of the exit at time t, persons/m^2.

3.3. The Number of Stranded Crowd. Till the evacuation ends, the number of stranded crowds in the bottleneck zone Q is

$$\begin{aligned} Q = Q_i - Q_e &= \sum_{i=1}^{k} \int_0^{t} v_i(t)\, w_i \rho_i(t)\, dt \\ &\quad - \left(\sum_{i=1}^{k} \int_0^{t_0} v_i(t)\, w_i \rho_i(t)\, dt + \int_{t_0}^{t} v(t)\, w \rho(t)\, dt \right) \quad (6) \\ &= \sum_{i=1}^{k} \int_{t_0}^{t} v_i(t)\, w_i \rho_i(t)\, dt - \int_{t_0}^{t} v(t)\, w \rho(t)\, dt. \end{aligned}$$

3.4. The Total Evacuation Time. When crowd streams all flow out of the exit to the next space, the evacuation ends. The time is

$$t_e = t_0 + \frac{\left(Q_z - \sum_{i=1}^{k} \int_0^{t_0} v_i(t)\, w_i \rho_i(t)\, dt \right)}{v' f' w}, \quad (7)$$

where Q_z is the number of total evacuation people; v', f' are the outflow velocity and flow-rate of the exit at free flow, m/s, person/(ms).

The model is supposed that, when the evacuation velocity of the exit is equal to zero, no one is moving, and the exit is absolutely blocked, crowd density in the bottleneck is very high and arch-like clogging emerges, which must be prevented. This model also can calculate the smallest safe egress width based on the safe density and can calculate total evacuation time.

4. Application Example

In some stands of Tianjin Olympic Center Stadium as an example, evacuation routes include three longitudinal passages, one transverse passage, the hall in front of the egress, and the egress. The hall in front of the egress is a typical building bottleneck and becomes our main study field. The passage exit structure is shown in Figure 5.

4.1. Parameters Selection. Assuming the discussed stands have a seating capacity of 1400; that is, $Q_z = 1400$; all persons are evenly distributed in the stands. Three longitudinal passages have the same width w_l, $w_l = 1.1$ m, enough to allow up to two spectators to walk simultaneously. Transverse passage width w_t is $w_t = 1.6$ m, stranded-crowd zone width w is $w = 3.3$ m and egress width w_e, is a varying value.

The distance from one longitudinal passage entrance to the stranded-crowd zone is l_t, $l_t = 3$ m. The length from the egress to the edge of transverse passage is l_l, $l_l = 1.4$ m. The area of stranded-crowd zone is A, $A = (w \times (l_l + w_t))$ m^2.

According to the Design Code for Sports Building (JCJ31-2003) of China [28], assuming one stream of crowd flow is 40 persons/min and one crowd stream breadth is 0.55 m, then flow-rate is

$$f = \frac{40}{(60 \times 0.55)} = 1.2 \text{ person/(ms)}. \quad (8)$$

In this paper, we ignore the time of spectators leaving seats to longitudinal passages and only consider the process of spectators beginning to flow into the transverse passage from stands entrances.

4.2. The Number of Stranded Crowd. At the very beginning of the evacuation, because the number of people flowing into the bottleneck is small, they can freely flow out of the egress. As people numbers increase, stranded crowd appears.

The number of stranded crowd can be calculated as follows:

(1) $t \le t_0$ $(t_0 = L/v_0)$

$$\begin{aligned} Q(t) &= \sum_{i=1}^{k} \int_0^{t} v_i(t)\, w_i \rho_i(t)\, dt \\ &= \sum_{i=1}^{k} \int_0^{t} f_i(t)\, w_i dt = k f w_l t, \end{aligned} \quad (9)$$

where $Q(t)$ is the initial number of stranded crowd in the bottleneck during the time $t \le t_0$; v_0 is the desired velocity

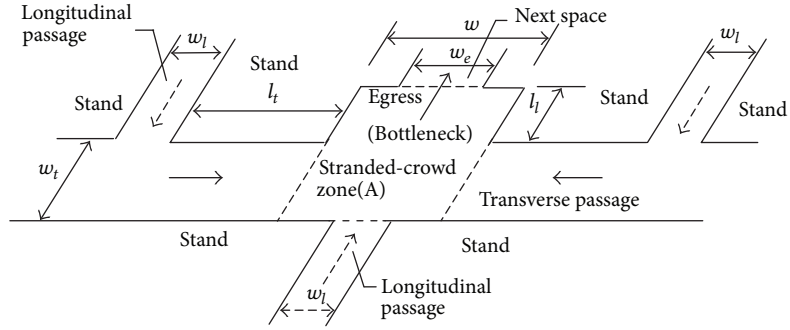


FIGURE 5: The plan of stands exit of Tianjin Olympic Center Stadium.

during the evacuation, $v_0 = 1.5$ m/s; L is the nearest distance from stands entrances to the bottleneck egress, $L = l_l + w_t$; f is inflow rate, $f = 1.21$ person/(ms) and $k = 3$;

(2) $t_0 < t \le t_1$

$$Q_e(t) = \int_{t_0}^{t} v(t)\, w\rho(t)\, dt = v(t)\,\rho(t)\, w_e t, \tag{10}$$

$$\rho(t) = \frac{Q(t)}{A} = \frac{Q(t)}{(w \times (l_l + w_t))}. \tag{11}$$

Because this model is based on crowd dynamics, the relationship between crowd density and crowd velocity of the egress of the bottleneck as shown in (1) is used, which is more suitable for the high-density crowd of buildings. We adopt the parameters from HMSO [24]; that is, $\alpha = 0.32$, $\beta = 0.021$, and $\gamma = 0.25$, and substituting (1) into (10), then

$$\begin{aligned} Q_e(t) &= v(t)\,\rho(t)\, w_e t \\ &= \left(1.669\left(0.32\left(1.32 - 0.83\ln\rho(t)\right)\right.\right. \\ &\quad \left.\left. +0.021\left(3 - 0.76\rho(t)\right) + 0.25\right)\right)\rho(t)\, w_e t \\ &= \left(1.23\rho(t) - 0.44\rho(t)\ln\rho(t) - 0.03\rho(t)^2\right) w_e t, \\ Q(t) &= \sum_{i=1}^{k} \int_0^t v_i(t)\, w_i \rho_i(t)\, dt \\ &\quad - \int_{t_0}^{t} v(t)\, w\rho(t)\, dt \\ &= Q(t-1) + k w_l f \Delta t - Q_e(t-1). \end{aligned} \tag{12}$$

At time t_1, the egress crowd velocity is equal to zero, which means no one can pass through the egress. $Q_e(t)$ is the number of outflow crowd during the time $t_0 < t \le t_1$; Δt is the time step, because $\rho(t)$ and $v(t)$ are variables of time, and $Q(t)$ is the dynamic value with time, and $Q(t-1)$, $Q_e(t-1)$ are the number of stranded and outflow crowd at the prior time, the time $t - 1$.

5. Results and Analysis

5.1. The Curves of the Number of Stranded Crowd and Outflow Crowd. From the above calculation, the curves of the number of stranded crowd and outflow crowd varying with time are obtained and compared with four different egress widths ($w_{e1} = 1.1$ m, $w_{e2} = 1.8$ m, $w_{e3} = 2.2$ m, and $w_{e4} = 3.3$ m).

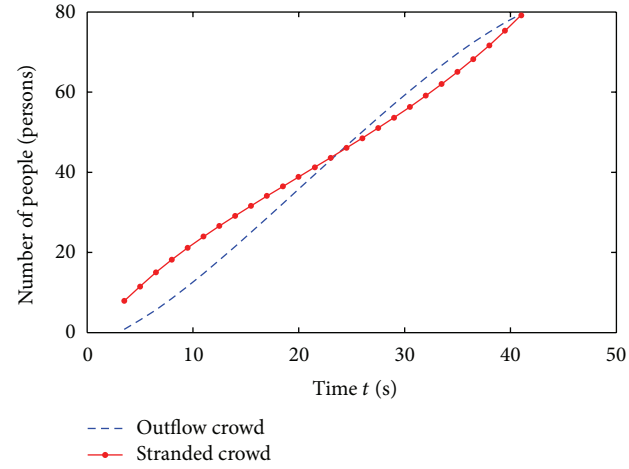


FIGURE 6: The curves of people number ($w_{e1} = 1.1$ m).

(1) $w_{e1} = 1.1$ m: Figure 6 shows that the changes of two curves are not linear, and their gradients are similar. In the initial time, stranded-crowd number in the bottleneck is greater than outflow number, but when the time $t = 24.5$ s, they are equal; then outflow number is greater than stranded-crowd number. When v' (the velocity of outflow) approaches zero, stranded-crowd number increases quickly, and the density of the zone increases up to $\rho = 8$ (p/m^2); at that time, $v' = 0$, $t = 41$ s, and the egress is jammed.

(2) $w_{e2} = 1.8$ m: Figure 7 shows that the gradients of two curves are apparently different. Outflow crowd number soon exceeds the stranded and becomes linear, but the number of stranded-crowd has been slow in growth until the time $t = 104$ s, and the rate of stranded-crowd number growth begins to have a sensible rise. When the time $t = 135.5$ s, $v' = 0$, and $\rho = 8.71$ (p/m^2), the egress is jammed.

(3) $w_{e3} = 2.2$ m: Figure 8 shows the curve of outflow crowd number is simply linear and increasing rapidly, and the curve of the stranded number is linear initially. However, when the time $t = 20$ s, the curve is horizontal, which means the stranded number is constant and the density of bottleneck $\rho = 2.2$ (p/m^2). In that situation, we consider the crowd is safe, and evacuation crowd could freely flow out without any obstacle; $v' = 0.82$ m/s.

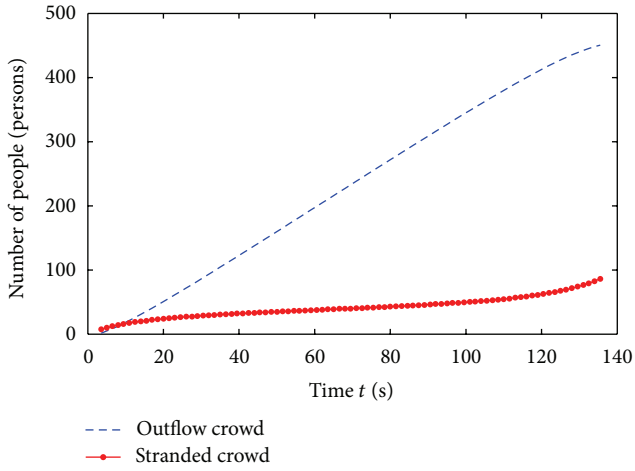


FIGURE 7: The curves of people number (w_{e2} = 1.8 m).

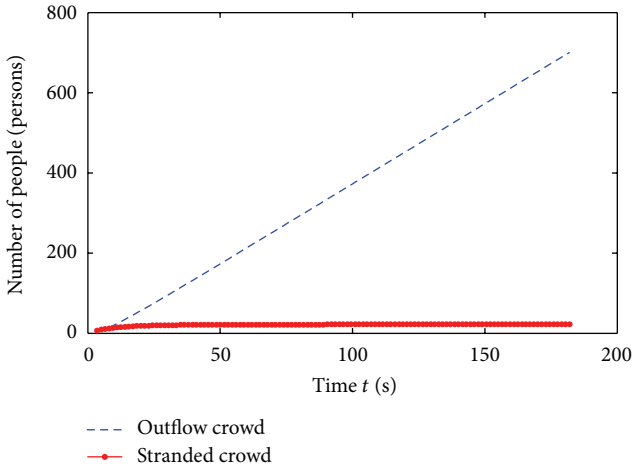


FIGURE 8: The curves of people number (w_{e3} = 2.2 m).

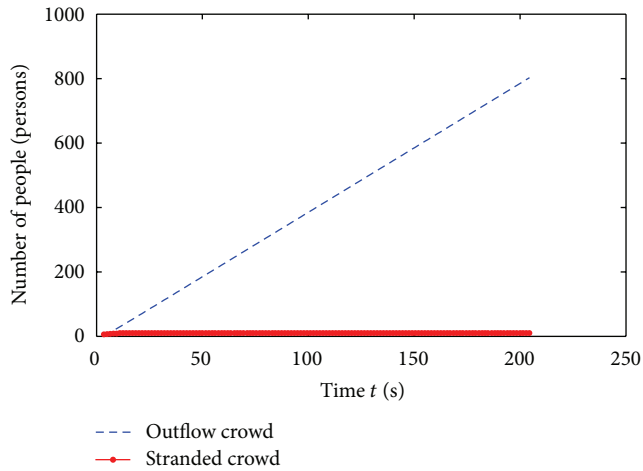


FIGURE 9: The curves of people number (w_{e4} = 3.3 m).

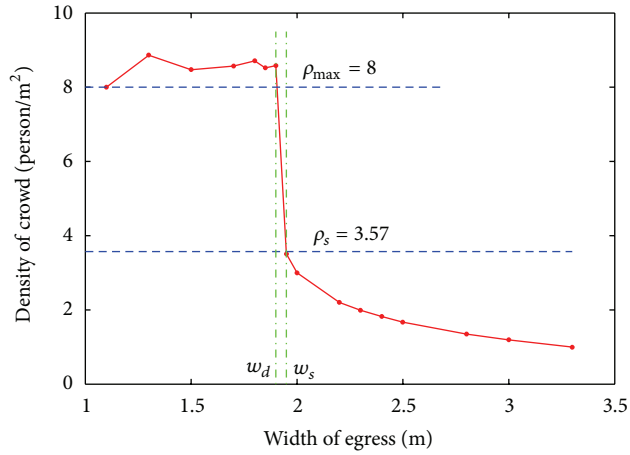


FIGURE 10: Relationship between crowd density and egress width.

(4) w_{e4} = 3.3 m: The egress width in a stadium should be greater than or equal to the sum of the widths of served stand passages or egresses [28]; so we choose $w_{e4} = w = 3.3$ m for calculation and that is equivalent to the egress completely opened. Figure 9 shows the calculation results are similar to the curves when the width w_{e3} = 2.2 m, except for when the density decreases, $\rho = 1$ (p/m^2), $v' = 1.2$ m/s. All evacuation is not obstructed at all. In addition, the results are according to the regulation of the Design Codes for Sports Building (JCJ31-2003) of China [28], which further verifies this model's effectiveness.

5.2. The Relationship of Crowd Density and Egress Width in the Bottleneck. Comparing the density with the egress width from 1.8 m to 2.2 m in the bottleneck, we observed that it slips very fast (Table 1); so we consider there may be a critical width which is the threshold value of safe width. In order to solve this problem, the density with different egress widths is calculated, and the results are shown in Figure 10.

From the above analysis, ρ_s, ρ_{max}, the corresponding safe width w_s, and dangerous width w_d are labeled, respectively, in Figure 10.

When $\rho \geq \rho_{max}$ and $w_d \leq 1.9$ m, the density of the crowd is extremely high, and the physical movement of pedestrians is almost impossible. This situation is terribly dangerous, because while pedestrians are still standing, when crushing occurs, the high pressure within the crowd compresses the lungs of some individuals in the crowd, producing fatalities from asphyxiation [17]. So the width of egress must be bigger than w_d, and inspection of the bottleneck density is very significant.

When $\rho_s \leq \rho < \rho_{max}$, $1.95 < w \leq 1.9$ m; although the width increases a little, the density decreases significantly, which seemingly small changes to the architecture of the surroundings can have such a large impact on the behavior of crowd, particularly when a crowd is panicked. After the density suddenly inclines up to $\rho_s = 3.57$ persons/m^2, with the width increasing, density is smoothly decreasing; so the safe width, $w_s = 1.95$, is the threshold width value, which does not lead to high density in theory.

Table 1: Parameters with different width.

Egress width (m)	Jamming time (s)	Crowd density (p/m^2)	Velocity (m/s)
1.1	41	8	0
1.8	136	8.71	0
2.2	—	2.2	0.82
3.3	—	1	1.2

6. Conclusion

The people evacuation calculation model, based on the crowd dynamics when the exit jammed, is presented in this paper. The main characteristics of the evacuation streams, people's safety under emergency conditions, and comparison of the results using dynamics formula with different egress widths for the movement in a stadium example are all discussed.

Safe egress width is within a stone's throw of dangerous egress width, when a crowd is panicked. Small changes to the architecture of the surroundings can have large effects on the behavior of crowd.

This stranded-crowd model can calculate total evacuation time, which will be carried out in future work.

References

[1] Y. Liu, D. M. Liu, B. Norman et al., "Analysis of evacuation performance of merging points in stadiums based on crowd simulation," in *Proceedings of the 12th Conference of the In-ternational Building Performance Simulation Association (IBPSA, BS'11)*, pp. 2651–2658, November 2011.

[2] X. Pan, C. S. Han, K. Dauber, and K. H. Law, "Human and social behavior in computational modeling and analysis of egress," *Automation in Construction*, vol. 15, no. 4, pp. 448–461, 2006.

[3] S. Chen, Y. Zheng, C. Cattani, and W. Wang, "Modeling of biological intelligence for SCM system optimization," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 769702, 10 pages, 2012.

[4] S. Wen, W. Zheng, J. Zhu, X. Li, and S. Chen, "Elman fuzzy adaptive control for obstacle avoidance of mobile robots using hybrid force/position incorporation," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 42, no. 4, pp. 603–608, 2012.

[5] K. J. Zhu and L. Z. Yang, "The effects of exit position and internal layout of classroom on evacuation efficiency," *Acta Physica Sinica*, vol. 59, no. 11, pp. 7701–7707, 2010.

[6] W. G. Song, Y. F. Yu, and T. Chen, "Influences of exit condition on pedestrian evacuation," *Fire Safety Science*, vol. 12, no. 2, pp. 100–105, 2003.

[7] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4288, 1995.

[8] M. Isobe, T. Adachi, and T. Nagatani, "Experiment and simulation of pedestrian counter flow," *Journal of Fire Protection Engineering*, vol. 4, no. 3, pp. 81–92, 1992.

[9] V. Papinigis, E. Geda, and K. Lukošius, "Design of people evacuation from rooms and buildings," *Journal of Civil Engineering and Management*, vol. 16, no. 1, pp. 131–139, 2010.

[10] S. Chen, W. Huang, C. Cattani, and G. Altieri, "Traffic dynamics on complex networks: a survey," *Mathematical Problems in Engineering*, vol. 2012, Article ID 732698, 23 pages, 2012.

[11] C. Cattani, R. Badea, S. Chen, and M. Crisan, "Biomedical signal processing and modeling complexity of living systems," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 298634, 2 pages, 2012.

[12] Q. S. Zhang, M. Liu, C. H. Wu, and G. Zhao, "A stranded-crowd model (SCM) for performance-based design of stadium egress," *Building and Environment*, vol. 42, no. 7, pp. 2630–2636, 2007.

[13] Q. S. Zhang, M. Liu, J. Liu, and G. Zhao, "Modification of evacuation time computational model for stadium crowd risk analysis," *Process Safety and Environmental Protection*, vol. 85, no. 6, pp. 541–548, 2007.

[14] Q. Liu, H. Yang, H. P. Lu, and S. Jing, "Study on the dispersal of people flow in the stadium," *China Civil Engineering Journal*, vol. 37, no. 10, pp. 92–98, 2004.

[15] S. Y. Chen, Y. H. Wang, and C. Cattani, "Key issues in modeling of complex 3D structures from video sequences," *Mathematical Problems in Engineering*, vol. 2012, Article ID 856523, 17 pages, 2012.

[16] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487–490, 2000.

[17] R. S. C. Lee and R. L. Hughes, "Exploring trampling and crushing in a crowd," *Journal of Transportation Engineering*, vol. 131, no. 8, pp. 575–582, 2005.

[18] V. M. Predtechenskii and A. I. Milinski, *Planning for Foot Traffic Flow in Buildings*, Stroiizdat, Moscow, Russia, 1969.

[19] J. J. Fruin, *Pedestrian Planning and Design*, Metropolitan Association of Urban Designers and Environmental Planners, New York, NY, USA, 1971.

[20] K. Ando, H. Ota, and T. Oki, "Forecasting the flow of people," *Railway Research Review*, vol. 8, no. 45, pp. 8–14, 1988.

[21] J. L. Paul, "Effective width model for evacuation flow in buildings," *Society of Fire Protection Engineers*, vol. 4, no. 6, pp. 215–232, 1980.

[22] H. E. Nelson and H. A. Maclennan, *Emergency Movement*, NFPA, Quincy, Mass, USA, 1995.

[23] K. Smith and D. N. Petley, *Environmental Hazards Assessing Risk and Reducing Disaster*, Routledge, London, UK, 5th edition, 2008.

[24] J. A. Lu, Z. Fang, Z. M. Lu et al., "Mathematical model of evacuation speed for personnel in buildings," *Engineering Journal of Wuhan University*, vol. 35, no. 2, pp. 66–70, 2002.

[25] P. A. Thompson and E. W. Marchant, "Computer and fluid modelling of evacuation," *Safety Science*, vol. 18, no. 4, pp. 277–289, 1995.

[26] *Guide to Safety at Sports Grounds (Green Guide)*, Football Licensing Authority Department for Culture, Media and Sport (DCMS), TSO (The Stationery Office), 5th edition, 2008.

[27] J. Chen, P. J. Zhang, W. Yang, and Q. Li, "Personnel evacuation strategy for theater based on systematic dynamics model," *Journal of Natural Disasters*, vol. 14, no. 6, pp. 125–132, 2005.

[28] *Design Code for Sports Building (JCJ31-2003) of China*, The Industry Standard of the People's Republic of China, Beijing, China, 2003.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 564214, 7 pages
http://dx.doi.org/10.1155/2013/564214

Research Article

Human Model Adaptation for Multiview Markerless Motion Capture

Dianyong Zhang,[1,2] Zhenjiang Miao,[1,2] and Shengyong Chen[3]

[1] *Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China*
[2] *Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China*
[3] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Dianyong Zhang; 07112064@bjtu.edu.cn

Received 28 November 2012; Accepted 17 January 2013

Academic Editor: Carlo Cattani

An approach to automatic modeling of individual human bodies using complex shape and pose information. The aim is to address the need for human shape and pose model generation for markerless motion capture. With multi-view markerless motion capture, three-dimensional morphable models are learned from an existing database of registered body scans in different shapes and poses. We estimate the body skeleton and pose parameters from the visual hull mesh reconstructed from multiple human silhouettes. Pose variation of body shapes is implemented by the defined underlying skeleton. The shape parameters are estimated by fitting the morphable model to the silhouettes. It is done relying on extracted silhouettes only. An error function is defined to measure how well the human model fits the input data, and minimize it to get the good estimate result. Further, experiments on some data show the robustness of the method, where the body shape and the initial pose can be obtained automatically.

1. Introduction

In the area of computer vision, researchers have been working for many years to obtain and analyze real 3D information of an object. Lots of new methods have been explored by researchers in the recent years [1–5]. As a special subject, the detection and recovery of human shapes and their 3D poses from images or videos are important problems in computer vision area. There are many potential applications in diverse fields such as motion capture, interactive computer games, industry design, sports or medical purpose, interfaces for human computer interaction (HCI), surveillance, and robotics. Accurate detailed human model can be used for markerless motion capture tracking a subject individual model, which includes information on both body shape and pose. Model-based motion capture is especially suited to markerless motion capture because it can constrain the search space by defining the degrees of freedom of the human skeleton. Initialization of human motion capture always requires the definition of humanoid model approximating the shape, appearance, kinematic structure, and initial pose of the subject to be tracked.

A lot of model-based pose estimation algorithms are based on minimizing an error function that measures how well the 3D model fits the images. So, the good initialization result is very important. The majority of algorithms for human pose estimation use a manually initialized generic model with limb lengths and shapes which approximate the individual. Automate initialization will improve the quality of tracking result. Accurate initialization is base of the right pose estimation. A limited number of researchers do the automate initialization to improve the tracking result by reconstructing 3D model of the subject from a single or multiple view images. Detailed 3D human shape and pose estimations from multi-view images are still difficult problems that does not exist a satisfactory solution.

In this paper, as illustrated in Figure 1, we use multi-views to recover a detailed 3D representation of the person. Our parametric representation of the body is based on a 3D morphable human model with an underlying predefined

(a) (b) (c) (d) (e)

Figure 1: Overview of our system workflow: (a) input multi-view images, (b) visual hull mesh from multi images and get the skeleton, (c) original 3D morphable model projected to the image, (d) shape and pose parameters to fit the images, and (e) output estimated model.

skeleton. The models are generated from a template mesh given by a deformable human model that is learned for a scan database [6] over 550 full body 3D scans taken of 114 undressed subjects. Previous contribution [7] estimates human shape and pose from multiple camera views using the popular SCAPE model, but it does not provide an underlying skeleton. We propose a method to automatically adjust a morphable human body model to fit the first frame of the markerless motion capture data. The goal of our job is to obtain a human body model as much adjusted as possible to the individual subject we are going to track. The estimated refined shape and skeleton pose serve as initialization for the next frame to be tracked.

The remaining sections of this paper are organized as follows. In the following section, we will present the relevant previous work on detailed human model estimation. In Section 3, we describe the morphable model with PCA method and the predefined skeleton information. In Section 4, we present the visual hull model and get the initial pose parameters. In Section 5, we present the fitting algorithm in details. In Section 6, we demonstrate the estimation result of our solution. We will conclude this paper in Section 7.

2. Related Works

The papers [9, 10] present comprehensive survey of existing related techniques in motion capture research area of computer vision. A model-based markerless motion capture system can be broken down into four processes: initialization, tracking, pose estimation, and recognition. The initialization step is concerned with two things: the initial pose of a subject and the model representing the subject. Shape and pose initialization can be obtained by manual adaptation or using automatic methods, the latter methods still have some limitations, such as the requirement of a specific pose or predefined motion. The priori model can be of several kinds: kinematic skeleton, shape, and color priors. Many approaches employ kinematic body models, it is hard for them to capture motion, let alone detailed body shape. The majority of algorithms continue to use a manually initialized generic model with limb lengths and shapes which approximate the individual. Because of the difference of vision, the manual initialization often cannot get the well initialization result. Over the past 10 years, there have been substantial research works [11–14] in the automatic initialization of human model shape from multiple view images. For improved accuracy in tracking, these approaches reconstruct an articulated model which approximates the shape of a specific subject. Because of few images, people cannot get the accurate body shape information, and furthermore, the shape of the subject can differ from person to person.

Our goal is to get the detailed shape and pose body model fast and accurately. Body shape is estimated from a single-view or a multi-view, which is performed by projecting the 3D model onto the images and building a cost function that minimizes the distance between the projected model and the 2D silhouettes. A popular parametric model SCAPE (Shape Completion and Animation for People) [15] is a data-driven method for building body shapes with different poses and individual body shapes. This model has recently been adopted as morphable model to estimate human body shape from monocular or multi-view images [7, 16–19]. Balan et al. [7] have adopted this model closer to observed silhouettes to capture more detailed body deformations.

Most recently, the approach has been used to infer pose and shape from a single image. Guan et al. [17] have considered more visual cues, shading cues, internal edges, and silhouettes to fit the SCAPE model to an uncalibrated single image with the body height constrained. Sigal et al. [16] describe a discriminative model based on the mixture of experts to estimate SCAPE model parameters from monocular and multicameras image silhouettes. Chen et al. [20] proposed a probabilistic generative method that models 3D deformable shape variations and infers 3D shapes form a single silhouette. They use nonlinear optimization to map the 3D shape data into a low-dimensional manifold, expressing shape variations by a few latent variables. Pons-Moll et al. [21] proposed a hybrid tracker approach that combined correspondence based local optimization with five inertial sensors placed at a human body, although they can obtain a much accurate and detailed human tracker, they need additional sensors, and they use the 3D scan model. Lee et al. [22] present a near-automatic method to get a 3D face model by PCA according to the set of input silhouettes. However, for the initialization phase, the pose is known approximately as the subject adopts a specified position in most of papers, such as T position or others. Many approaches still require manual or semiautomatic model positioning for bootstrapping the algorithm.

Benjamin et al. [14] presented a method for generic pose initialization, and it is for the skeleton pose and conclude the

coarse body shape information. Gall et al. [12] introduce an approach for global optimization that is for human motion capturing called as interacting simulated annealing (ISA), they use prior knowledge learned for training motion data as the soft constraint. Many algorithms are based on minimizing an error function that measures how well 3D model fits the image. Our approach is most similar to the work of Jain et al. [23] and Balan et al. [7], but they both estimate pose parameters simultaneously and this will cost more time than estimating pose and shape parameters separately.

Our main contribution is proposing a new approach to automatic generation 3D shape and pose model fitting the multicameras images using a 3D morphable model. In our system, we get the individual human model with the same pose as the subject we are going to track automatically. Human shape and pose are captured by multiple synchronized and calibrated cameras. And we use background subtraction algorithm to extract body silhouettes [24]. The overview of our system is showed in Figure 1.

3. 3D Morphable Model

Principal component analysis (PCA) is a popular statistical method to extract the most salient directions of data variation from large multidimensional datasets. Our morphable model is based on scan database of [6] over 550 full body 3D scans taken of 114 undressed subjects. All subjects are scanned in a based pose, some subjects are scanned in 9 poses chosen randomly from a set of 34 poses. We apply the PCA model to this database [6], a new human shape model can be generated by learning a linear mapping between shape parameters and PCA weights. Therefore, a human model is given by

$$M(\alpha) = m_0 + \sum_{i=1}^{n} \alpha_i m_i, \tag{1}$$

where the human shape parameter is $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, m_i is the ith eigen human model, and m_0 is the mean or average human model. Similar to [23, 25], the morphable model is a combination of a bone skeleton with joints. Like Jain et al. [23] and Gall et al. [25], we drive the body pose by a defined underlying skeleton manually, shown in Figure 2. And the shape parameters can be described by PCA parameters. In our paper, we define 20 human PCA components like [23]. We define a kinematic chain, so the motion of body model can be parameterized by the joint angles. For many years, kinematic chains are widely used in human tracking and motion capture systems. The mesh deformation can be controlled by linear blend skinning (LBS) technique. If V_i is the position of vertex i, T_j is the transformation of the bone j, and $\omega_{i,j}$ is the weight of the j bone for vertex i, LBS gives the position of the transformed vertex i as

$$v_i' = \sum_j \omega_{i,j} \left(T_j v_i\right). \tag{2}$$

The bone weights for the vertices mean how much each bone transform affects each vertex. These weights are normalized such that $\sum \omega_{i,j} = 1$.

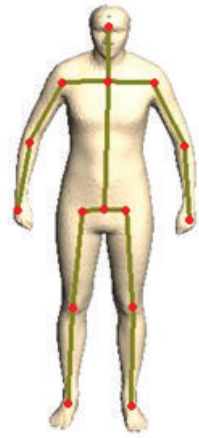


Figure 2: The template morphable human model with the underlying skeleton.

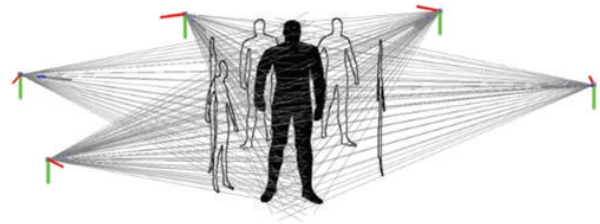


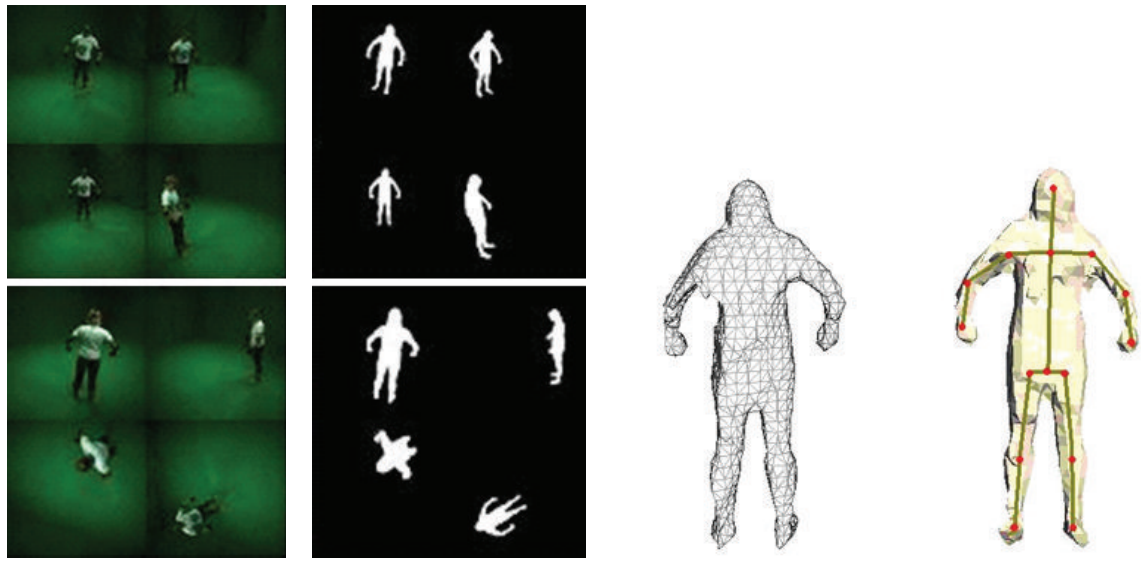
Figure 3: The concept of visuall hull reconstruction, picture taken from [8].

4. The Visual Hull and Pose Initialization

The pose of a human body model is modelled by a set of joint angles. The pose is computed using the skeleton curve from the visual hull based image silhouettes.

The visual hull is the volume obtained from the interaction of the silhouette cones. The visual hull for the image sequences analyzed is shown in Figure 3. Camera numbers and placements highly influenced the reconstruction quality of the visual hull. 3D volumetric reconstruction from images has long been an important and active research topic. Volumetric approach has been widely used to reconstruct the visual hull from silhouettes for its simplicity and robustness. We build our body model to fit multi-view images, and it just can be used for the markerless motion capture. Not like most of researches, people usually build the body visual hull model using voxel data. We want to get the visual hull mesh model. The visual hull construction process can be shown in Figure 4.

Though the visual hull model from the silhouettes tends to overestimate the real body volume of the subject, we can still compute the skeleton pose using it. We use the Baran's algorithm [26] to rig the visual hull mesh model. We define a template skeleton and adjust the skeleton to fit inside the mesh by resizing and positioning its bones and joints. This adjustment can be treated as an optimization problem, which is to compute the skeleton adjustment that can fit better



(a) (b) (c) (d)

Figure 4: Processing images to compute silhouettes and visual hull mesh: (a) input multi-view images, (b) segmentation result, (c) reconstruction visual hull, and (d) extract the skeleton and joints.

inside the mesh while maintaining as much as possible its resemblance with the defined template skeleton.

The recovered position and joint angles of the visual hull can be used to initialize the optimization of the 3D morphable model pose parameters. The number of joints extract from visual hull mesh is 18, we just consider 15 joints not include the foot and waist joints. Like Pishchulin et al. [27], we retarget the skeleton of 3D morphable model to the extracted 3D pose by computing inverse kinematics through minimizing the Euclidean distance between a set of corresponding 3D joint positions: namely, head, neck, thorax, pelvis, left/right knees, ankles, hips, elbows, and wrists. In Section 5, we describe the method of getting the shape parameters in details.

5. Fitting Human Model to Images

In this section, we describe our method for fitting model shape parameters to a set of input silhouette images. This step aims to deform the 3D morphable model to fit the given silhouettes. We assume the human initial pose of the articulated skeleton is known from Section 4. Then, the shape PCA parameters are the only variables that need to be estimated.

5.1. Correspondence between 2D Images and 3D Model. The fitting is fundamental to 2D-3D correspondences. Let $M(\alpha)$ be an individual human model given a vector parameter $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, the fitting procedure determines the parameter vector α so that the proposed 3D morphable model fits the given 2D silhouettes as well as possible. In the multiple cameras way, each camera has its own coordinates. We should transform the human body model local coordinates (x, y, and z) into the image coordinates (x', y'), this can be done by three steps: the first step is to transform the local coordinates into the world coordinates; the second is to transform the world coordinates into the camera coordinates and the last is to project the camera coordinates, into the image coordinates. The initial location parameters $o = (x_0, y_0, z_0)$, the center of the local human body coordinate in the world coordinates, which are estimated from the centroid of the visual hull generated from multiviews.

Let S^k, $k = 1 \cdots K$ be an input silhouette body image captured by camera C^k. $S_m^k(\alpha)$ is a silhouette image generated by projecting $M(\alpha)$ onto an image plane using the camera C^k calibration parameters.

5.2. Shape Deformation. If we define a cost function f that measures the difference of two silhouette images, our goal is to find α that minimizes this total penalty as follows:

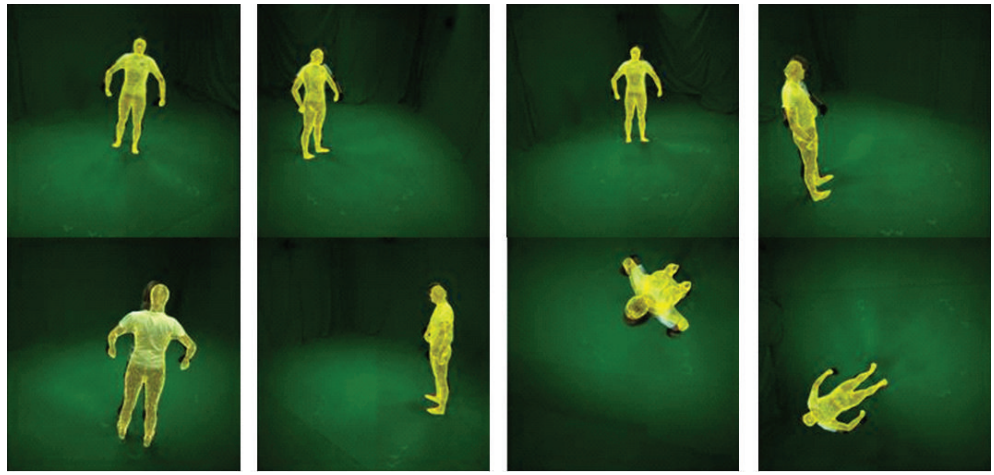
$$E_{\text{total}}(\alpha) = \sum_{k=1}^{K} f\left(S^k, S_m^k(\alpha)\right) \tag{3}$$

for a suitable cost function f.

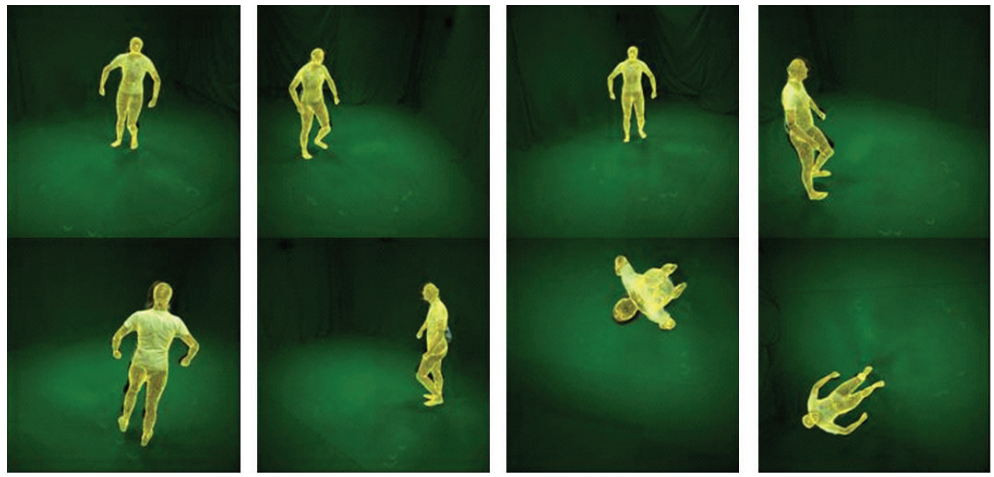
The solution of our problem can be expressed as a minimization of an error function depending on the correspondences shape parameters. We map the surface manifold of the vertex space of the body mesh onto the extract silhouette. And enable those pixels which correspond to contour vertices. We consider the problem of morphable model fitting body silhouette with image silhouettes.

We measure how close a given body shape hypothesis is to the input foreground silhouettes using the distance function between silhouette S and T like papers [7, 19], an asymmetric distance between them in this case,

$$\tilde{d}(S, T) = \frac{\sum_{i,j}\left(S_{i,j} D_{i,j}(T)\right)}{\left(\sum_{i,j} S_{i,j}\right)^{\delta}} \tag{4}$$



(a)



(b)

Figure 5: Shape estimation results based eight cameras: (a) fitting result of the first frame, (b) fitting result of the 25th frame.

for the numerator, where $S_{i,j} = 1$ for the pixels inside silhouette S otherwise $S_{i,j} = 0$. $D_{i,j}(T)$ is a distance transform function and when the pixel (i, j) inside T, the function $D_{i,j}(T) = 0$, for the points outside T, $D_{i,j}(T)$ is defined as the Euclidean distance to the closest point on the boundary of T. For the denominator, it is a normalization term based on the size of the silhouette S. As the Balan and Black [19], we set the $\delta = 1.5$ to achieve the effect of cameras depth invariance.

Then, the distance between the projection of the 3D model points and the image observation is minimized with the model PCA weights parameters α. We define the objective function like Balan and Black [19] as follows:

$$E_{\text{total}}(\alpha) = \sum_{k=1}^{K} \tilde{d}\left(S_m^k(\alpha), S^k\right) + \tilde{d}\left(S^k, S_m^k(\alpha)\right) \quad (5)$$

this function uses a symmetric distance to match the estimated and observed silhouettes over the K camera views, but we only consider the shape parameters not including the pose parameters. We minimize it just to optimize a set of 20 principal shape components from all camera perspectives.

6. Experimental Results

In this section, we provide experimental results for our 3D morphable model silhouette fitting process described in Section 5. We test our system using a database of MPI08 [21, 28] (hb data set) provided by the University of Hannover, Germany, a person is captured with 8 HD cameras with a resolution of $1004 * 1004$ pixels. The initial position of human model is computed from the visual hull centroid, and the human skeleton pose can be computed by the visual hull mesh. Our goal is to match the boundary contours between input and morphable models silhouettes. For computational efficientncy, we use the distance between the outline of one silhouette and the outline of the other and vice-versa can be used. Therefore, we have to measure the distance between the model contours and image contours. We choose the error function to be minimized, we fit our morphable model to images by iteratively minimizing the error in all views, and we estimate 20 PCA components for the 3D morphable human model. Figure 5 shows the final estimated results, we choose the first frame and the 25th frame to test.

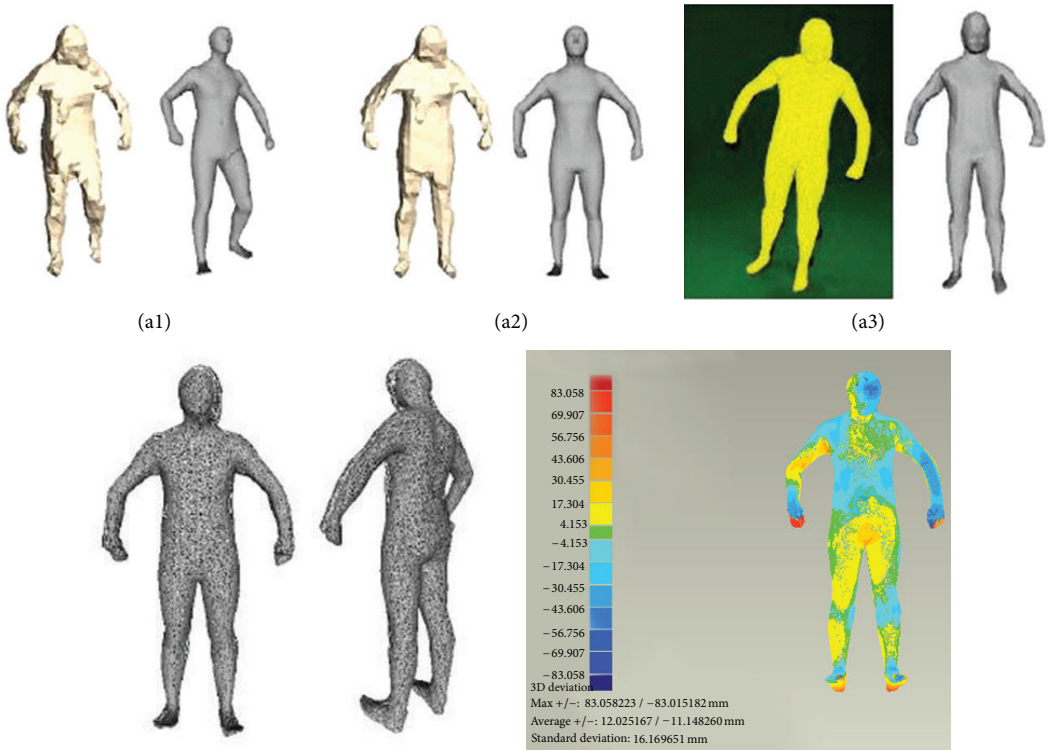


Figure 6: The compare result with visual hull and scan model: (a1) and (a2) are compared result with visual hull mesh, (a3) is the 3D scan model, (b) the scan model overlaid with the estimated model fit in, and (c) the compare difference with the scan model using corresponding mesh points, the unit is millimeter.

As we know, it is a difficult problem that estimates biometric parameters from a given 3D human model, and we do not know the ground truth values of this subject. We just compare it to the 3D scan model provided by Pons-Moll et al. [21, 28]. Figure 6 presents the compare result with the visual hull mesh and the scan model: the average distance is 0.075 mm, standard deviation is 16.16 mm, we can see that estimated human shape is very similar to the scan model except the hands part, for most of application, the hands parameter is not important.

7. Conclusion

Our system implements automatic initialization and gives right estimates. We have presented a method for estimating 3D human pose and shape from multi-view imagery. The approach based a learned 3D morphable human model using PCA method. The pose variation is due to the underlying skeleton. When we get the approximate body shape and pose we take into account not only the skeleton pose but also the nonrigid deformations of the human body shape. Then, we can obtain detailed human model shapes with full correspondence. The shape we computed can replace the 3D scan model in motion capture areas. However, in these cases the human body silhouette should be easy to extract from the image data. The subject should be in tight clothes in multi-view cameras (more than 5 cameras) and have no occlusion. And we do not use tracking algorithm, we fit the 3D morphable model for every frame independently. In the future work, we will consider more constraints on body shape, for the exposed parts of the body, the skin color can be constrained to add the fitting precision. And surface tracking will also be considered in the future.

Acknowledgments

The authors would like to thank Hasler [6], Gall [25], and Pons-Moll [21, 28] for providing their database for research purpose. This work is supported by the National Key Technology R&D Program of China (2012BAH01F03), National Natural Science Foundation of China (60973061, 61173096, R1110679), National 973 Key Research Program of China (2011CB302203), Doctoral Fund of Ministry of Education of China (20100009110004, 20113317110001), Beijing Natural Science Foundation (4123104), and Tsinghua-Tencent Joint Lab for IIT.

References

[1] F. Moreno-Noguer and P. Fua, "Stochastic exploration of ambiguities for non-rigid shape recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[2] M. Carlini and S. Castellucci, "Modelling and simulation for energy production parametric dependence in greenhouses,"

Mathematical Problems in Engineering, vol. 2010, Article ID 590943, 28 pages, 2010.

[3] D. Vazquez, A. M. Lopez, and D. Ponsa, "Unsupervised domain adaptation of virtual and real worlds for pedestrian detection," in *Proceedings of the 21st International Conference on Pattern Recognition*, pp. 3492–3495, Barcelona, Spain, November 2012.

[4] P. Fan, S.-Y. Chen, S.-J. Lu et al., "High dynamic range color image rendering with human visual adaptation simulation," in *Proceedings of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 234–237, Taipei, Taiwan, September 2009.

[5] D. Zhang, Z. Miao, and M. Wang, "Images based human volumetric model reconstruction and animation," in *Proceedings of the 2nd International Conference on Image Analysis and Signal Processing (IASP '10)*, pp. 210–213, April 2010.

[6] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H. P. Seidel, "A statistical model of human pose and body shape," *Computer Graphics Forum*, vol. 28, no. 2, pp. 337–346, 2009.

[7] A. Balan, L. Sigal, M. Black, J. Davis, and H. Haussecker, "Detailed human shape and pose from images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.

[8] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[9] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[10] M. Sunkel, B. Rosenhahn, and H. P. Seidel, "Silhouette-based generic model adaptation for marker-less motion capturing," in *Proceeding of the 2nd Conference on Human Motion: Understanding, Modeling, Capture and Animation*, pp. 119–135, 2007.

[11] J. Gall, B. Rosenhahn, and H. P. Seidel, "An introduction to interacting simulated annealing," in *Human Motion-Understanding, Modeling, Capture and Animation, Computational Imaging and Vision*, vol. 36, pp. 319–345, Springer, Berlin, Germany, 2008.

[12] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, and H. P. Seidel, "Multilinear pose and body shape estimation of dressed subjects from image sets," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 1823–1830, June 2010.

[13] R. Benjamin, C. Michel, and N. Vincent, "Generic initialization for motion capture from 3D shape," in *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR '10)*, pp. 306–315, June 2010.

[14] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: shape completion and animation of people," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 241–253, 2005.

[15] L. Sigal, A. O. Balan, and M. J. Black, "Combined discriminative and generative articulated pose and nonrigid shape estimation," in *Proceedings of the 21st Annual Conference on Neural Information Processing Systems (NIPS '07)*, December 2007.

[16] P. Guan, A. Weiss, A. O. Balan, and M. J. Black, "Estimating human shape and pose from a single image," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '09)*, pp. 1381–1388, September 2009.

[17] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han, "Parameter reshaping of human bodies in images," *ACM Transactions on Graphics*, vol. 29, no. 4, article 126, 2010.

[18] A. O. Balan and M. J. Black, "The nake truth: estimating body shape under clothing," in *Proceedings of the 10th European Conference on Computer Vision (ECCV '08)*, D. A. Forsyth, P. H. S. Torr, and A. Zisserman, Eds., vol. 5303 of *Lecture Notes in Computer Science*, pp. 15–29, Springer, Marseille, France, 2008.

[19] Y. Chen, T. Kim, and R. Cipolla, "Inferring 3D shapes and deformations from single views," in *Computer Vision—ECCV*, vol. 6313, pp. 300–313, Springer, Berlin, Germany, 2010.

[20] G. Pons-Moll, A. Baak, T. Helten, M. Müller, H. P. Seidel, and B. Rosenhahn, "Multisensor-fusion for 3D full-body human motion capture," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 663–670, June 2010.

[21] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju, "Silhouette-based 3D face shape recovery," in *Graphics Interface Proceedings*, pp. 21–30, Canadian Computer-Human Communications Society, 2003.

[22] A. Jain, T. Thormahlen, H. P. Seidel, and C. Theobalt, "Movie re-shape: tracking and reshaping of humans in videos," *ACM Transactions on Graphics*, vol. 29, no. 5, article 148, 2010.

[23] S. Y. Chen, J. H. Zhang, Y. F. Li, and J. W. Zhang, "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 118–127, 2012.

[24] J. Gall, C. Stoll, E. De Aguiar, C. Theobalt, B. Rosenhahn, and H. P. Seidel, "Motion capture using joint skeleton tracking and surface estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 1746–1753, June 2009.

[25] S. Corazza, L. Mündermann, A. M. Chaudhari, T. Demattio, C. Cobelli, and T. P. Andriacchi, "A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach," *Annals of Biomedical Engineering*, vol. 34, no. 6, pp. 1019–1029, 2006.

[26] I. Baran and J. Popovi, "Automatic rigging and animation of 3D characters," *ACM Transactions on Graphics*, vol. 26, no. 3, article 72, 2007.

[27] L. Pishchulin, A. Jain, M. Andriluka, T. Thormählen, and B. Schiele, "Articulated people detection and pose estimation: reshaping the future," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 3178–3185, June 2012.

[28] A. Baak, T. Helten, M. Mueller, G. Pons-Moll, H. P. Seidel, and B. Rosenhahn, "Analyzing and evaluating markerless motion tracking using inertial sensors," in *Proceedings of the 11th European Conference of Computer Vision (ECCV)*, Crete, Greece, September 2010.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 786749, 14 pages
http://dx.doi.org/10.1155/2013/786749

Review Article

Modeling for Deformable Body and Motion Analysis: A Review

Hailang Pan,[1] Hongwen Huo,[2] Guoqin Cui,[3] and Shengyong Chen[4]

[1] *School of Electronic and Optical Engineering, Nanjing University of Science and Technology, 200 Xiao Ling Wei, Nanjing 210094, China*
[2] *Information Technology Research Center, China Electronics Standardization Institute, No. 1 Andingmen East Street, Beijing 100007, China*
[3] *National Key Laboratory for Digital Multimedia Chip Technology, Vimicro Corporation, Beijing 100191, China*
[4] *College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Hailang Pan; panhailang@126.com

Received 5 December 2012; Revised 26 January 2013; Accepted 26 January 2013

Academic Editor: Carlo Cattani

This paper surveys the modeling methods for deformable human body and motion analysis in the recent 30 years. First, elementary knowledge of human expression and modeling is introduced. Then, typical human modeling technologies, including 2D model, 3D surface model, and geometry-based, physics-based, and anatomy-based approaches, and model-based motion analysis are summarized. Characteristics of these technologies are analyzed. The technology accumulation in the field is outlined for an overview.

1. Introduction

Human body modeling is experiencing a continuous and accelerated growth. This is partly due to the increasing demand from computer graphics and computer vision communities. Computer graphics pursues a realistic modeling of both the human body geometry and its associated motion. This will benefit applications such as games, virtual reality, or animations, which demand highly realistic human body models (HBMs).

Recently, computer vision has been used for the automatic generation of HBMs from a sequence of images by incorporating and exploiting prior knowledge of the human appearance. Computer vision also addresses human body modeling, but in contrast to computer graphics it seeks more for an efficient than an accurate model for applications, such as intelligent video surveillance, motion analysis, telepresence, or human-machine interface. Computer vision applications rely on vision sensors for reconstructing HBMs. Obviously, the rich information provided by a vision sensor, containing all the necessary data for generating a HBM, needs to be processed. Approaches such as tracking segmentation-model fitting or motion prediction-segmentation-model fitting or other combinations have been proposed showing different performances according to the nature of the scene to be processed (e.g., indoor environments, studio-like environments, outdoor environments, single-person scenes, etc.). The challenge is to produce a HBM able to faithfully follow the movements of a real person [1, 2].

Modeling a human is a great challenge if we consider the numerous parts needed to compose a body. The first step is the basic structure modeling, the definition of the joints, their positions, orientations, and the geometric model that will describe the body hierarchy. Next, we need to think about the body volume and on top of this, we can use a parametric surface to simulate skin for example. With these three elements, we can reach a good representation of a body.

Methods used to deform the human skin layer include curve-, contour-, surface-, geometry-, physics-, and anatomy-based approaches. 2D models such as curve and contour models offer representational and computational simplicity and are often preferred over 3D models for applications involving monocular images and video. These models typically represent the shape of the human body coarsely. None

of these methods explicitly models how clothing influences human shape. Surface models used in animation have a highly structured mesh to give high-resolution representation in areas of deformation and efficient representation in other areas. Preserving this vertex parameterization is important in reconstructing models that can be used for animation. Geometry-based approach such as free form deformation (FFD) provides flexibility for the users to control deformations of models. The method is simple and fast but requires considerable skill to model realistic model. Physics-based approaches such as the finite element method (FEM) can model skin layers according to their physical properties accurately. Anatomy-based methods create an accurate human body model based on a precise representation of the skeleton, muscles, and fatty tissues. These techniques generate realistic and dynamic deformation of an articulated body using physical simulation but, due to their high computational cost, applications are mainly in offline simulation and animation [3].

Model-based pose estimation algorithms aim at recovering human motion from one or more camera views and a 3D model representation of the human body. The model pose is usually parameterized with a kinematic chain, and thereby the pose is represented by a vector of joint angles. The majority of algorithms are based on minimizing an error function that measures how well the 3D model fits the image. This category of algorithms usually has two main stages, namely, defining the model and fitting the model to image observations. The model image association problem for pose estimation is usually formulated as the minimization/maximization of an error/likelihood function. The two main strategies have been described, namely, local and particle-based optimization. Local optimization methods are faster and more accurate but in practice, if there are visual ambiguities, or really fast motions, the tracker might fail catastrophically. To achieve more robustness, particle filters can be used because they can represent uncertainty through a rigorous Bayesian paradigm [4].

In the past, there are several reviews on the study of human motion capture, detection, and analysis. Aggarwal et al. [5] wrote a review about articulated and elastic nonrigid motion in 1994. There is no other review about non-rigid, elastic, or deformable human motion analysis in the past 19 years. This paper is the first such review after 1995.

This paper is organized as follows Section 2 presents the 2D curve and contour human models. Section 3 presents the 3D human surface models such as quadrics, superquadrics, implicit surface, spline surface, and mesh surface models. Section 4 presents geometry-based human models. Section 5 presents physics-based human models. Anatomically based human models are presented in Section 6.

2. Two-Dimensional Models

2.1. Curve Models. Tabia et al. proposed an approach to matching 3D objects in the presence of nonrigid transformation and partially similar models [6]. They adopt square-root elastic (SRE) framework because it simplifies the elastic shape analysis. They define a space of closed curves of interest, impose a Riemannian structure on this space using the elastic metric, and compute geodesic paths under this metric. These geodesic paths can then be interpreted as optimal elastic deformations of curves. Mori and Malik [7] took a single two-dimensional image containing a human figure, locate the joint positions, and use these to estimate the body configuration and pose in three-dimensional space (see Figure 1(a)). They match the input image to each stored view using the technique of shape context matching in conjunction with a kinematic chain-based deformation model and then use joint positions to estimate the body configuration and pose in three-dimensional space. In their approach, a shape is represented by a discrete set of n points sampled from the internal and external contours on the shape. They first perform edge detection on the image, using the boundary detector to obtain a set of edge pixels on the contours of the body. They then sample some number of points (300–1000) from these edge pixels to be used as the sample points for the body. Srivastava et al. [8] introduced a square-root velocity (SRV) representation for analyzing shapes of curves in Euclidean spaces using an elastic metric (see Figure 1(b)). Huang et al. [9] presented a variational and statistical approach for shape registration (see Figure 1(c)). Shapes of interest are implicitly embedded in a higher-dimensional space of distance transforms. In this implicit embedding space, registration is formulated in a hierarchical manner: the mutual information criterion supports various transformation models and is optimized to perform global registration; then, a B-spline-based incremental free form deformations (IFFD) model is used to minimize a sum-of-squared-differences (SSD) measure and further recover a dense local nonrigid registration field.

2.2. Contour Models. Liu et al. [10] presented boosted deformable model for human body alignment. Their model representation consists of a shape component represented by a point distribution model and an appearance component represented by a collection of local features, trained discriminatively as a two-class classifier using boosting (see Figure 2(a)). Freifeld et al. [11] defined a new "contour person" model of the human body that has the expressive power of a detailed 3D model and the computational benefits of a simple 2D part-based model. The contour person (CP) model is learned from a 3D SCAPE model of the human body that captures natural shape and pose variations; the projected contours of this model, along with their segmentation into parts, form the training set (see Figure 2(b)). The CP model factors deformations of the body into three components: shape variation, viewpoint change, and part rotation. This latter model also incorporates a learned non-rigid deformation model. Zuffi et al. [12] defined a new deformable structures (DS) model that is a natural extension of previous pictorial structures (PS) models and that captures the non-rigid shape deformation of the parts (see Figure 2(c)). Each part in a DS model is represented by a low-dimensional shape deformation space, and pairwise potentials between parts capture how the shape varies with pose and the shape of neighboring parts. A key advantage of such a model is that it more accurately models object boundaries. This enables

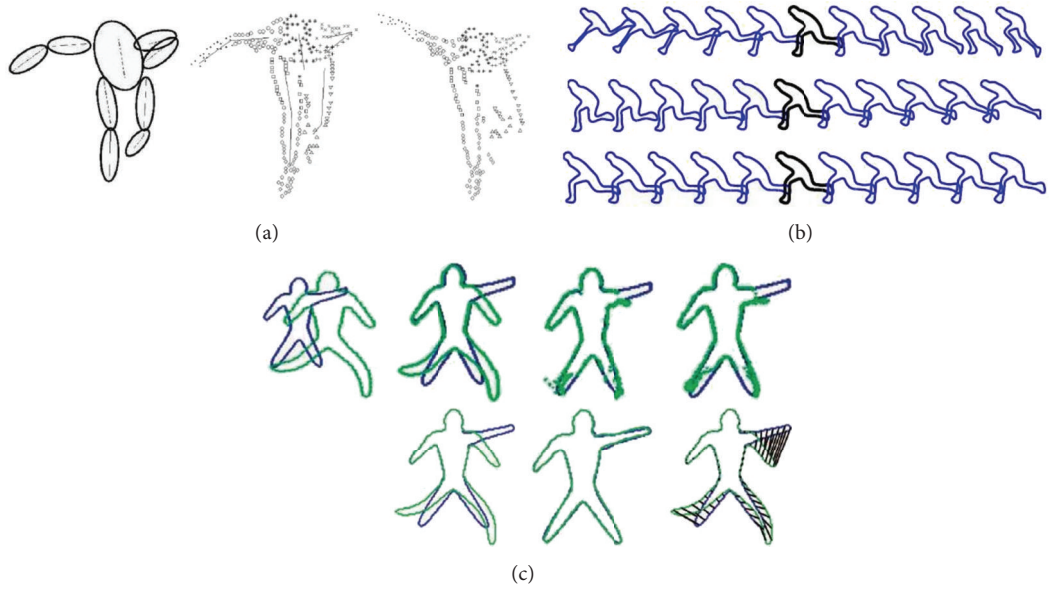


Figure 1: Motion analysis based on 2D curve models in [7–9].

image likelihood models that are more discriminative than the previous PS likelihoods. They focus on a human DS model learned from 2D projections of a realistic 3D human body model and use it to infer human poses in images using a form of nonparametric belief propagation. Guan et al. [13] studied detection, tracking, segmentation, and pose estimation of people in monocular images. They start with a contour person (CP) model (see Figure 2(d)), which is a low-dimensional, realistic, parameterized generative model of 2D human shape and pose. The CP model is learned from examples created by 2D projections of multiple shapes and poses generated from a 3D body model such as SCAPE. The CP model is based on a template, corresponding to a reference contour that can be deformed into a new pose and shape. This deformation is parameterized and factors the changes of a person's 2D shape due to pose, body shape, and the parameters of the viewing camera. This factorization allows different causes of the shape change to be modeled separately.

3. 3D Surface Models

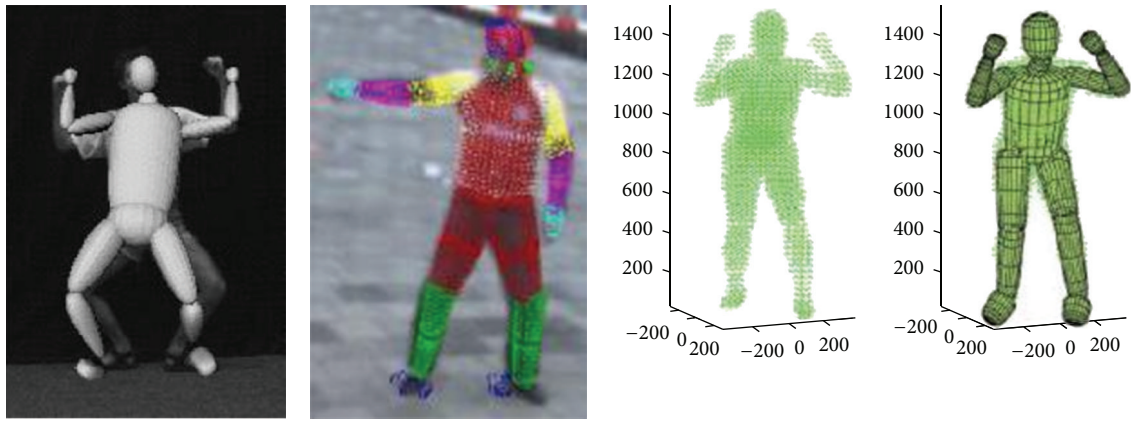
3.1. Quadrics. Park and Hodgins [14] presented a technique for capturing and animating those motions using a commercial motion capture system and approximately 350 markers. They supplement these markers with a detailed, actor-specific surface model (see Figure 3(a)). The motion of the skin can then be computed by segmenting the markers into the motion of a set of rigid parts and a residual deformation (approximated first as a quadratic transformation and then with radial basis functions). Fayad et al. [15] proposed a more general shape model that accounts for quadratic deformations. Their approach takes motion capture (MOCAP) data as input and enables the extraction of more accurate estimates for the rigid component of the different body segments using a factorization framework. The parameters of the model are computed using a Levenberg-Marquardt nonlinear optimization scheme. Hyun et al. [16] presented a new approach to the modeling and deformation of a human or virtual character's arms and legs. Each limb is represented as a set of ellipsoids of varying sizes interpolated along a skeleton curve (see Figure 3(b)). A base surface is generated by approximating these ellipsoids with a swept ellipse, and the difference between that and the detailed shape of the arm or leg is represented as a displacement map. Pan and Liu [17] presented a model of elastic articulated objects based on revolving conic surface and a method of model-based motion estimation (see Figure 3(c)). The model includes 3D object skeleton and deformable surfaces that can represent the deformation of human body surfaces. In each limb, surface deformation is represented by adjusting one or two deformation parameters. Then, the 3D deformation parameters are determined by corresponding 2D image points and contours with volume invariable constraint. The 3D motion parameters are estimated based on the 3D model.

3.2. Superquadrics. Terzopoulos and Metaxas [18] presented a physically based approach to fitting complex three-dimensional shapes using a class of dynamic models that can deform both locally and globally. They formulate the deformable superquadrics, which incorporate the global shape parameters of a conventional superellipsoid with the local degrees of freedom of a spline. The model's six global deformational degrees of freedom capture gross shape features from visual data and provide salient part descriptors for efficient indexing into a database of stored models. The local deformation parameters reconstruct the details of complex shapes that the global abstraction misses. The equations of motion which govern the behavior of deformable superquadrics make them responsive to externally applied forces. The authors fit models to visual data by transforming the data into forces and simulating the equations of motion through time to adjust the translational, rotational, and

Figure 2: Motion analysis based on 2D contour models in [10–13].

Figure 3: Motion analysis based on 3D quadrics models in [14, 16, 17].

deformational degrees of freedom of the models. Sminchisescu [19] built a human body model which consists of a kinematic "skeleton" of articulated joints controlled by angular joint parameters, covered by "flesh" built from superquadric ellipsoids with additional tapering and bending parameters (see Figure 4(a)). A typical model has around 30 joint parameters, plus 8 internal proportion parameters encoding the positions of the hip, clavicle, and skull tip joints, plus 9 deformable shape parameters for each body part, gathered into a vector. A complete model can be encoded as a single large parameter vector. During tracking, they usually estimate only joint parameters, but during initialization the most important internal proportions and shape parameters are also optimized, subject to a soft prior based on



(a) (b) (c)

Figure 4: Motion analysis based on 3D superquadrics models in [19–21].

standard humanoid dimensions and updated using collected image evidence. Although this model is far from being photo realistic, it suffices for high-level interpretation and realistic occlusion prediction, and it offers a good trade-off between computational complexity and coverage. Hofmann and Gavrila [20] presented an approach for 3D human body shape model adaptation to a sequence of multi-view images. They use an articulated model with linearly tapered superquadrics as geometric primitives for torso, neck, head, upper and lower arm, hand, upper and lower leg, and foot, assuming body symmetry (see Figure 4(b)). The parameter space of each superquadric comprises parameters for length, squareness, and tapering. They implement automatic pose and shape estimation using a three-step procedure: first, they recover initial pose over a sequence using an initial (generic) body model. Both model and poses then serve as input to the above-mentioned adaptation process. Finally, a more accurate pose recovery is obtained by means of the adapted model. Sundaresan and Chellappa [21] proposed a general approach using Laplacian Eigenmaps and a graphical model of the human body to segment 3D voxel data of humans into different articulated chains. They select the superquadric model to represent human bodies (see Figure 4(c)). They use a hierarchical approach beginning with a skeletal model (joint locations and limb lengths) and then proceeding to increase the model complexity and refining parameters to obtain a volumetric model (superquadric parameters). Yang and Lee [22] reconstructed a 3D human body pose from stereo image sequences based on a top-down learning method. The 3D human body model consists of 17 body components. The human body model has 40 degrees of freedom (DOF). Tapered superquadrics are employed to represent body components.

3.3. Implicit Surface. Matsuda and Nishita [23] modeled the human body by layered metaballs, which correspond to the horizontal cross section of the body, in their cloth simulation system. For each cross section, metaballs are generated by measured sample points on the boundary of the cross section. In order to fit the metaball surface with the sampling points, they employed the steepest descent method. For body deformation, the sampling points on the cross section are smoothly moved using Bezier curves. Blinn [24] presented a new algorithm applicable to other functional forms, in particular to the summation of several Gaussian density distributions. He models human body by using implicit surface (see Figure 5(a)), but he did not model from images. Thalmann et al. [25] presented different methods for representing realistic deformations for virtual humans with various characteristics: sex, age, height, and weight. Their methods based on a combination of metaballs and splines could be applied to frame-by-frame computer generated films and virtual environments. Smooth implicit surfaces, known as metaballs, are attached to an articulated skeleton of the human body and are arranged in an anatomically based approximation. This particular human body model includes 230 metaballs. D'Apuzzo et al. [26] outlined techniques for fitting a simplified model to the noisy 3D data extracted from the images and a new tracking process based on least squares matching is presented. They present a simplified model of a limb. Ellipsoidal metaballs are used to simulate the gross behavior of bone, muscle, and fat tissue. Only three ellipsoidal metaballs are attached to each limb skeleton and arranged in an anatomically based approximation (see Figure 5(b)). Each ellipsoidal metaball has four deformation parameters and each limb has three ellipsoidal metaballs, so each limb has 12 deformation parameters. Fua et al. [27] presented a comprehensive concept to fit animation models to a variety of different data derived from multiimage video sequences. Their research includes setting up and calibrating a system of three CCD cameras, extracting image silhouettes, tracking individual key body points in 3D, and generating surface data by stereo or multi-image matching. To reduce the number of degrees of freedom (DOFs) and to be able to robustly estimate the skeleton's position, they replace the multiple metaballs by one ellipsoid attached to each bone in the skeleton (see Figure 5(c)). Plänkers and Fua [28] developed a framework for 3D shape and motion recovery of articulated deformable objects. They propose a formalism that incorporates the use of implicit surfaces into earlier robotics approaches that

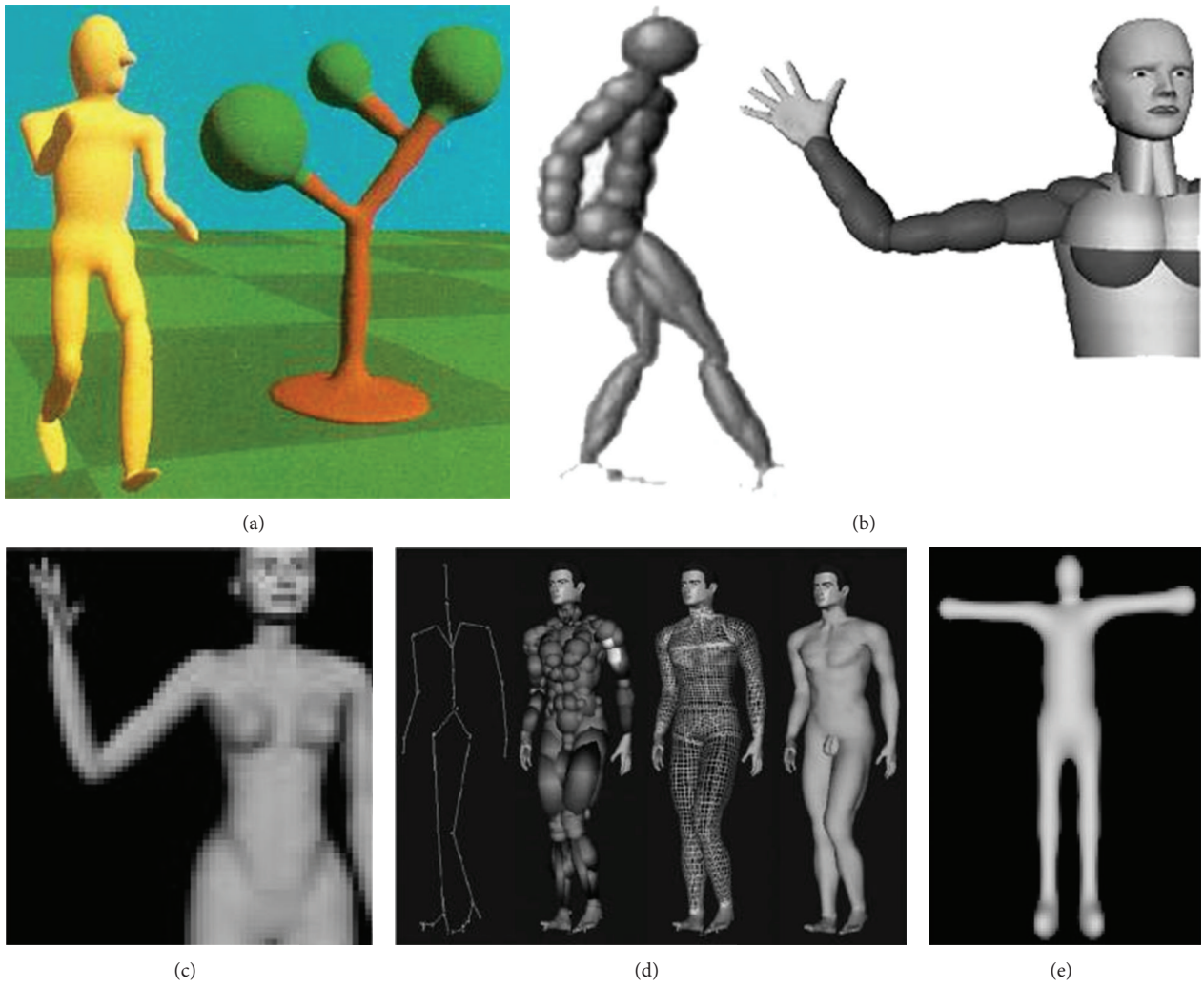
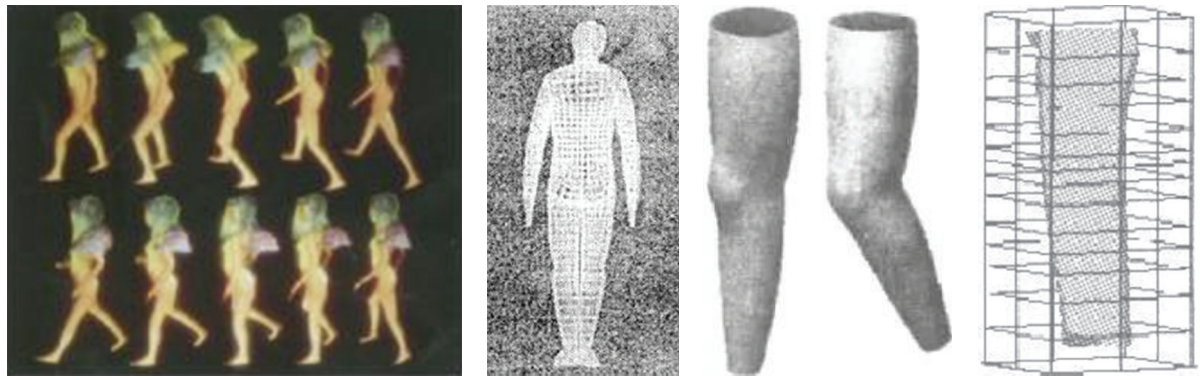


Figure 5: Motion analysis based on implicit surface models in [24, 26–29].

is designed to handle articulated structures. Their human body model also includes 230 metaballs (see Figure 5(d)). Tong et al. [29] constructed a human body model using convolution surface with articulated kinematic skeleton (see Figure 5(e)). The human body's pose and shape in a monocular image can be estimated from convolution curve through nonlinear optimization.

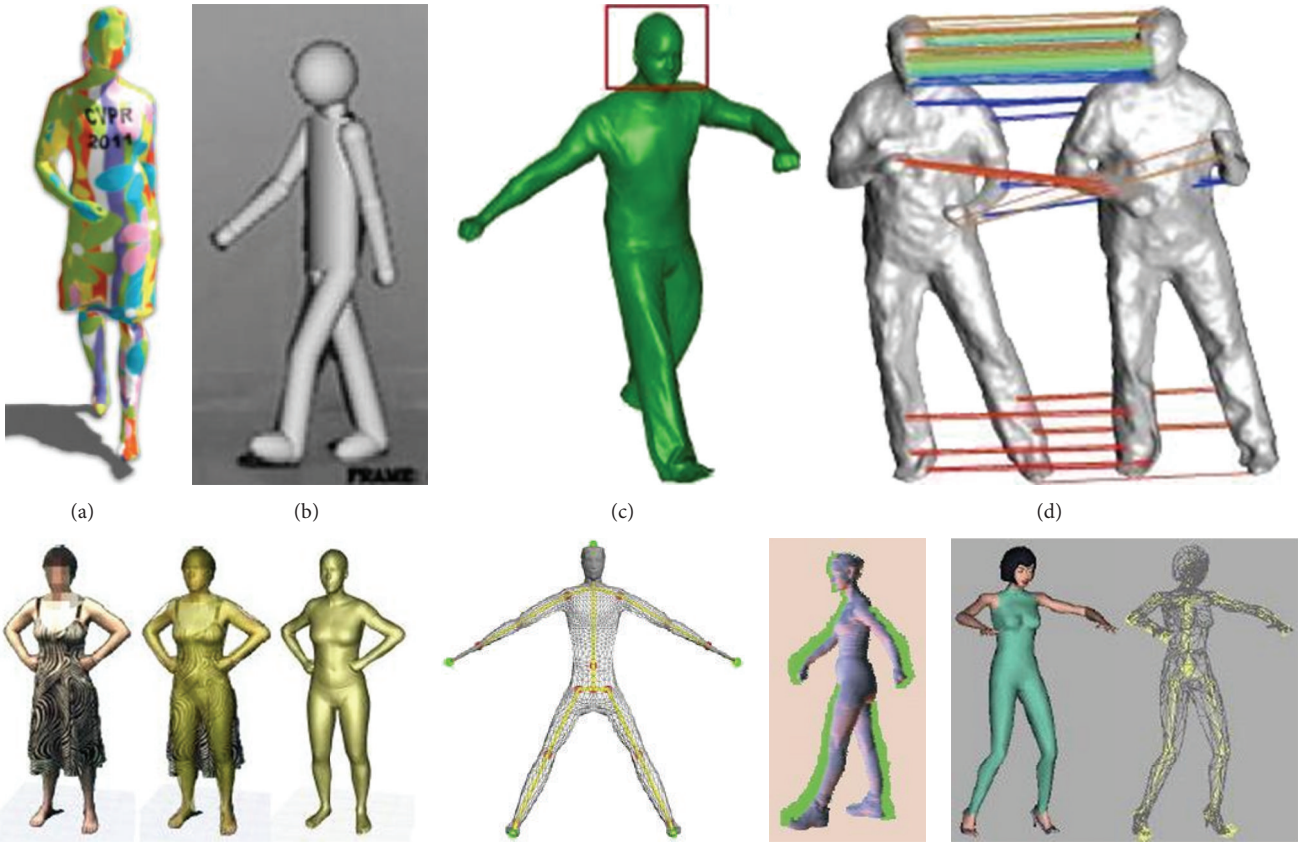
3.4. Spline Surface. Nahas et al. [30] described how the use of B-spline surfaces allows lissome movements of body and face (see Figure 6(a)). Their method is empirical, based on a parametrical animation. It can be combined with a muscles model for animation. Fu and Yuan [31] introduced the establishment of human body based on B-spline (see Figure 6(b)). According to the military criteria, they divide the whole body into sixteen limbs and use the method of multiplying of matrix to establish the equation of human body's movement. It constructs the blending surface between two limbs by transfinite interpolant. Huang et al. [32] discussed a motion modeling method to simulate the bend of human leg and corresponding deformations of muscles on the basis of NURBS FFD (free form deformation) [33]. Generally, FFD uses a trivariate tensor product spline to transmit deformations, and it is feasible to choose only two order B-spline basis functions in this case (see Figure 6(c)). According to the anatomic structure of joints, the simulation formulas are presented, and they coincide with the motion characteristics of knee joint and muscles. Wang and Jiang [34] simulated three-dimensional human's leg bending by using free form deformation on the basis of NURBS (see Figure 6(d)); leg looked as combo of rigid body bone and flexible body muscle. It improves Barr's deformation methods, weight is used to control deformation, and a good visual effect of simulation is achieved.

3.5. Mesh Surface. Huang et al. [35] considered the problem of aligning multiple non-rigid surface mesh sequences into a single temporally consistent representation of the shape and motion (see Figure 7(a)). A global alignment graph structure is introduced, which uses shape similarity to identify frames for intersequence registration. Graph optimization is performed to minimize the total non-rigid deformation required



(a) (b) (c) (d)

Figure 6: Motion analysis based on spline surface models in [30–32, 34].



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 7: Motion analysis based on mesh models in [35–41, 43].

to register the input sequences into a common structure. Chang and Lin [36] presented a 3D model-based tracking algorithm called the progressive particle filter to decrease the computational cost in high degrees of freedom by employing hierarchical searching. A 3D virtual human model is developed to simulate human movement. The proposed 3D human model is constructed from deformable flesh (see Figure 7(b)). Deformable flesh can be deformed to precisely fit the target body to achieve accurate tracking results. Liao et al. [37] reconstructed complete 3D deformable models over time by a single depth camera, provided that most parts of the models are observed by the camera at least once. A mesh warping algorithm based on linear mesh deformation is used to align different partial surfaces. A volumetric method is then used to combine partial surfaces, fix missing holes, and smooth alignment errors (see Figure 7(c)). Varanasi et al. [38] addressed the problem of surface tracking in multiple camera environments and over time sequences. In order to

fully track a surface undergoing significant deformations, they cast the problem as a mesh evolution over time. Such an evolution is driven by 3D displacement fields estimated between meshes recovered independently at different time frames. The contribution is a novel mesh evolution-based framework that allows to fully track, over long sequences, an unknown surface encountering deformations, including topological changes (see Figure 7(d)).

Balan and Black [39] estimated the detailed 3D shape of a person from images of that person wearing clothing. They employ a parametric body model called SCAPE that is able to capture variability of body shapes between people, as well as articulated and non-rigid pose deformations. The model is derived from a large training set of human laser scans, which have been brought in full correspondence with respect to a reference mesh (see Figure 7(e)). Starck and Hilton [40] presented a model-based approach to recover animated models of people from multiple view video images. A prior humanoid surface model is first decomposed into multiple levels of detail and represented as a hierarchical deformable model for image fitting. A novel mesh parameterization is presented that allows propagation of deformation in the model hierarchy and regularization of surface deformation to preserve vertex parameterization and animation structure (see Figure 7(f)). De Aguiar et al. [41] jointly captured the motion and the dynamic shape of humans from multiple video streams without using optical markers. Their approach uses a deformable high-quality mesh of a human as scene representation. It jointly uses an image-based 3D correspondence estimation algorithm and a fast Laplacian mesh deformation scheme to capture both motion and surface deformation of the actor from the input video footage (see Figure 7(g)). Wang et al. [42] developed an efficient and intuitive deformation technique for virtual human modeling by silhouettes input. With their method, the reference silhouettes and the target silhouettes are used to modify the synthetic human model, which is represented by a polygonal mesh. The system moves the vertices of the polygonal model so that the spatial relation between the original positions and the reference silhouettes is identical to the relation between the resulting positions and the target silhouettes. Their method is related to the axial deformation. Seo et al. [43] aimed to carry out realistic deformations on the human body models as well as make its usage simple. Their system is composed of several modules: skin attachment to an H-Anim skeleton is carried out first in order to get deformation in skeletal shape modification as well as in animation; volumetric deformation module deals with the volumetric scale of body parts (see Figure 7(h)). These deformation operators, together with the skeletal deformation, allow the automatic adaptation of the body model to different sizes and proportions to accommodate anthropometrical variations; surface optimization is used to simplify the model in consideration of not only geometric features, but also the animation aspect of it; finally, the BDP (MPEG-4 format) generation module describes the geometry of the model as well as how to animate it according to the MPEG-4 BDP specifications.

4. Geometry-Based Approaches

Kokkinos et al. [44] presented intrinsic shape context (ISC) descriptors for 3D shapes (see Figure 8(a)). They generalize to surfaces the polar sampling of the image domain used in shape contexts: for this purpose, they chart the surface by shooting geodesic outwards from the point being analyzed; "angle" is treated as tantamount to geodesic shooting direction and radius as geodesic distance. To deal with orientation ambiguity, they exploit properties of the Fourier transform. For the analysis of deformable 3D shapes, Raviv et al. [45] introduced an (equi)affine invariant diffusion geometry by which surfaces that go through squeeze and shear transformations can still be properly analyzed (see Figure 8(b)). The definition of an affine invariant metric enables them to construct an invariant Laplacian from which local and global geometric structures are extracted. Castellani et al. [46] exploited a new generative model for encoding the variations of local geometric properties of 3D shapes. Surfaces are locally modeled as a stochastic process, which spans a neighborhood area through a set of circular geodesic pathways, captured by a modified version of a Hidden Markov Model (HMM), named multicircular HMM (MC-HMM). The approach proposed consists of two main phases: local geometric feature collection and MC-HMM parameter estimation. Akhter et al. [47] proposed a dual approach to describe the evolving 3D structure in trajectory space by a linear combination of basis trajectories. They describe the dual relationship between the two approaches, showing that they both have equal power for representing 3D structure. They further show that the temporal smoothness in 3D trajectories alone can be used for recovering nonrigid structure from a moving camera (see Figure 8(c)). The principal advantage of expressing deforming 3D structure in trajectory space is that they can define an object on independent basis. This results in a significant reduction in unknowns and corresponding stability in estimation. Gotardo and Martinez [48] addressed the classical computer vision problems of rigid and nonrigid structure from motion (SFM) with occlusion. They assume that the columns of the input observation matrix W describe smooth 2D point trajectories over time. They then derive a family of efficient methods that estimate the column space of W using compact parameterizations in the Discrete Cosine Transform (DCT) domain. In non-rigid SFM, they propose a 3D shape trajectory approach that solves for the deformable structure as the smooth time trajectory of a single point in a linear shape space.

Raviv et al. [49] presented a generalization of symmetries for non-rigid shapes and a numerical framework for their analysis (see Figure 8(d)), addressing the problems of full and partial exact and approximate symmetry detection and classification. Zhu et al. [50] formulated a hierarchical configurable deformable template (HCDT) to model articulated visual objects—such as horses and baseball players—for tasks such as parsing, segmentation, and pose estimation. HCDTs represent an object by an AND/OR graph where the OR nodes act as switches, which enables the graph topology to vary adaptively. This hierarchical representation is compositional, and the node variables represent positions

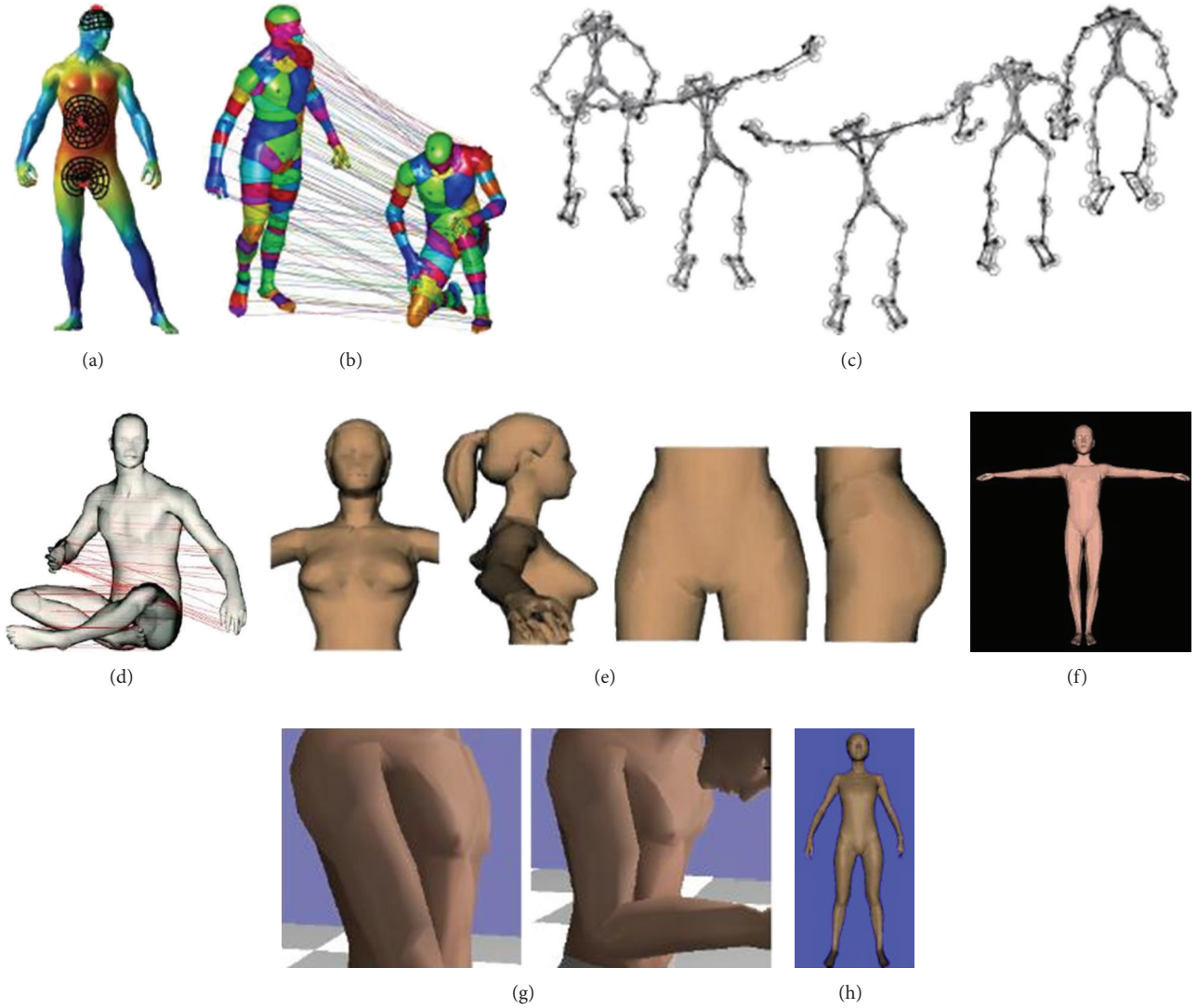


Figure 8: Motion analysis by geometry-based approaches in [44, 45, 47, 49, 51–54].

and properties of subparts of the object. The graph and the node variables are required to obey the summarization principle, which enables an efficient compositional inference algorithm to rapidly estimate the state of the HCDT. Cui et al. [51] reported a parameterized model for virtual human body (see Figure 8(e)). In this model, the virtual human body was partitioned into several parts. Based on the partitioned human model, the proportional characteristics of the human body were used to calculate the offset of the vertices to implement the deformation on specific part of the body. The interpolation method was used to smoothen the deformed surfaces.

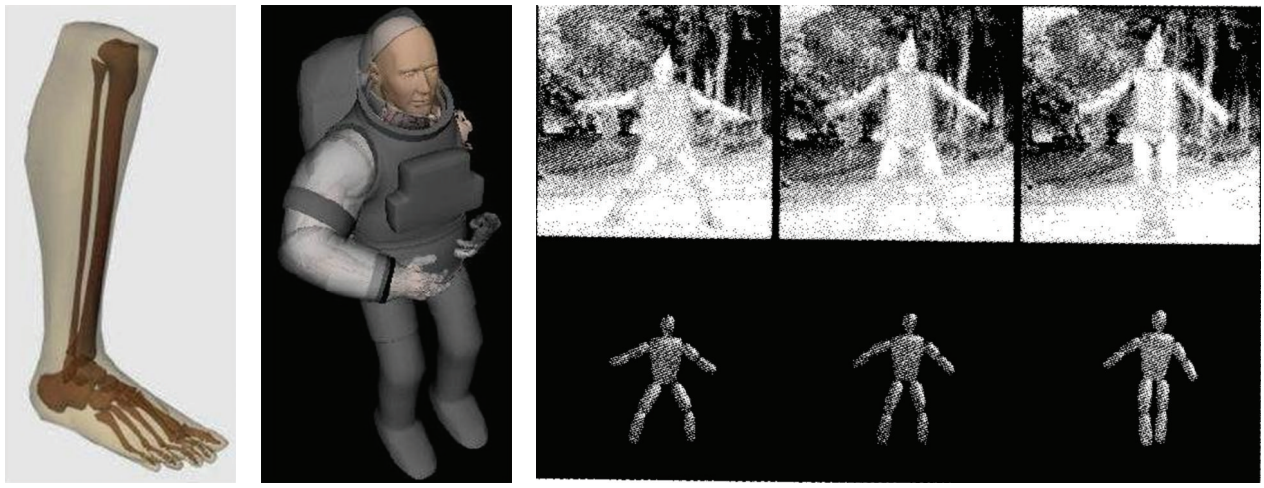
Liu and Shang [52] presented the example-based method for generating realistic, controllable human models (see Figure 8(f)). Users are assisted in automatically generating an example body data by controlling the parameters. The examples from the Poser and 3D Max are preprocessed as templates. The modeling method learns from these examples. After this learning process, the synthesizer translates the mesh of vertices to the generation of appropriate shape and proportion of the body geometry through free form deformation method. Oshita and Suzuki [53] proposed an easy-to-use real-time method to simulate realistic deformation of human skin. They utilize the fact that various skin deformations can be categorized into various deformation patterns. A deformation pattern for a local skin region is represented using a dynamic height map (see Figure 8(g)). Users of their system specify a region of the body model in the texture space to which each deformation pattern should be applied. Then, during animation, the skin deformation of the body model is realized by changing the height patterns based on the joint angles and applying bump mapping or displacement mapping to the body model. Tian et al. [54] presented an improved skinning method that can effectively reduce traditional flaws of this method. After the key terminologies are introduced, the improvement for skinning deformation is illustrated in detail. The main measures include that adding extra joints on JSL to minimize the distance between adjacent joints and

thus joint's importance attribute is introduced, using joint cluster to replace single joint; creating the corresponding relationship between skin and JSLs based on flexible model and multijoints-binding method (MJBM), that is to say binding one skin vertex to several joints using distance criterion and weight coefficient, is the function of distances between the skin vertex and its related joints. All these improvements can make the skin deformation more smooth (see Figure 8(h)). Zhou and Zhao [55] presented a skin deformation algorithm for creating 3D characters or virtual human models. The algorithm can be applied to rigid deformation, joint-dependent localized deformation, skeleton driven deformation, cross-contour deformation, and free-form deformation (FFD). These deformations are computed and demonstrated with examples, and the algorithm is applied to overcome the difficulties in mechanically simulating the motion of the human body by club-shape models. The techniques enable the reconstruction of dynamic human models that can be used in defining and representing the geometrical and kinematical characteristics of human motion. Shen et al. [56] presented an approach for human skin modeling and deformation based on cross-sectional methods. Internally, the authors use dynamic trimmed parametric patches for describing the smooth deformation of skin pieces; then they polygonalize parametric patches for final body skin synthesis and rendering. Simple and intuitive, their method combines the advantages of both parametric and polygonal representations, produces very realistic body deformations, and allows the display of surface models at several levels of detail. Smeets et al. [57] used an isometric deformation model. The geodesic distance matrix is used as an isometry-invariant shape representation. Two approaches are described to arrive at a sampling order invariant shape descriptor: the histogram of geodesic distance matrix values and the set of largest singular values of the geodesic distance matrix. Shape comparison is performed by the comparison of the shape descriptors using the χ^2-distance as dissimilarity measure.

Rumpf and Wirth [58] introduced the covariance of a number of given shapes if they are interpreted as boundary contours of elastic objects. Based on the notion of nonlinear elastic deformations from one shape to another, a suitable linearization of geometric shape variations is introduced. Once such a linearization is available, a principal component analysis can be investigated. This requires the definition of a covariance metric—an inner product on linearized shape variations. The resulting covariance operator robustly captures strongly nonlinear geometric variations in a physically meaningful way and allows to extract the dominant modes of shape variation. The underlying elasticity concept represents an alternative to Riemannian shape statistics. Fundana et al. [59] proposed a method for variational segmentation of image sequences containing nonrigid, moving objects. The method is based on the classical Chan-Vese model augmented with a novel frame-to-frame interaction term, which allows them to update the segmentation result from one image frame to the next using the previous segmentation result as a shape prior. The interaction term is constructed to be pose invariant and to allow moderate deformations in shape. Mio et al. [60] studied shapes of planar arcs and closed contours modeled on elastic curves obtained by bending, stretching, or compressing line segments nonuniformly along their extensions. Shapes are represented as elements of a quotient space of curves obtained by identifying those that differ by shape-preserving transformations. The elastic properties of the curves are encoded in Riemannian metrics on these spaces. Geodesics in shape spaces are used to quantify shape divergence and to develop morphing techniques. The shape spaces and metrics constructed offer an environment for the study of shape statistics. Elasticity leads to shape correspondences and deformations that are more natural and intuitive than those obtained in several existing models. Cremers [61] tackled the challenge of learning dynamical statistical models for implicitly represented shapes. They show how these can be integrated as dynamical shape priors in a Bayesian framework for level set-based image sequence segmentation. They propose learning the temporal dynamics of a deforming shape by approximating the shape vectors of a sequence of level set functions by a Markov chain of order k.

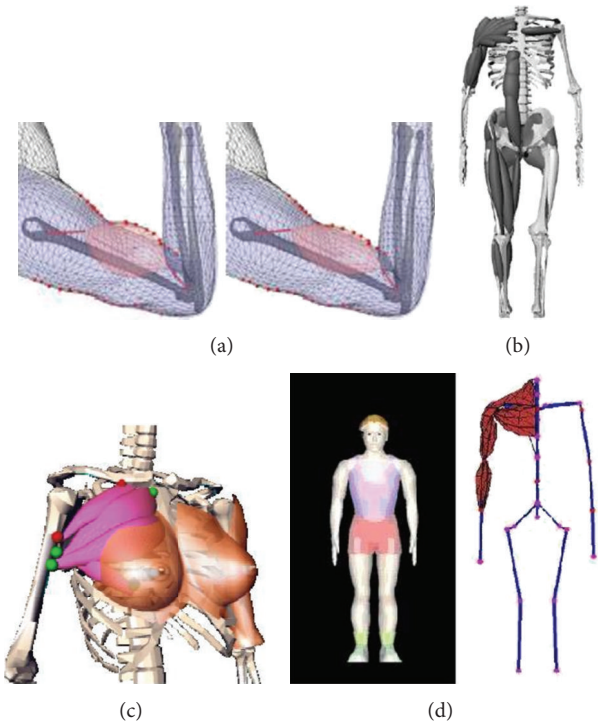
5. Physics-Based Approaches

Tang [62] presented a physics-based approach to model human skin deformation using boundary element method (BEM). Given the magnitude of displacement between the skin layer and the underlying skeleton at the anatomical landmarks, the approach determines the displacement of each vertex of the human skin model by using the BEM (see Figure 9(a)). They demonstrated their results by modeling the skin deformation of human lower limb with jumping and walking motions. Shin and Badler [63] modeled a deformable human arm to improve the accuracy of constrained reach analysis. Their research is largely composed of two parts. The first part is modeling a deformable human arm based on these empirical biomechanical properties and calculating the deformation due to various contact areas (see Figure 9(b)). The second part is evaluating the reachable space (reachability) from the arm deformation in a given geometric (CAD) environment. Using the empirical force-displacement relation, they have built a simple human arm model that deforms using a finite element method. Pentland and Horowitz [64] introduced a physically correct model of elastic nonrigid motion. This model is based on the finite element method (see Figure 9(c)), but it decouples the degrees of freedom by breaking down object motion into rigid and nonrigid vibration or deformation modes. The result is an accurate representation for both rigid and nonrigid motions that has greatly reduced dimensionality. Because of the small number of parameters involved, they have been able to use this representation to obtain accurate overconstrained estimates of both rigid and nonrigid global motions. These estimates can be integrated over time by the use of an extended Kalman filter [65], resulting in stable and accurate estimates of both 3D shape and 3D velocity. The formulation was then extended to include constrained nonrigid motion. Examples of tracking single nonrigid objects and multiple constrained objects were presented.



(a) (b) (c)

Figure 9: Motion analysis by physics-based approaches in [62–64].



(a) (b) (c) (d)

Figure 10: Motion analysis by anatomy-based approaches in [66, 68–70].

6. Anatomy-Based Approaches

Hyun et al. [66] presented a sweep-based approach to human body modeling and deformation. A rigid 3D human model, given as a polygonal mesh, is approximated with control sweep surfaces. The vertices on the mesh are bound to nearby sweep surfaces and then follow the deformation of the sweep surfaces as the model bends and twists its arms, legs, spine, and neck (see Figure 10(a)). Anatomical features including bone protrusion, muscle bulge, and skin folding are supported by a GPU-based collision detection procedure. The volumes of arms, legs, and torso are kept constant by a simple control using a volume integral formula for sweep surfaces. Zuo et al. [67] proposed a new method of muscle modeling based on both anatomical and real-time considerations. In the muscle modeling system, muscle can be constructed and edited easily through appointing some radial and transverse cross-section control parameters. Deformation of muscle model can be achieved through axial deformation and cross section's deformation. The user can adjust the precision of models to meet different requirements. Nedel and Thalmann [68] proposed a method to simulate human bodies based on anatomy concepts. Their model is divided into three layers and presented in three steps: the concept of a rigid body from a real skeleton, the muscle design and deformation based on physical concepts, and skin generation. Muscles are represented at two levels: the action lines and the muscle shape. The action line represents the force produced by a muscle on the bones, while the muscle shapes used in the simulation consist of surface-based models. To physically simulate deformations, they used a mass-spring system with a new kind of springs, called "angular springs," which were developed to control the muscle volume during simulation (see Figure 10(b)). Aubel and Thalmann [69] proposed a new, generic, multilayered model for automating the deformations of the skin of human characters based on physiological and anatomical considerations. Muscle motion and deformation are automatically derived from an action line that is deformed using a 1D mass-spring system. They cover the muscle layer with a viscoelastic fat layer that concentrates the crucial dynamics effects of the animation (see Figure 10(c)). Min et al. [70] proposed an anatomically based modeling and animation scheme for a human body model whose shape was created from 3D scan data of a human body. The proposed human body model is composed of three layers: a skeleton layer, a muscle layer, and a skin layer. The skeleton layer, represented as a set of joints and bones, controls the animation of the human body model. The muscle layer deforms the skin

layer realistically during animation. They create the muscles in that layer using soft objects, also known as blobby objects or metaballs, and deform them through the insertion/origin points of the muscles and the volume-preserving constraints. To deform the skin layer during animation, they bind the skin layer to both the skeleton layer and the muscle layer by finding corresponding joints and muscles of the vertices on the skin layer. They applied the proposed scheme for modeling the upper limb and shoulder of human body (see Figure 10(d)).

7. Conclusion

This paper attempts to provide a comprehensive survey of research on deformable human modeling and motion analysis and to provide some structural categories for the methods described in over 60 papers. This work can be generally categorized as 2D model, 3D surface model, and geometry-based, physics-based, and anatomy-based approaches. Compared with rigid motion analysis, the analysis of nonrigid and elastic articulated motion is still in its infancy at the current stage. The main difficulties in developing algorithms for human shape and motion analysis stem from the complex 3D nonrigid motions of human. The motivations of human motion analysis based on deformable models are driven by application areas such as medical imaging, biomedical applications, gesture recognition, choreography, video conferencing, material deformation studies, and image compression.

Although over the last decade much progress has been made in human pose estimation based on deformable model, there remain a number of open problems and challenges. First, one important motivation of this research is to build a body surface model that properly describes human body deformation from a small number of parameters and human 3D shape analysis from 2D image sequences. Second, sports biomechanics analysis is made to determine temporal and spatial parameters, kinematic variables, and kinetic variables of human body. Sports biomechanics analysis has been restricted to rigid models. There is less research using nonrigid model. Image-based sports biomechanics analysis of deformable human body model includes volume, gravity center, force of gravity and moment of inertia determination, and kinetic and rotational dynamics analysis. Third, while tracking walking motions in semicontrolled settings is more or less reliable, robust tracking of arbitrary and highly dynamic motions is still challenging even in controlled setups. Fourth, tracking arbitrary motions in outdoor settings has been mostly unaddressed and remains one of the open problems in computer vision. Outdoor human tracking would be allowed to capture sport motions in their real competitive setup. Fifth, tracking people in the office or in the streets interacting with the environment is still an extremely challenging problem to be solved. We expect that novel schemes will be presented to deal with human motion analysis based on deformable models in the future.

Acknowledgments

This research is supported by the National Science Foundation of China 61075031, 31270998, and 61173096, China Postdoctoral Science Foundation 2012M511321, Jiangsu Postdoctoral Science Foundation 1102169C, and National Science and Technology Support Program of China 2012BAZ04319.

References

[1] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: survey and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2179–2195, 2009.

[2] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[3] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, no. 3, pp. 585–601, 2003.

[4] G. Pons-Moll and B. Rosenhahn, "Model-based pose estimation," in *Model-Based Pose Estimation. Visual Analysis of Humans: Looking at People*, pp. 139–170, Springer, 2011.

[5] J. K. Aggarwal, Q. Cai, W. Liao, and B. Sabata, "Articulated and elastic non-rigid motion: a review," in *Proceedings of the Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 2–14, November 1994.

[6] H. Tabia, M. Daoudi, J. P. Vandeborre, and O. Colot, "A new 3D-matching method of nonrigid and partially similar models using curve analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 852–858, 2011.

[7] G. Mori and J. Malik, "Recovering 3D human body configurations using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1052–1062, 2006.

[8] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn, "Shape analysis of elastic curves in euclidean spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1415–1428, 2011.

[9] X. Huang, N. Paragios, and D. N. Metaxas, "Shape registration in implicit spaces using information theory and free form deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1303–1318, 2006.

[10] X. Liu, T. Yu, T. Sebastian, and P. Tu, "Boosted deformable model for human body alignment," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.

[11] O. Freifeld, A. Weiss, S. Zuffi, and M. J. Black, "Contour people: a parameterized model of 2D articulated human shape," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 639–646, June 2010.

[12] S. Zuffi, O. Freifeld, and M. J. Black, "From pictorial structures to deformable structures," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3546–3553, 2012.

[13] P. Guan, O. Freifeld, and M. J. Black, "A 2D human body model dressed in eigen clothing," in *Proceedings of the European Conference on Computer Vision, part 1*, vol. 6311 of *Lecture Notes in Computer Science*, pp. 285–298, 2010.

[14] S. I. Park and J. K. Hodgins, "Capturing and animating skin deformation in human motion," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 881–889, 2006.

[15] J. K. Fayad, A. D. Bue, L. Agapito, and P. M. Q. Aguiar, "Human body modelling using quadratic deformations," in *Proceedings*

of the 7th EUROMECH Solid Mechanics Conference, pp. 1–19, Lisbon, Portugal, 2009.

[16] D. E. Hyun, S. H. Yoon, M. S. Kim, and B. Juttler, "Modeling and deformation of arms and legs based on ellipsoidal sweeping," in *Proceedings of the 11th Pacific Conference on Computer Graphics and Applications*, p. 204, 2003.

[17] H. Pan and Y. Liu, "Motion estimation of elastic articulated objects from points and contours with volume invariable constraint," *Pattern Recognition*, vol. 41, no. 2, pp. 458–467, 2008.

[18] D. Terzopoulos and D. Metaxas, "Dynamic 3D models with local and global deformations: deformable superquadrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 703–714, 1991.

[19] C. Sminchisescu, *Estimation algorithms for ambiguous visual models—three dimensional human modeling and motion reconstruction in monocular video sequences [Ph.D. thesis]*, Inria, 2002.

[20] M. Hofmann and D. M. Gavrila, "3D human model adaptation by frame selection and shape-texture optimization," *Computer Vision and Image Understanding*, vol. 115, no. 11, pp. 1559–1570, 2011.

[21] A. Sundaresan and R. Chellappa, "Model driven segmentation of articulating humans in Laplacian Eigenspace," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1771–1785, 2008.

[22] H. D. Yang and S. W. Lee, "Reconstruction of 3D human body pose from stereo image sequences based on top-down learning," *Pattern Recognition*, vol. 40, no. 11, pp. 3120–3131, 2007.

[23] R. Matsuda and T. Nishita, "Modeling and deformation method of human body model based on range data," in *Proceedings of the International Conference on Shape Modeling and Applications*, pp. 80–87, 272, 1999.

[24] J. F. Blinn, "A generalization of algebraic surface drawing," *ACM Transactions on Graphics*, vol. 1, no. 3, pp. 235–256, 1982.

[25] D. Thalmann, J. Shen, and E. Chauvineau, "Fast realistic human body deformations for animation and VR applications," in *Proceedings of the Computer Graphics International (CGI '96)*, pp. 166–174, June 1996.

[26] N. D'Apuzzo, R. Plänkers, P. Fua, A. Gruen, and D. Thalmann, "Modeling human bodies from video sequences," in *Electronic Imaging*, Proceedings of SPIE, pp. 1–12, Bellingham, Wash, USA, 1999.

[27] P. Fua, A. Gruen, R. Plänkers, N. D'Apuzzo, and D. Thalmann, "Human body modeling and motion analysis from video sequences," *International Archives of Photogrammetry and Remote Sensing*, vol. 32, pp. 866–873, 1998.

[28] R. Plänkers and P. Fua, "Articulated soft objects for multiview shape and motion capture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1182–1187, 2003.

[29] M. L. Tong, Y. C. Liu, and T. Huang, "Recover human pose from monocular image under weak perspective projection," in *Proceedings of Lecture Notes in Computer Science*, vol. 3766, pp. 102–111, Springer, Heidelberg, Germany, 2007.

[30] M. Nahas, H. Huitric, and M. Saintourens, "Animation of a B-Spline figure," *The Visual Computer*, vol. 3, no. 5, pp. 272–276, 1988.

[31] S. B. Fu and X. G. Yuan, "The establishment of human body based on B-spline surface," *Chinese Journal of Computers*, vol. 21, no. 12, pp. 1131–1135, 1998.

[32] H. Y. Huang, F. H. Qi, and Z. H. Yao, "A leg motion modeling method based on NURBS free-form deformation," *Journal of Computer Research & Development*, vol. 37, no. 6, pp. 697–702, 2000.

[33] J. Sanchez-Reyes and J. M. Chacon, "Hermite approximation for free-form deformation of curves and surfaces," *Computer-Aided Design*, vol. 44, no. 5, pp. 445–456, 2012.

[34] J. Z. Wang and Y. M. Jiang, "Three-dimension human body free form deformation method based on NURBS," *Computer Engineering and Applications*, vol. 38, no. 17, pp. 95–97, 2002.

[35] P. Huang, C. Budd, and A. Hilton, "Global temporal registration of multiple non-rigid surface sequences," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3473–3480, 2011.

[36] I. C. Chang and S. Y. Lin, "3D human motion tracking based on a progressive particle filter," *Pattern Recognition*, vol. 43, no. 10, pp. 3621–3635, 2010.

[37] M. Liao, Q. Zhang, H. Wang, R. Yang, and M. Gong, "Modeling deformable objects from a single depth camera," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 167–174, Kyoto, Japan, October 2009.

[38] K. Varanasi, A. Zaharescu, E. Boyer, and R. Horaud, "Temporal surface tracking using mesh evolution," *Proceedings of the 10th European Conference on Computer Vision, part 2*, pp. 30–43, 2008.

[39] A. O. Balan and M. J. Black, "The naked truth: estimating body shape under clothing," in *Proceedings of the European Conference on Computer Vision, part 2*, vol. 5303 of *Lecture Notes in Computer Science*, pp. 15–29, 2008.

[40] J. Starck and A. Hilton, "Model-based human shape reconstruction from multiple views," *Computer Vision and Image Understanding*, vol. 111, no. 2, pp. 179–194, 2008.

[41] E. De Aguiar, C. Theobalt, C. Stoll, and H. P. Seidel, "Marker-less deformable mesh tracking for human shape and motion capture," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.

[42] C. C. L. Wang, Y. Wang, and M. M. F. Yuen, "View-dependent deformation for virtual human modeling from silhouettes," in *Proceedings of the IASTED International Conference Visualization, Imaging, and Image Processing*, pp. 140–144, ACTA Press, Marbella, Spain, 2001.

[43] H. Seo, F. Cordier, L. Philippon, and N. M. Thalmann, "Interactive modeling of MPEG-4 deformable human body models," in *Proceedings of the IFIP TC5/WG5. 10 Workshop on Deformable Avatars (Deform/Avatars '00)*, pp. 120–131, 2000.

[44] I. Kokkinos, M. Bronstein, R. Litman, and A. Bronstein, "Intrinsic shape context descriptors for deformable shapes," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 159–166, 2012.

[45] D. Raviv, A. M. Bronstein, M. M. Bronstein, R. Kimmel, and N. Sochen, "Affine-invariant diffusion geometry for the analysis of deformable 3D shapes," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2361–2367, 2011.

[46] U. Castellani, M. Cristani, and V. Murino, "Statistical 3D shape analysis by local generative descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2555–2560, 2011.

[47] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Trajectory space: a dual representation for nonrigid structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1442–1456, 2011.

[48] P. F. U. Gotardo and A. M. Martinez, "Computing smooth time-trajectories for camera and deformable shape in structure from motion with occlusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 2051–2065, 2011.

[49] D. Raviv, A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Full and partial symmetries of non-rigid shapes," *International Journal of Computer Vision*, vol. 89, no. 1, pp. 18–39, 2010.

[50] L. Zhu, Y. Chen, C. Lin, and A. Yuille, "Max margin learning of hierarchical configural deformable templates (HCDTs) for efficient object parsing and pose estimation," *International Journal of Computer Vision*, vol. 93, no. 1, pp. 1–21, 2011.

[51] H. J. Cui, R. M. Wang, and Y. Li, "Parameterized model for virtual human deformation," *Journal of Fiber Bioengineering & Informatics*, vol. 4, no. 4, pp. 371–381, 2011.

[52] Z. Liu and S. Shang, "Free-form deformation algorithm of human body model for garment," in *Proceedings of the International Conference on Computer Application and System Modeling (ICCASM '10)*, vol. 11, pp. V11602–V11605, October 2010.

[53] M. Oshita and K. Suzuki, "Artist-oriented real-time skin deformation using dynamic patterns," in *Arts and Technology*, vol. 30 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 88–96, 2010.

[54] Q. G. Tian, X. J. Li, and B. Z. Ge, "Implementation of digital human modeling and skin deformation based on flexible model and multi-joints-binding method," in *Proceedings of the MIPPR Pattern Recognition and Computer Vision*, vol. 7496 of *Proceedings of the SPIE*, pp. 74961K-1–74961K-6, 2009.

[55] X. J. Zhou and Z. X. Zhao, "The skin deformation of a 3D virtual human," *International Journal of Automation and Computing*, vol. 6, no. 4, pp. 344–350, 2009.

[56] J. H. Shen, N. M. Thalmann, and D. Thalmann, "Human skin deformation from cross-sections," in *Proceedings of the Computer Graphics International*, pp. 1–17, Melbourne, Australia, 1994.

[57] D. Smeets, J. Hermans, D. Vandermeulen, and P. Suetens, "Isometric deformation invariant 3D shape recognition," *Pattern Recognition*, vol. 45, no. 7, pp. 2817–2831, 2012.

[58] M. Rumpf and B. Wirth, "An elasticity-based covariance analysis of shapes," *International Journal of Computer Vision*, vol. 92, no. 3, pp. 281–295, 2011.

[59] K. Fundana, N. C. Overgaard, and A. Heyden, "Variational segmentation of image sequences using region-based active contours and deformable shape priors," *International Journal of Computer Vision*, vol. 80, no. 3, pp. 289–299, 2008.

[60] W. Mio, A. Srivastava, and S. Joshi, "On shape of plane elastic curves," *International Journal of Computer Vision*, vol. 73, no. 3, pp. 307–324, 2007.

[61] D. Cremers, "Dynamical statistical shape priors for level set-based tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1262–1273, 2006.

[62] Y. M. Tang, "Modeling skin deformation using boundary element method," *Computer-Aided Design and Applications*, vol. 7, no. 1, pp. 101–108, 2010.

[63] H. Shin and N. I. Badler, "Modeling deformable human arm for constrained research analysis," in *Proceedings of the Digital Human Modeling Conference*, pp. 217–228, 2002.

[64] A. Pentland and B. Horowitz, "Recovery of nonrigid motion and structure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 730–742, 1991.

[65] P. Frogerais, J. Bellanger, and L. Senhadji, "Various ways to compute the continuous-discrete extended kalman filter," *IEEE Transactions on Automatic Control*, vol. 57, no. 4, pp. 1000–1004, 2012.

[66] D. E. Hyun, S. H. Yoon, J. W. Chang, J. K. Seong, M. S. Kim, and B. Jüttler, "Sweep-based human deformation," *The Visual Computer*, vol. 21, no. 8–10, pp. 542–550, 2005.

[67] L. Zuo, J. T. Li, and Z. Q. Wang, "Anatomical human musculature modeling for real-time deformation," in *Proceedings of Computer Graphics, Visualization and Computer Vision*, pp. 1–7, 2003.

[68] L. P. Nedel and D. Thalmann, "Anatomic modeling of deformable human bodies," *The Visual Computer*, vol. 16, no. 6, pp. 306–321, 2000.

[69] A. Aubel and D. Thalmann, "Realistic deformation of human body shapes," in *Proceedings of the Computer Animation and Simulation*, pp. 125–135, 2000.

[70] K. H. Min, S. M. Baek, G. A. Lee, H. Choi, and C. M. Park, "Anatomically-based modeling and animation of human upper limbs," in *Proceedings of the International Conference on Human Modeling and Animation*, pp. 1–14, IEEE Computer Society Press, Washington, DC, USA, 2000.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 391765, 8 pages
http://dx.doi.org/10.1155/2013/391765

Research Article

Chaotic Charged System Search with a Feasible-Based Method for Constraint Optimization Problems

B. Nouhi,[1] S. Talatahari,[2] H. Kheiri,[1] and C. Cattani[3]

[1] *Department of Mathematical Sciences, University of Tabriz, Tabriz, Iran*
[2] *Marand Faculty of Engineering, University of Tabriz, Tabriz, Iran*
[3] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

Correspondence should be addressed to S. Talatahari; talatahari@tabrizu.ac.ir

Received 15 December 2012; Accepted 23 January 2013

Academic Editor: Shengyong Chen

Recently developed chaotic charged system search was combined to feasible-based method to solve constraint engineering optimization problems. Using chaotic maps into the CSS increases the global search mobility for a better global optimization. In the present method, an improved feasible-based method is utilized to handle the constraints. Some constraint design examples are tested using the new chaotic-based methods, and the results are compared to recognize the most efficient and powerful algorithm.

1. Introduction

The charged system search (CSS) is known as one of the efficient optimization algorithms among metaheuristic algorithms in recent years. The large number of researches on applying and enhancing this method proves this [1–11]. This algorithm uses the governing laws of electrostatics in physics and the governing laws of motion from the Newtonian mechanics [1].

In physics, the space surrounding an electric charge has a property known as the electric field. This field exerts a force on other electrically charged objects. The electric field surrounding a point charge is specified by Coulomb's law. Coulomb confirmed that the electric force between any two small charged spheres is inversely proportional to the square of the separation distance between the particles directed along the line joining them and proportional to the product of the charges of the two particles. Also, the magnitude of the electric field at a point inside a charged sphere can be obtained using Gauss's law that is proportional to the separation distance between the particles. In the CSS, charged particles (CPs) or solution candidates are treated as a charged sphere that can exert electrical forces on each other according to the Coulomb and Gauss laws of electrostatics. The resultant force acts on each CP creating an acceleration according to Newton's second law by which, in combination with Newtonian mechanics, the position of each CP can be determined [1].

Recently, Talatahari et al. [7] developed chaotic charged system search algorithms (CCSS) based on the idea of using chaotic systems instead of random processes. In the CCSS algorithms, the role of randomness can be played by a chaotic dynamics. Experimental studies show the benefits of using chaotic signals instead of random signals. Nine chaos-based CSS methods were developed, and then, for each variant, the performance of ten different chaotic maps was investigated to identify the most powerful variant. As suggested in [7], here more elaborated experiments are performed to discover the adaptive algorithm for solving optimization problems. Also, this paper combines the most powerful CCSS algorithms with feasible-based method in order to solve constraint engineering design problems. In order to evaluate these algorithms, some well-studied numerical engineering examples are analyzed using some best CCSS methods and compared to the other methods.

2. Standard and Chaotic-Based Charged System Search Algorithm

2.1. Standard Charged System Search Algorithm. The charged system search contains a number of charged particle (CPs)

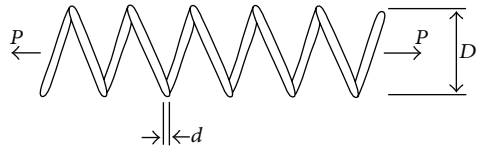


Figure 1: Schematic of tension/compression string.

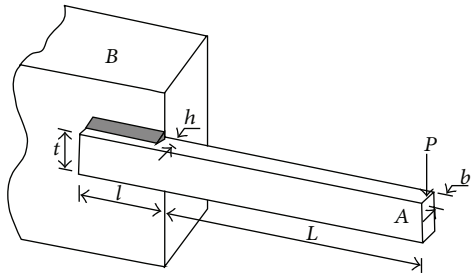


Figure 2: Schematic of welded beam structure.

where each one is treated as a charged sphere and can insert an electric force to the others. The pseudocode for the CSS algorithm is summarized as follows [1].

Step 1 (initialization). The magnitude of charge for each CP is defined as

$$q_i = \frac{\text{fit}(i) - \text{fitworst}}{\text{fitbest} - \text{fitworst}} \quad i = 1, 2, \ldots, N, \tag{1}$$

where fitbest and fitworst are the best and the worst fitness of all the CPs, fit(i) represents the fitness of the agent i, and N is the total number of CPs. The separation distance r_{ij} between two charged particles is defined as follows:

$$r_{ij} = \frac{\left\|\mathbf{X}_i - \mathbf{X}_j\right\|}{\left\|\left(\mathbf{X}_i + \mathbf{X}_j\right)/2 - \mathbf{X}_{\text{best}}\right\| + \varepsilon}, \tag{2}$$

where $\mathbf{X}_i$ and $\mathbf{X}_j$ are the positions of the ith and jth CPs, respectively, $\mathbf{X}_{\text{best}}$ is the position of the best current CP, and ε is a small positive number. The initial positions of CPs are determined randomly as

$$x_{i,j}^{(o)} = x_{i,\min} + \text{rand}_{ij} \cdot \left(x_{i,\max} - x_{i,\min}\right), \quad i = 1, 2, \ldots, N, \tag{3}$$

where $x_{i,j}^{(o)}$ determines the initial value of the ith variable for the jth CP; $x_{i,\min}$ and $x_{i,\max}$ are the minimum and the maximum allowable values for the ith variable.

Step 2 (CM creation). A number of the best CPs and the values of their corresponding fitness functions are saved in the charged memory (CM).

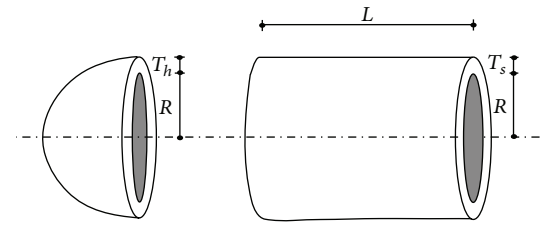


Figure 3: Schematic of pressure vessel.

Step 3 (the probability function and kind of the forces determination). The probability of moving each CP toward the others is determined using the following function:

$$p_{ij} = \begin{cases} 1, & \dfrac{\text{fit}(i) - \text{fitbest}}{\text{fit}(j) - \text{fit}(i)} > \text{rand} \vee \text{fit}(j) > \text{fit}(i), \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

The kind of the forces can be attractive or repelling determined by using the force parameter ar_{ij} defined as

$$ar_{ij} = \begin{cases} +1, & k_t < \text{rand}_{ij}, \\ -1, & k_t > \text{rand}_{ij}, \end{cases} \tag{5}$$

where ar_{ij} determines the type of the force, in which +1 represents the attractive force and –1 denotes the repelling force, and k_t is a parameter to control the effect of the kind of the force.

Step 4 (forces determination). The resultant force vector for each CP is calculated as

$$\mathbf{F}_j = q_j \sum_{i,i \neq j} \left(\frac{q_i}{a^3} r_{ij} \cdot i_1 + \frac{q_i}{r_{ij}^2} \cdot i_2\right) ar_{ij} p_{ij} \left(\mathbf{X}_i - \mathbf{X}_j\right), \quad j = 1, 2, \ldots, N \quad i_1 = 1, \quad i_2 = 0 \Longleftrightarrow r_{ij} < a \quad i_1 = 0, \quad i_2 = 1 \Longleftrightarrow r_{ij} \geq a. \tag{6}$$

Step 5 (solution construction). Each CP moves to the new position as

$$\mathbf{X}_{j,\text{new}} = \text{rand}_{j1} \cdot k_a \cdot \frac{\mathbf{F}_j}{m_j} \cdot \Delta t^2 + \text{rand}_{j2} \cdot k_v \cdot \mathbf{V}_{j,\text{old}} \cdot \Delta t + \mathbf{X}_{j,\text{old}}, \tag{7}$$

$$\mathbf{V}_{j,\text{new}} = \frac{\mathbf{X}_{j,\text{new}} - \mathbf{X}_{j,\text{old}}}{\Delta t},$$

where k_a and k_v are the acceleration and the velocity coefficients, respectively, and rand_{j1} and rand_{j2} are two random numbers uniformly distributed in the range $(0, 1)$.

Step 6 (CP position correction). If each CP swerves off the predefined bounds, its position is corrected using the harmony search-based handling approach.

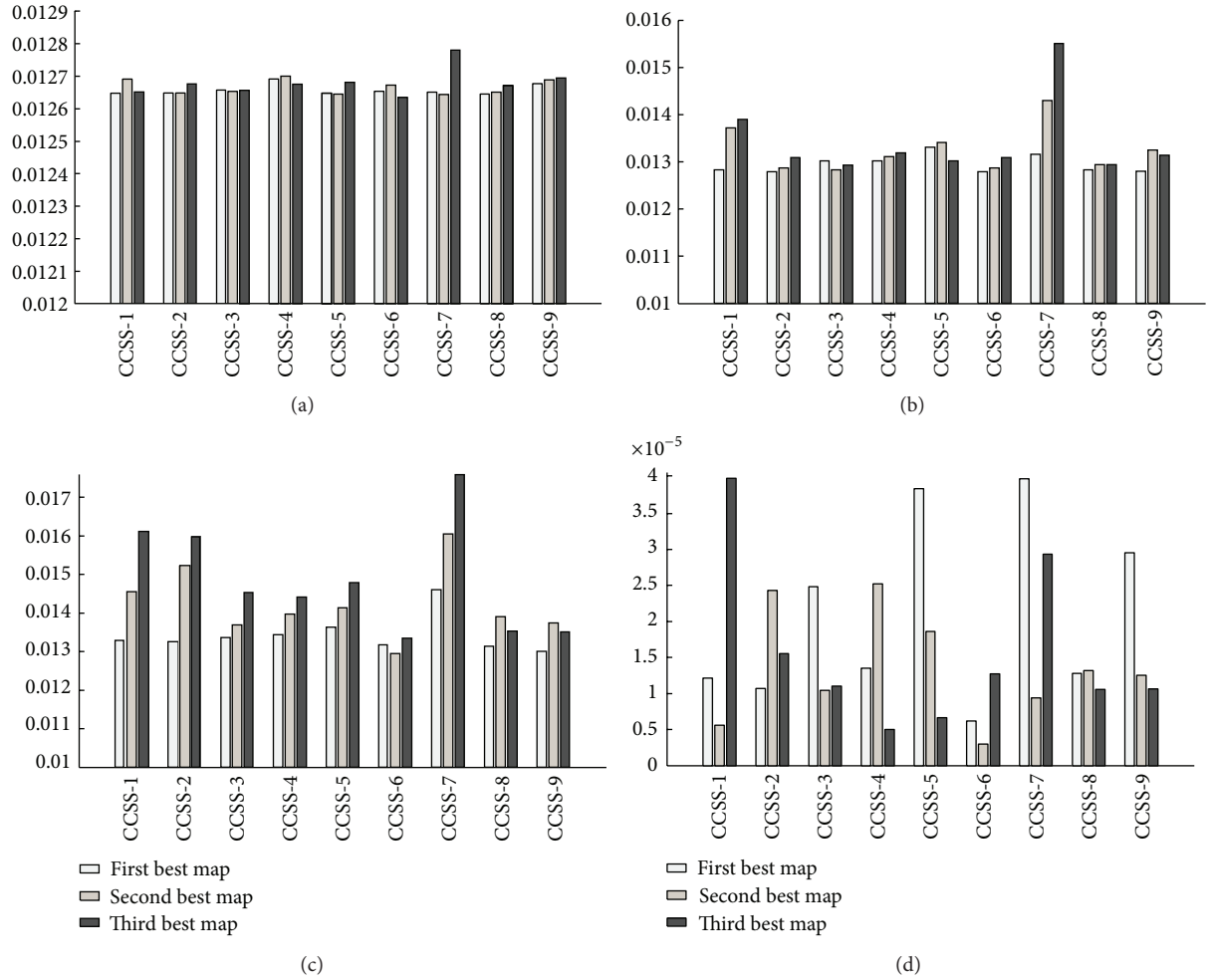


Figure 4: The results obtained by the CCSS methods for the string problem: (a) best results (b) mean of results, (c) worst result, and (d) Standard deviation.

Step 7 (CM updating). The better new vectors are included in the CM, and the worst ones are excluded from the CM.

Step 8 (terminating criterion control). Steps 3–7 are repeated until a terminating criterion is satisfied.

2.2. Chaotic Charged System Search Algorithm. To the best of our knowledge, random initialization of the CSS and the adjusted limit parameters may affect the performance of the algorithm and reduce or increase its convergence speed. In other words, the parameters of the algorithm such as k_t, k_a, and k_v are the key factors to control the balance of the exploration and exploitation of the algorithm. This problem was solved by introducing chaotic charged system search [7] where the suitable values or the initial population can be generated chaotically by using chaotic maps. Chaos is a deterministic, random-like process found in nonlinear, dynamical system, which is nonperiod, nonconverging, and bounded [12]. The nature of chaos looks to be random and unpredictable, possessing an element of regularity and randomness of a simple deterministic dynamical system, and chaotic system may be considered as the source of randomness [13, 14].

The chaotic CSS algorithms, denoted by CCSS, can be obtained by using the values generated by a chaotic map instead of one or more random parameters needed in the CSS algorithm. Therefore, considering which parameter is defined chaotically, we can specify different algorithms. Table 1 summarizes different chaotic-based CSS algorithms. There are different chaotic maps to be used as listed in Table 2.

3. Present Algorithms

In [7], by hybridizing the charged system search and chaos, different methods were developed to solve numerical global optimization problems. The proposed approaches utilized different chaotic maps to adapt the parameters of the CSS algorithm. Nine chaotic CSS algorithms by using ten different chaotic maps are defined resulting in 90 different methods. These methods were then analyzed for the benchmark mathematical functions. The simulation resulting demonstrated that some tested CCSS approaches are efficient methods to

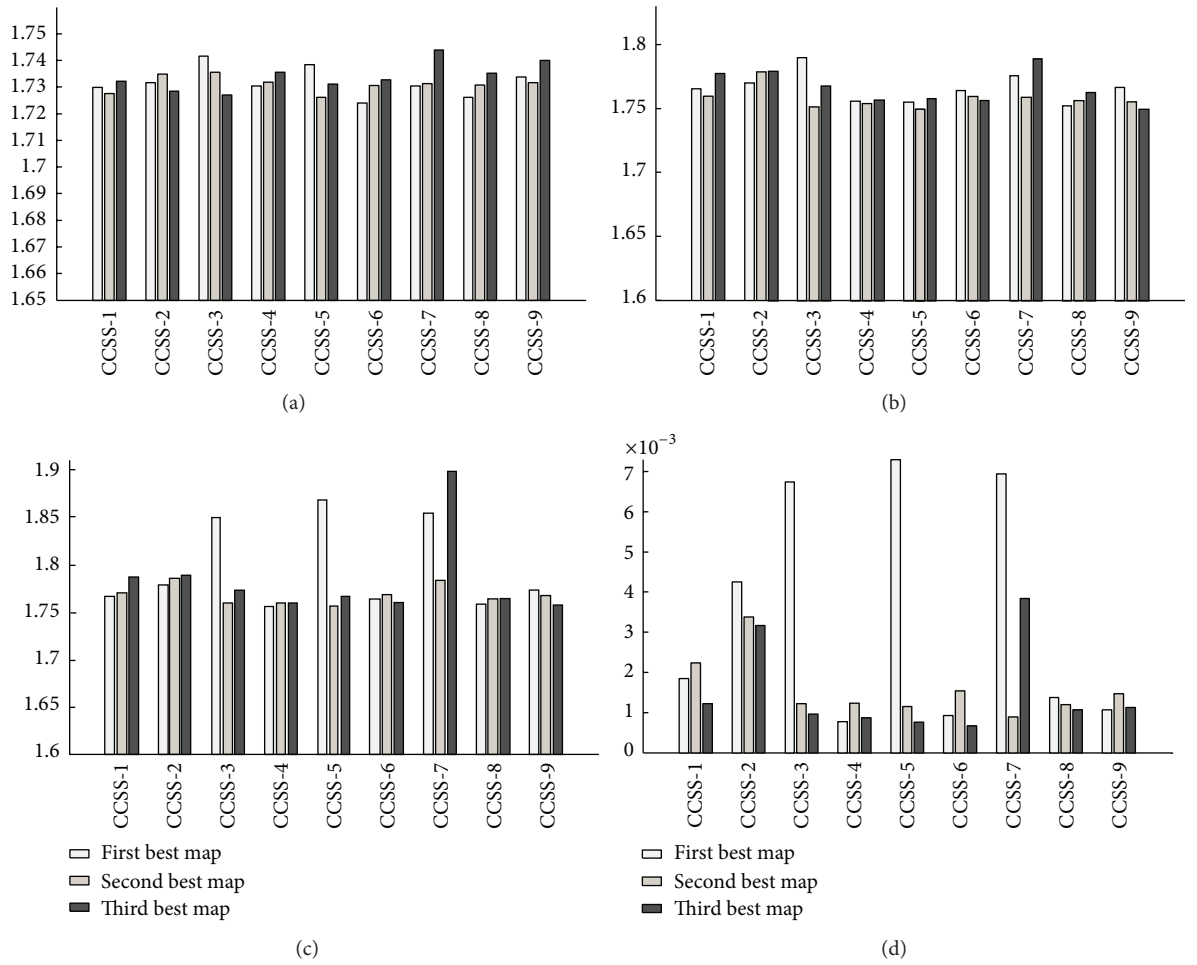


Figure 5: The results obtained by the CCSS methods for the welded beam problem: (a) best results (b) mean of results, (c) worst result, (d) Standard deviation.

Table 1: Different chaotic-based CSS algorithms.

Algorithm	Equation	Condition
CCSS-1	$x_{i,j} = x_{i,\min} + cm_{ij} \cdot (x_{i,\max} - x_{i,\min})$	The initial positions of CPs are determined chaotically.
CCSS-2	$ar_{ij} = \begin{cases} +1 & k_t < cm_{ij} \\ -1 & k_t > cm_{ij} \end{cases}$	The kind of the force is determined chaotically.
CCSS-3	$p_{ij} = \begin{cases} 1 & \dfrac{\text{fit}(i) - \text{fitbest}}{\text{fit}(j) - \text{fit}(i)} > cm_{ij} \vee \text{fit}(j) > \text{fit}(i) \\ 0 & \text{otherwise} \end{cases}$	The probability of moving each CP toward the others is determined chaotically.
CCSS-4	CCSS-2 + CCSS-3	The a_{ij} and p_{ij} are determined chaotically.
CCSS-5	$\mathbf{X}_{j,\text{new}} = cm_{j1} \cdot \dfrac{\mathbf{F}_j}{m_j} \cdot \Delta t^2 + \text{rand}_{j2} \cdot k_v \cdot \mathbf{V}_{j,\text{old}} \cdot \Delta t + \mathbf{X}_{j,\text{old}}$	The coefficient of the force is determined chaotically.
CCSS-6	$\mathbf{X}_{j,\text{new}} = \text{rand}_{j1} \cdot k_a \cdot \dfrac{\mathbf{F}_j}{m_j} \cdot \Delta t^2 + cm_{j2} \cdot \mathbf{V}_{j,\text{old}} \cdot \Delta t + \mathbf{X}_{j,\text{old}}$	The coefficient of the velocity is determined chaotically.
CCSS-7	$\mathbf{X}_{j,\text{new}} = cm_{j1} \cdot \dfrac{\mathbf{F}_j}{m_j} \cdot \Delta t^2 + cm_{j2} \cdot \mathbf{V}_{j,\text{old}} \cdot \Delta t + \mathbf{X}_{j,\text{old}}$	The coefficients of the force and velocity are determined chaotically.
CCSS-8	CCSS-4 + CCSS-7	The initial positions of CPs are determined randomly, and the rest random generators are placed to chaotic maps.
CCSS-9	CCSS-1 + CCSS-8	All random generators are placed to chaotic maps.

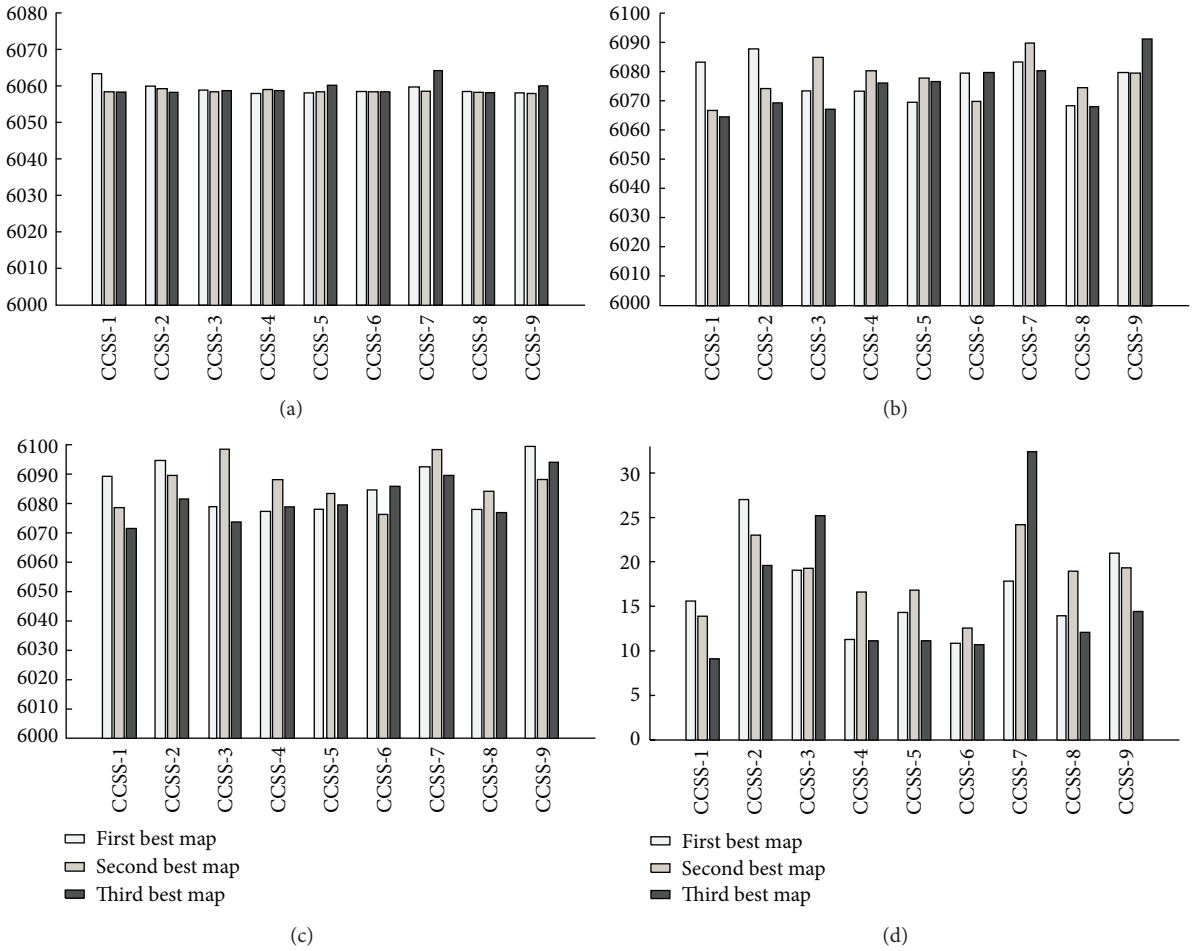


Figure 6: The results obtained by the CCSS methods for the pressure vessel problem: (a) best results (b) mean of results, (c) worst result, and (d) Standard deviation.

explore the search space and discover the global solution. From the statistical investigation, it was shown that the CCSS-8, utilizing the chaotic maps, instead of a_{ij}, p_{ij}, k_v, and k_a, has better performance than other approaches. The second best methods were CCSS-6 and CCSS-9. In order to simplify determining the most adaptive maps, for each algorithm three maps are chosen as the best ones as summarized in Table 3. In this paper, the selected maps are utilized for solving constraint engineering design examples.

On the other hand, it is necessary to handle the constraints of the problem by using a suitable method [20–22]. The feasible-based approach [23] deals with constrained search spaces by using the separation of constraints and objectives. In this method, the idea is to avoid the combination of the value of the objective function and the constraints of a problem to assign fitness, like when using a penalty function [23–26]. Kaveh and Talatahari [27] have presented a modified feasible-based method which employs the following four rules as.

Rule 1. Any feasible solution is preferred to any infeasible solution.

Rule 2. Infeasible solutions with slight violations of the constraints are treated as feasible ones.

Rule 3. Between two feasible solutions, the one with better objective function value is preferred.

Rule 4. Between two infeasible solutions, the one having smaller sum of constraint violations is preferred.

By using the first and fourth rules, the search tends to the feasible region rather than infeasible region, and by employing the third rule, the search tends to the feasible region with good solutions [28]. For most of the engineering optimization problems, the global minimum locates on or close to the boundary of a feasible design space. By applying Rule 2, the CPs approach to the boundaries and can go near to the global minimum in a great probability.

If the location of CPs become out of the variable boundaries, the solutions cannot be used. In this paper, using the harmony search-based handling approach, this problem is dealt with. According to this mechanism, any component

Table 2: Different utilized chaotic maps.

Name	Equation	Condition
Logistic map [15]	$cm_{k+1} = a \cdot cm_k(1 - cm_k)$	$cm_o \in (0, 1)$, $cm_o \notin \{0.0, 0.25, 0.5, 0.75, 1.0\}$ and $a = 4$
Tent map [16]	$cm_{k+1} = \begin{cases} \dfrac{cm_k}{0.7} & cm_k < 0.7 \\ \dfrac{10}{3cm_k(1 - cm_k)} & \text{otherwise} \end{cases}$	
Sinusoidal map [16]	$cm_{k+1} = a \cdot cm_k^2 \sin(\pi \cdot cm_k)$	$a = 2.3$ and $cm_o = 0.7$
Gauss map [16]	$cm_{k+1} = \begin{cases} 0 & cm_k = 0 \\ \dfrac{1}{cm_k \bmod(1)} & \text{otherwise} \end{cases}$	$\dfrac{1}{cm_k \bmod(1)} = \dfrac{1}{cm_k} - \left[\dfrac{1}{cm_k}\right]$
Circle map [17]	$cm_{k+1} = cm_k + b - \left(\dfrac{a}{2\pi}\right)\sin(2\pi \cdot cm_k) \bmod(1)$	$a = 0.5$ and $b = 0.2$
Sinus map	$cm_{k+1} = 2.3(cm_k)^{2\sin(\pi \cdot cm_k)}$	
Hénon map	$cm_{k+1} = 1 - a \cdot cm_k^2 + b \cdot cm_{k-1}$	$a = 1.4$ and $b = 0.3$
Ikeda map [18]	$x_{n+1} = 1 + 0.7\left(x_n \cos(\theta_n) - y_n \sin(\theta_n)\right)$ $y_{n+1} = 0.7\left(x_n \sin(\theta_n) + y_n \cos(\theta_n)\right)$	$\theta_n = 0.4 - \dfrac{6}{1 + x_n^2 + y_n^2}$
Liebovtech map [19]	$x_{k+1} = \begin{cases} \alpha_1 x_k & 0 < x_k \le d_1, \\ \dfrac{d_2 - x_k}{d_2 - d_1} & d_1 < x_k \le d_2, \\ 1 - \alpha_2(1 - x_k) & d_2 < x_k \le 1 \end{cases}$	$\alpha_1 = \dfrac{d_2}{d_1}(1 - (d_2 - d_1))$, $\alpha_2 = \dfrac{1}{d_2 - 1}\left((d_2 - 1) - d_1(d_2 - d_1)\right)$.
Zaslavski map [16]	$y(k+1) = \left[y(k) + v + az(k+1)\right] \pmod 1$ $z(k+1) = \cos(2\pi y(k)) + e^{-r} z(k)$	$v = 400$, $r = 3$, $a = 12.6695$ and $z(t) \in [-1.0512, 1.0512]$

Table 3: Best maps for each CCSS.

Best maps	CCSS								
	1	2	3	4	5	6	7	8	9
1	Tent	Tent	Sinusoidal	Tent	Liebovtech	Liebovtech	Liebovtech	Sinus	Sinus
2	Circle	Gauss	Ikeda	Ikeda	Sinus	Sinusoidal	Zaslavski	Liebovtech	Tent
3	Sinus	Ikeda	Sinus/Gauss	Circle	Tent	Zaslavski	Sinus	Tent	Gauss

of the CP's vector violating the variable boundaries can be generated randomly from CM as

$$x_{i,j} = \begin{cases} \text{w.p. HMCR ==> select a new value} \\ \quad \text{for variable from CM} \\ \quad\quad \text{==> w.p. (1 – PAR)} \\ \quad\quad\quad \text{do nothing} \\ \quad\quad \text{==> w.p. PAR choose} \\ \quad\quad\quad \text{a neighboring value} \\ \text{w.p. (1 – HMCR) ==> select a new value} \\ \quad \text{randomly from} \\ \quad \text{the allowable list.} \end{cases} \tag{8}$$

4. Numerical Engineering Problems

The population size $N = 20$ is sufficient for most of the problems while the maximum number of function evolutions for the examples is set to 4,000 (200 iterations). The explanation of the examples is presented in Section 4.1, and the performance of the CCSS algorithms to optimize these functions is investigated in the next subsection.

4.1. Description of the Examples. Three engineering design problems which have been previously solved using a variety of other techniques are considered to perform investigation on efficiency of the proposed algorithms. The description of these examples is as the following.

4.1.1. A Tension/Compression String Design Problem. This problem consists of minimizing the weight of a tension/compression spring subject to constraints on shear stress, surge frequency, and minimum deflection as shown

in Figure 1. The design variables are the mean coil diameter $D(= x_1)$, the wire diameter $d(= x_2)$, and the number of active coils $N(= x_3)$. The problem is defined with details in [1].

4.1.2. A Welded Beam Design Problem. The welded beam structure, shown in Figure 2, is a practical design problem that often has been used as a benchmark problem for testing different optimization methods. The objective is to find the minimum fabricating cost of the welded beam subject to constraints on shear stress (τ), bending stress (σ), buckling load (P_c), end deflection (δ), and side constraint. There are four design variables, namely, $h(= x_1)$, $l(= x_2)$, $t(= x_3)$, and $b(= x_4)$. The detailed information about the constraints and objective function is presented in [1].

4.1.3. A Pressure Vessel Design Problem. A cylindrical vessel is capped at both ends by hemispherical heads as shown in Figure 3. The objective is to minimize the total cost, including the cost of material, forming, and welding.

The variables contain x_1 as the thickness of the shell (T_s), x_2 as the thickness of the head (T_h), x_3 as the inner radius (R), and x_4 as the length of cylindrical section of the vessel, not including the head (L). Details related to the objective and constraint functions are available in [1].

4.2. Experimental Results

4.2.1. Results for the Tension/Compression String Problem. For this example, performances are assessed on the basis of the best fitness values and the statistics results of the new approaches as reported in Figure 4. Simulation results show that all the proposed methods perform satisfactorily. However, the best feasible solutions obtained by the methods CCSS-8 (with Sinus map), CCSS-6 (with Zaslavski map), and CCSS-5 (with Sinus map) are the best results among the others. Meanwhile, the CSS-8 and CCSS-6 methods (almost with all the three maps) have better performance in relation to the mean and worst values. Moreover, the standard deviation of the results by the CSS-6 in 50 independent runs for this problem is the smallest among the others.

4.2.2. Results for the Welded Beam Problem. Figure 5 presents the statistical information of this problem obtained by the proposed algorithms. Clearly, the worst solutions belong to the CCSS-7, while the best ones belong to the CCSS-6 and CCSS-8. The CCSS-4 as well as CCSS-9 stand in the second place. Although some new methods have small differences in relation to best, mean, or worst results, obviously many of them perform very well in relation to the standard deviation. Similar to the previous example, the CCSS-6 has the best standard deviation compared to the other methods.

4.2.3. Results for the Pressure Vessel Problem. The obtained results using 9 variants of the presented algorithms are shown in Figure 6. Similar to the previous examples, many of the CCSS methods can find a suitable result successfully; however, the best solution found by CCSS-6 and CCSS-4 is better than the best solutions found by the other techniques. Also, it can be seen that the average searching quality of the CCSS-8 (with Sinus and Tent maps) and CCSS-6 (with Sinusoidal map) algorithms is better than those of other methods. Many of the proposed methods improve the reliability of the algorithm by reducing the standard deviation values.

5. Conclusion Remarks

As suggested in [7] due to the superiorities of the CCSS methods, here more elaborated experiments are performed to discover the better methods which can be utilized in solving engineering problems. From 90 different chaotic charged system search algorithms (obtained by using 10 chaotic maps and 9 different methods), 27 most efferent methods (3 different chaotic maps for each algorithm) are chosen and investigated to solve constraint engineering problems.

Performances are assessed on the basis of the best fitness values and the statistics results of the new approaches from 50 runs with different seeds. Simulation results show that for all examples, the proposed methods perform satisfactorily. Almost all of the proposed methods improve the reliability of the algorithm by reducing the standard deviation values. From numerical results, it is clear that CCSS-6, in which the coefficient of the velocity is determined chaotically, is the most reliable algorithm having the smallest standard deviation values, while, the algorithm with chaotic coefficients for the force and velocity (CCSS-7) is the worst one. Meanwhile, the CSS-8 and CCSS-6 methods have better performance in relation to the best, mean, and worst values. In CCSS-8, all parameters of the algorithm are determined chaotically, but the initial positions of agents are defined randomly. To sum up, the coefficient of the velocity plays a key role in reliability of the algorithm and the results show that a chaotic velocity coefficient can improve the performance of the algorithm; in addition, using all chaotic parameters of the algorithm can improve the performance, as well; however, chaotic initialization has very small or even no influence on the final results. This chaotic charged model, which is a coupled system, for constraint optimization problems in future could also benefit from other coupled systems such as the kinetic models of competition and the corresponding hybrid competition models [29, 30].

References

[1] A. Kaveh and S. Talatahari, "A novel heuristic optimization method: charged system search," *Acta Mechanica*, vol. 213, no. 3-4, pp. 267–289, 2010.

[2] A. Kaveh and S. Talatahari, "A charged system search with a fly to boundary method for discrete optimum design of truss structures," *Asian Journal of Civil Engineering*, vol. 11, no. 3, pp. 277–293, 2010.

[3] A. Kaveh and S. Talatahari, "Optimal design of skeletal structures via the charged system search algorithm," *Structural and Multidisciplinary Optimization*, vol. 41, no. 6, pp. 893–911, 2010.

[4] A. Kaveh and S. Talatahari, "Charged system search for optimum grillage system design using the LRFD-AISC code,"

Journal of Constructional Steel Research, vol. 66, no. 6, pp. 767–771, 2010.

[5] A. Kaveh and S. Talatahari, "Geometry and topology optimization of geodesic domes using charged system search," *Structural and Multidisciplinary Optimization*, vol. 43, no. 2, pp. 215–229, 2011.

[6] A. Kaveh and S. Talatahari, "An enhanced charged system search for configuration optimization using the concept of fields of forces," *Structural and Multidisciplinary Optimization*, vol. 43, no. 3, pp. 339–351, 2011.

[7] S. Talatahari, A. Kaveh, and R. Sheikholeslami, "An efficient charged system search using chaos for optimization problems," *International Journal of Optimization in Civil Engineering*, vol. 1, no. 2, pp. 305–325, 2011.

[8] S. Talatahari, R. Sheikholeslami, M. Shadfaran, and M. Porbaba, "Charged system search algorithm for optimum design of gravity retaining walls subject to seismic loading," *Mathematical Problems in Engineering*, vol. 2012, Article ID 301628, 10 pages, 2012.

[9] S. Talatahari, A. Kaveh, and R. Sheikholeslami, "Engineering design optimization using chaotic enhanced charged system search algorithms," *Acta Mechanica*, vol. 223, no. 10, pp. 2269–2285, 2012.

[10] S. Talatahari, A. Kaveh, and N. Mohajer Rahbari, "Parameter identification of Bouc-Wen model for MR fluid dampers using adaptive charged system search optimization," *Journal of Mechanical Science and Technology*, vol. 26, no. 8, pp. 2523–2534, 2012.

[11] T. Niknam, F. Golestaneh, and M. Shafiei, "Probabilistic energy management of a renewable microgrid with hydrogen storage using self-adaptive charge search algorithm," *Energy*, vol. 49, pp. 252–267, 2013.

[12] B. Alatas, "Chaotic bee colony algorithms for global numerical optimization," *Expert Systems with Applications*, vol. 37, pp. 5682–5687, 2010.

[13] G. G. Schuster, *Deterministic Chaos: An Introduction*, Federal Republic of Germany: Physick, GmnH, Weinheim, Germany, 2nd edition, 1988.

[14] L. D. S. Coelho and V. C. Mariani, "Use of chaotic sequences in a biologically inspired algorithm for engineering design optimization," *Expert Systems with Applications*, vol. 34, no. 3, pp. 1905–1913, 2008.

[15] R. M. May, "Simple mathematical models with very complicated dynamics," *Nature*, vol. 261, no. 5560, pp. 459–467, 1976.

[16] H.-O. Peitgen, H. Jürgens, and D. Saupe, *Chaos and Fractals*, Springer, Berlin, Germany, 1992.

[17] W. M. Zheng, "Kneading plane of the circle map," *Chaos, Solitons and Fractals*, vol. 4, no. 7, pp. 1221–1233, 1994.

[18] U. Dressler and J. D. Farmer, "Generalized Lyapunov exponents corresponding to higher derivatives," *Physica D. Nonlinear Phenomena*, vol. 59, no. 4, pp. 365–377, 1992.

[19] A. Erramilli, R. P. Singh, and P. Pruthi, *Modeling Packet Traffic with Chaotic Maps*, Royal Institute of Technology, Stockholm-Kista, Sweden, 1994.

[20] C. Cattani, S. Y. Chen, and G. Aldashev, "Information and modeling in complexity," *Mathematical Problems in Engineering*, vol. 2012, Article ID 868413, 4 pages, 2012.

[21] S. C. Lim, C. H. Eab, K. H. Mak, M. Li, and S. Y. Chen, "Solving linear coupled fractional differential equations by direct operational method and some applications," *Mathematical Problems in Engineering*, vol. 2012, Article ID 653939, 28 pages, 2012.

[22] P. Lu, S. Y. Chen, and Y. Zheng, "Artificial intelligence in civil engineering," *Mathematical Problems in Engineering*, vol. 2012, Article ID 145974, 22 pages, 2012.

[23] K. Deb, "An efficient constraint handling method for genetic algorithms," *Computer Methods in Applied Mechanics and Engineering*, vol. 186, pp. 311–338, 2000.

[24] C. A. Coello Coello, "Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art," *Computer Methods in Applied Mechanics and Engineering*, vol. 191, no. 11-12, pp. 1245–1287, 2002.

[25] S. Y. Chen and Y. F. Li, "Automatic sensor placement for model-based robot vision," *IEEE Transactions on Systems, Man and Cybernetics B*, vol. 34, no. 1, pp. 393–408, 2004.

[26] Y. Zheng, H. Shi, and S. Chen, "Fuzzy combinatorial optimization with multiple ranking criteria: a staged tabu search framework," *Pacific Journal of Optimization*, vol. 8, no. 3, pp. 457–472, 2012.

[27] A. Kaveh and S. Talatahari, "Engineering optimization withhybrid particle swarm and ant colony optimization," *Asian Journal of Civil Engineering*, vol. 10, no. 6, pp. 611–628, 2009.

[28] E. M. Montes and C. A. C. Coello, "An empirical study about the usefulness of evolution strategies to solve constrained optimization problems," *International Journal of General Systems*, vol. 37, no. 4, pp. 443–473, 2008.

[29] C. Cattani and A. Ciancio, "Hybrid two scales mathematical tools for active particles modelling complex systems with learning hiding dynamics," *Mathematical Models and Methods in Applied Sciences*, vol. 17, no. 2, pp. 171–187, 2007.

[30] C. Cattani and A. Ciancio, "Separable transition density in the hybrid model for tumor-immune system competition," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 610124, 6 pages, 2012.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 416186, 8 pages
http://dx.doi.org/10.1155/2013/416186

Research Article

Fast Texture Synthesis in Adaptive Wavelet Packet Trees

Ying-Shen Juang,[1] Hsi-Chin Hsin,[2] Tze-Yun Sung,[3] and Carlo Cattani[4]

[1] *Department of Business Administration, Chung Hua University, Hsinchu City 30012, Taiwan*
[2] *Department of Computer Science and Information Engineering, National United University, Miaoli 36003, Taiwan*
[3] *Department of Electronics Engineering, Chung Hua University, Hsinchu City 30012, Taiwan*
[4] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

Correspondence should be addressed to Tze-Yun Sung; bobsung@chu.edu.tw

Received 21 January 2013; Accepted 18 February 2013

Academic Editor: Shengyong Chen

Wavelet packet transform known as a substantial extension of wavelet transform has drawn a lot of attention to visual applications. In this paper, we advocate using adaptive wavelet packet transform for texture synthesis. The adaptive wavelet packet coefficients of an image are organized into hierarchical trees called adaptive wavelet packet trees, based on which an efficient algorithm has been proposed to speed up the synthesis process, from the low-frequency tree nodes representing the global characteristics of textures to the high-frequency tree nodes representing the local details. Experimental results show that the texture synthesis in the adaptive wavelet packet trees (TSIAWPT) algorithm is suitable for a variety of textures and is preferable in terms of computation time.

1. Introduction

The goal of texture synthesis is to generate an arbitrarily large image that resembles a given sample texture in appearance. With the emerging market in various applications ranging from computer graphics, machine vision to advanced imaging products [1–4], texture synthesis has been one of the increasingly active areas of research in recent years [5–7]. Though it is difficult to precisely define textures, two main categories have been commonly used to describe textures as either stochastic or structural patterns. For more complete surveys of texture analysis and synthesis, the reader is referred to [1, 8, 9].

One of the straightforward ways to synthesize a large textured image is to duplicate a given sample texture and stitch them together. In order to ameliorate the blocking seams [10], Cheng proposed a seamless montage method with suitable tiles designed by using the quilting algorithm [11] and Wang Tiles [12] in [13]. With the assumption of Markov random field (MRF) [14, 15], Efros and Leung estimated the conditional distributions of output pixels from the input pixels that are similar to the neighbors for nonparametric texture synthesis [16]. Wei and Levoy proposed a search-based order-independent algorithm to synthesize textures in an arbitrary order [17]. As MRF modeling is local and stationary, textures can be efficiently synthesized patch by patch, instead of pixel by pixel [18]. To speed up the process of synthesizing textures, many multiresolution based-texture synthesis algorithms were proposed [19–26]. In [19], an MRF-based similarity metric was used to minimize the difference between the synthesis image and the sample texture from lower to higher resolutions. The algorithm presented in [20] is a discrete version of [19] with an improvement in speed. In [21], the lowest-resolution texture was first synthesized, based on which a set of four candidate pixels were examined to find the best synthesis pixel at the next higher-resolution level; this process repeatedly proceeded until the highest-resolution texture was obtained. Fang presented a fast multiresolution image completion algorithm with improved convergence of the synthesis process [22]. De Bonet used the Laplacian pyramid [23] together with a filter bank to capture the characteristics of the input texture at multiple resolutions and then synthesized the output image with an efficient resolution recursive sampling procedure [24]. Wei proposed a scheme to

synthesize textures in the Gaussian pyramid [25] through the use of tree-structured vector quantization [26].

Wavelet transform provides an efficient multiresolution analysis [27–30]. It decomposes an image into subbands with orientation selectivity, in which the higher-frequency components are represented by shorter basis functions with higher spatial resolutions, and the lower frequency components are represented by larger basis functions with higher spectral resolutions; this property matches the human visual system [31]. Yu et al. proposed a simple scheme to synthesize wavelet coefficients by sampling from the wavelet coefficients of the input texture [32]. Cui proposed texture synthesis based on the lowest-frequency scaling coefficients only; the corresponding higher-frequency wavelet coefficients were obtained accordingly [33]. For images with textures, there are significant coefficients throughout wavelet subbands [34]; this needs to be taken into account for a more compact representation. Wavelet packet transform extends wavelet transform by including more basis functions [27]. In [35], we adopted an efficient scheme to organize wavelet packet coefficients into hierarchical trees, based on which an efficient algorithm had been proposed for texture synthesis. Though wavelet packet transform is preferable to wavelet transform in terms of the representational diversity, the dominant components of an image may only be distributed in parts of the wavelet subbands. Hence, it is not necessary to decompose all of the wavelet subbands of an image into wavelet packets [36–38]. In this paper, an adaptive wavelet packet transform has been proposed to represent textures in the adaptive wavelet packet trees, based on which a fast texture synthesis algorithm has also been proposed.

The remainder of this paper proceeds as follows. In Section 2, the construction of hierarchical wavelet packet trees is briefly reviewed. Section 3 presents the proposed adaptive-wavelet-packet-tree-based texture synthesis algorithm. Experimental results are given in Section 4. A conclusion can be found in Section 5.

2. Hierarchical Wavelet Packet Trees

One of the advantages of using wavelet transform is to represent signals at multiple resolutions. For various decompositions with more basis functions, the high-frequency wavelet subbands of a signal can be further decomposed using wavelet packet transform. More specifically, a sequence of wavelet coefficients (WC) $D_\ell(k)$ at resolution ℓ can be decomposed by

$$\begin{aligned} \widetilde{D}_{\ell,1}(n) &= \sum_k h(2n-k)\cdot D_\ell(k), \\ \widetilde{D}_{\ell,2}(n) &= \sum_k g(2n-k)\cdot D_\ell(k), \end{aligned} \tag{1}$$

where $h(n)$ and $g(n)$ are the low-pass and high-pass wavelet filters, respectively, and $\widetilde{D}_{\ell,1}(n)$ and $\widetilde{D}_{\ell,2}(n)$ are wavelet packet coefficients (WPC), which can be efficiently combined into a single sequence [35]:

$$\widetilde{D}_\ell(n) \begin{cases} \widetilde{D}_{\ell,1}\left(\dfrac{n}{2}\right), & \text{even } n, \\ \widetilde{D}_{\ell,2}\left(\dfrac{n-1}{2}\right), & \text{odd } n. \end{cases} \tag{2}$$

The combination of the above is based on the spatial relationships.

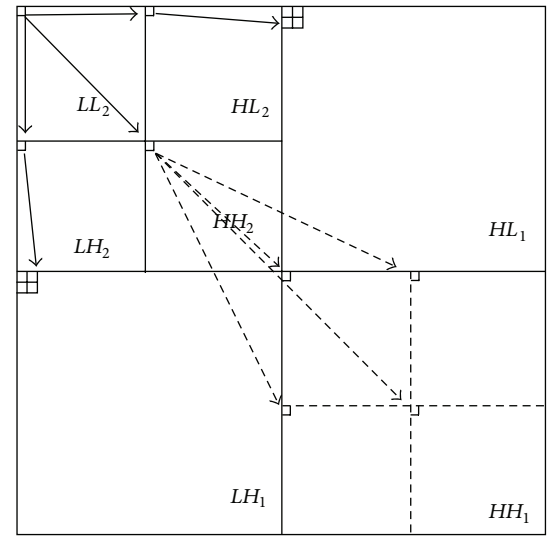


Figure 1: Example of an adaptive wavelet packet tree consisting of two wavelet subtrees in the horizontal and vertical directions (solid lines) and one wavelet packet subtree in the diagonal direction (dashed lines).

For image applications, 2D wavelet transform/wavelet packet transform can be obtained by using the tensor product of tow 1D wavelet transform/wavelet packet transform. Figure 1 shows a 2-level 2D wavelet transform, in which LL_2 is the lowest frequency subband representing the approximation of an image at the coarsest resolution 2; HL_ℓ, LH_ℓ, and HH_ℓ are the high frequency wavelet subbands representing the detail information in the horizontal, vertical, and diagonal directions, respectively, at resolution $\ell = 1, 2$; the wavelet subbands are delimited by solid lines, and two wavelet subtrees delineated by solid lines are also given to show the spatial relationships between subbands HL_1 and HL_2 (LH_1 and LH_2) at two successive resolution levels in the horizontal (vertical) direction. In addition, the wavelet subband HH_1 has been decomposed into 4 wavelet packet subbands (delimited by dashed lines), together with the subband HH_2; a 2D wavelet packet subtree delineated by dashed lines has been constructed by combining 2D WPC horizontally followed by vertically, or vice versa.

3. Adaptive Wavelet Packet Trees and Thier Application to Texture Synthesis

Wavelet packet transform has the advantage of representing images with more choices of basis functions. This section

Table 1: Numbers of significant wavelet coefficients (WC), wavelet packet coefficients (WPC), and adaptive wavelet packet coefficients (AWPC) from the most significant bit-plane 1 to the least significant bit-plane 8 and the reduced numbers of significant coefficients comparing WC to WPC and WPC to AWPC for the test image shown in Figure 2(a).

Bit plane	WC	WPC	AWPC	WC-WPC	WPC-AWPC
1	213	222	222	−9	0
2	1026	1084	1087	−58	−3
3	2558	2211	2250	347	−39
4	4647	3996	3880	651	116
5	6954	6271	5990	683	281
6	9663	9239	8858	424	381
7	12870	12616	12299	254	317
8	16683	16489	16264	194	225

presents an efficient algorithm to synthesize textures in the adaptive wavelet packet domain and a simple scheme to accelerate the synthesis process.

3.1. Adaptive Wavelet Packet Trees. For images with textures, it is likely that lots of wavelet coefficients are significant throughout the wavelet subbands, depending on the characteristics of the constituent textures. Though all of the wavelet subbands can be fully decomposed into wavelet packets, there may still be textures that are dominated by some of the wavelet subbands. Hence, we advocate the use of adaptive wavelet packet transform for texture synthesis. For the sake of simplicity, the following criterion has been used to determine the significance of a coefficient:

$$\mathrm{Sig}\left(P_i(x,y)\right) = \begin{cases} 1, & \dfrac{\left|P_i(x,y)\right|}{\max\left|P_i(x,y)\right|} > T_r, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

where the magnitude of a coefficient $P_i(x, y)$ at position (x, y) in subband B_i is first normalized and then compared to a threshold T_r. For each non-LL subband, it is further decomposed until the number of significant coefficients has no longer decreased; this leads to the desired adaptive wavelet packet transform.

Take the image shown in Figure 2(a) as an example, Table 1 shows the numbers of significant wavelet coefficients (WC), wavelet packet coefficients (WPC), and adaptive wavelet packet coefficients (AWPC) with respect to threshold 2^{-j} at bit-plane j, $j = 1, \ldots, 8$, together with the reduced numbers of significant coefficients comparing WC to WPC and WPC to AWPC. As noted, the numbers of significant coefficients have been reduced starting form bit-plane 3 by using wavelet packet transform and reduced more starting from bit-plane 4 by using the proposed adaptive wavelet packet transform. It implies that the dominant components of high-detailed textures are likely to be more concentrated in the adaptive wavelet packet domain.

As only the significant wavelet subbands of an image are decomposed into wavelet packet subbands, the resulting adaptive wavelet packet trees may consist of both wavelet subtrees and wavelet packet subtrees. Figure 1 shows an adaptive wavelet packet tree consisting of two wavelet subtrees in the horizontal and vertical directions, and one wavelet packet subtree in the diagonal direction. It still retains the same hierarchical structure as the conventional wavelet trees.

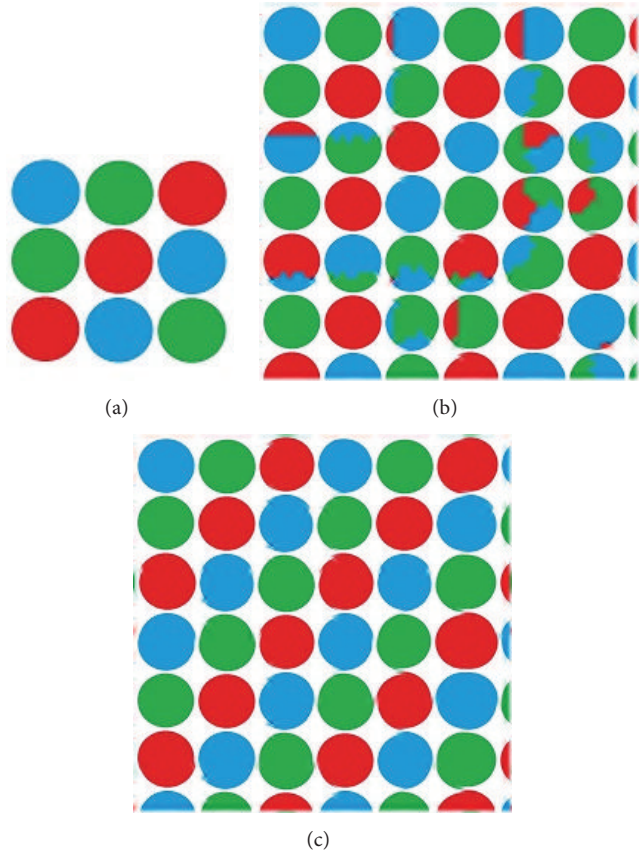


Figure 2: (a) Test image; (b) and (c) synthesis results using TSIWPT [35] and TSIAWPT.

3.2. Acceleration. In [35] we proposed the TSIWPT algorithm to synthesize a large textured image in wavelet packet trees. To reduce the computation time, TSIWPT minimizes the difference between the output wavelet packet trees and the input

wavelet packet trees based on their respective luminance components only. However, it may cause color distortions, especially for quasi-regular textures with different colors. For example; Figure 2(a) shows a regular texture consisting of blue, green, and red circular objects. The synthesis result using TSIWPT has some color mixes as shown in Figure 2(b). Though one can easily solve such color distortion problem by taking the color components into account directly while searching for the best output wavelet packet trees, it often leads to a great increase in computation time.

In this paper, we adopt the clustering approach to address the above speed and quality issue. The idea is to group the input adaptive wavelet packet trees into similarity sets, from which the best output adaptive wavelet packet trees are to be constructed. The running time can be reduced as the search of the best output adaptive wavelet packet trees is restricted to the similarity sets. In addition, the similarity sets of adaptive wavelet packet trees are constructed from the input adaptive wavelet packet trees; thus, a look-up table (LUT) can be used to store their locations in the input adaptive wavelet packet domain, which is independent of the size of the output image. As a result, the larger the synthesis image is, the greater the reduced computation time will be.

3.3. Proposed Algorithm. Figure 3 depicts a flowchart of the texture synthesis in adaptive wavelet packet trees (TSIAWPT) algorithm. We summarize TSIAWPT as follows.

Step 1. Construct the input the adaptive wavelet packet trees (AWPT) via adaptive wavelet packet transform.

Step 2. Cluster the input AWPT into similarity sets and build an LUT to store their locations.

Step 3. Take a patch of input AWPT as the initial patch of output AWPT. For the next patch of output AWPT to be synthesized, take the union of the neighboring similarity sets as the candidate set; evaluate the low-frequency tree nodes to refine the candidate set. Empirically, the tree nodes at the top two AWPT levels are suitable for the refinement of candidate sets.

Step 4. Search the refined candidate set for the best patch of output AWPT based on the high-frequency tree nodes.

Step 5. Repeat Step 3 followed by Step 4 until all the patches of output AWPT are synthesized.

Step 6. Take the inverse of the output AWPT to produce the synthesis image.

For a raster scan with rectangular patches, two types of neighboring patches, that is, the upper and left neighboring patches, are involved in clustering the input AWPT into similarity sets. The synthesized patches of output AWPT are thus obtained from the union of their respective upper and left neighboring similarity sets.

To determine a suitable patch size for textures with quasi-periodic structures, we adopt the use of autocorrelation

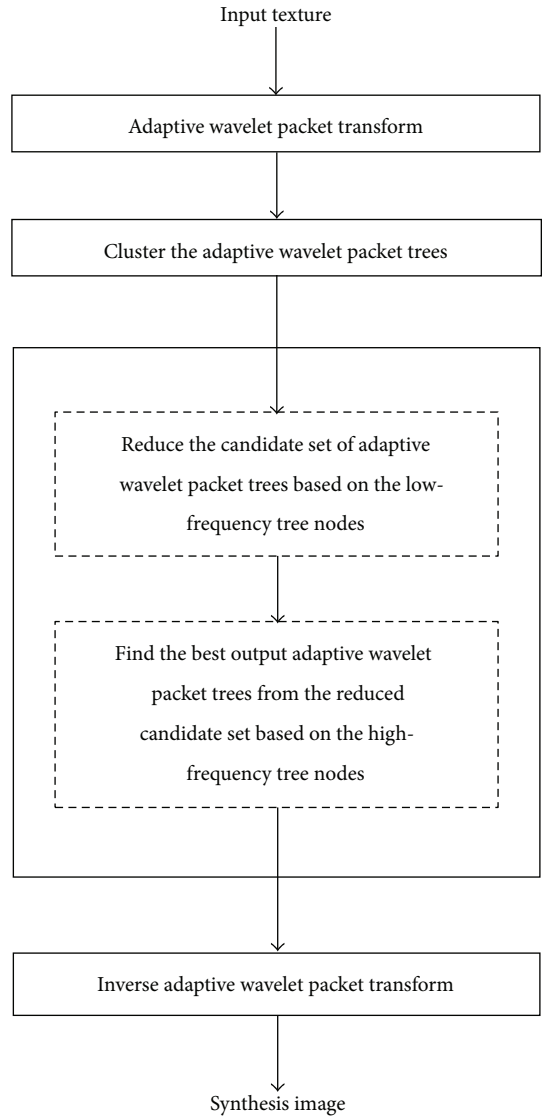


Figure 3: Flowchart of the TSIAWPT algorithm.

function (ACF) given by

$$R(m,n) = \frac{E\left[P_0(x,y)\,P_0(x+m,y+n)\right]}{E\left[P_0^2(x,y)\right]}, \tag{4}$$

where $P_0(x,y)$ is an adaptive wavelet packet coefficient at position (x,y) in the lowest-frequency subband 0 and m and n are the displacements in the horizontal and vertical directions, respectively. The ACF-based patch size $S_x \times S_y$ is determined by

$$\begin{aligned} S_x \times S_y &= \arg\max_{m,n} R(m,n), \\ \left\lfloor \frac{M_0}{8} \right\rfloor \le m \le \left\lfloor \frac{M_0}{2} \right\rfloor, &\qquad \left\lfloor \frac{N_0}{8} \right\rfloor \le n \le \left\lfloor \frac{N_0}{2} \right\rfloor, \end{aligned} \tag{5}$$

where $M_0 \times N_0$ is the size of the lowest-frequency subband of the input texture and $\lfloor z \rfloor$ denotes the largest integer less than z.

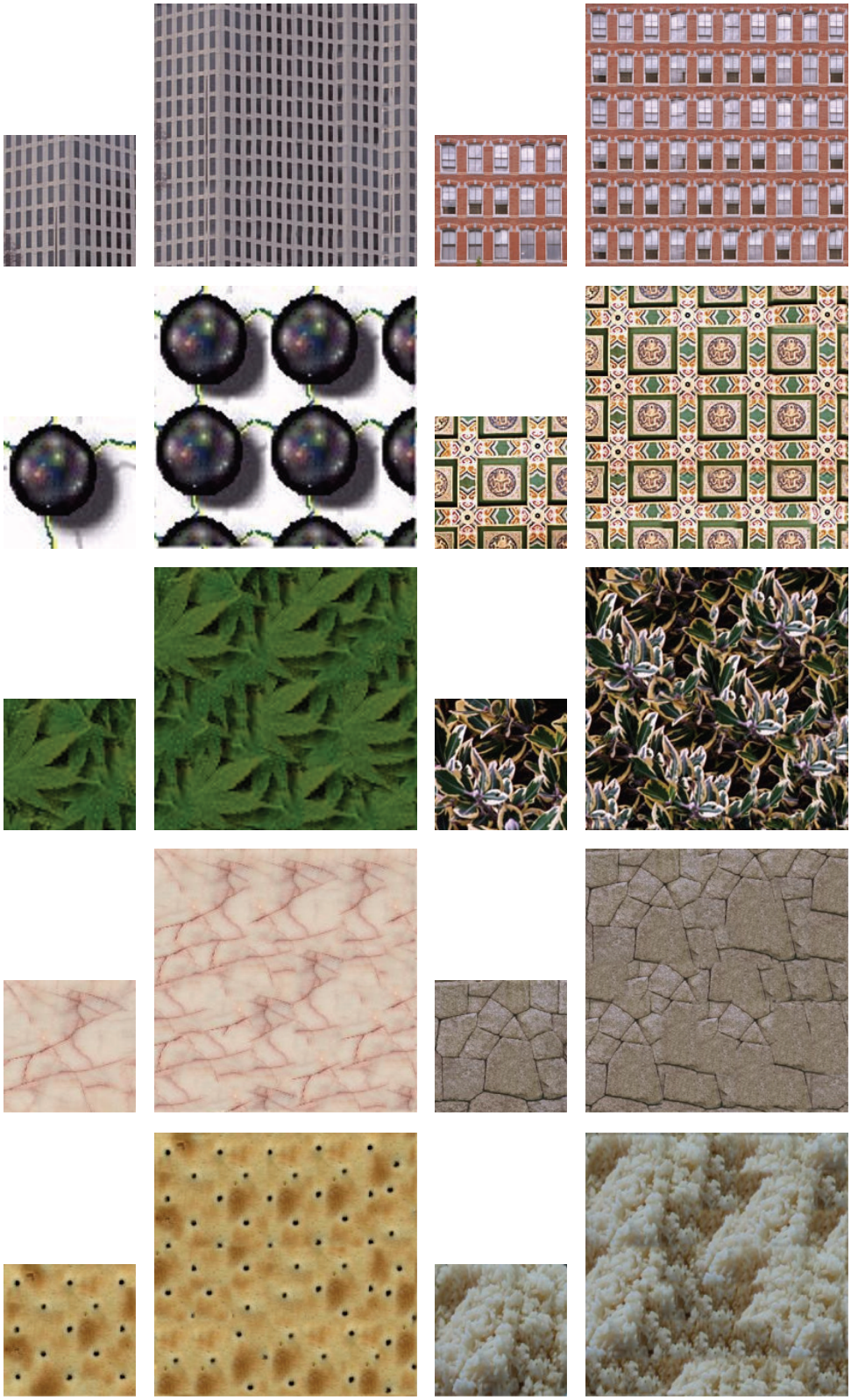


Figure 4: More synthesis results using TSIAWPT: architectures (1st row), regular structures (2nd row), leaves (3rd row), raw materials (4th row), and man-made materials (5th row).

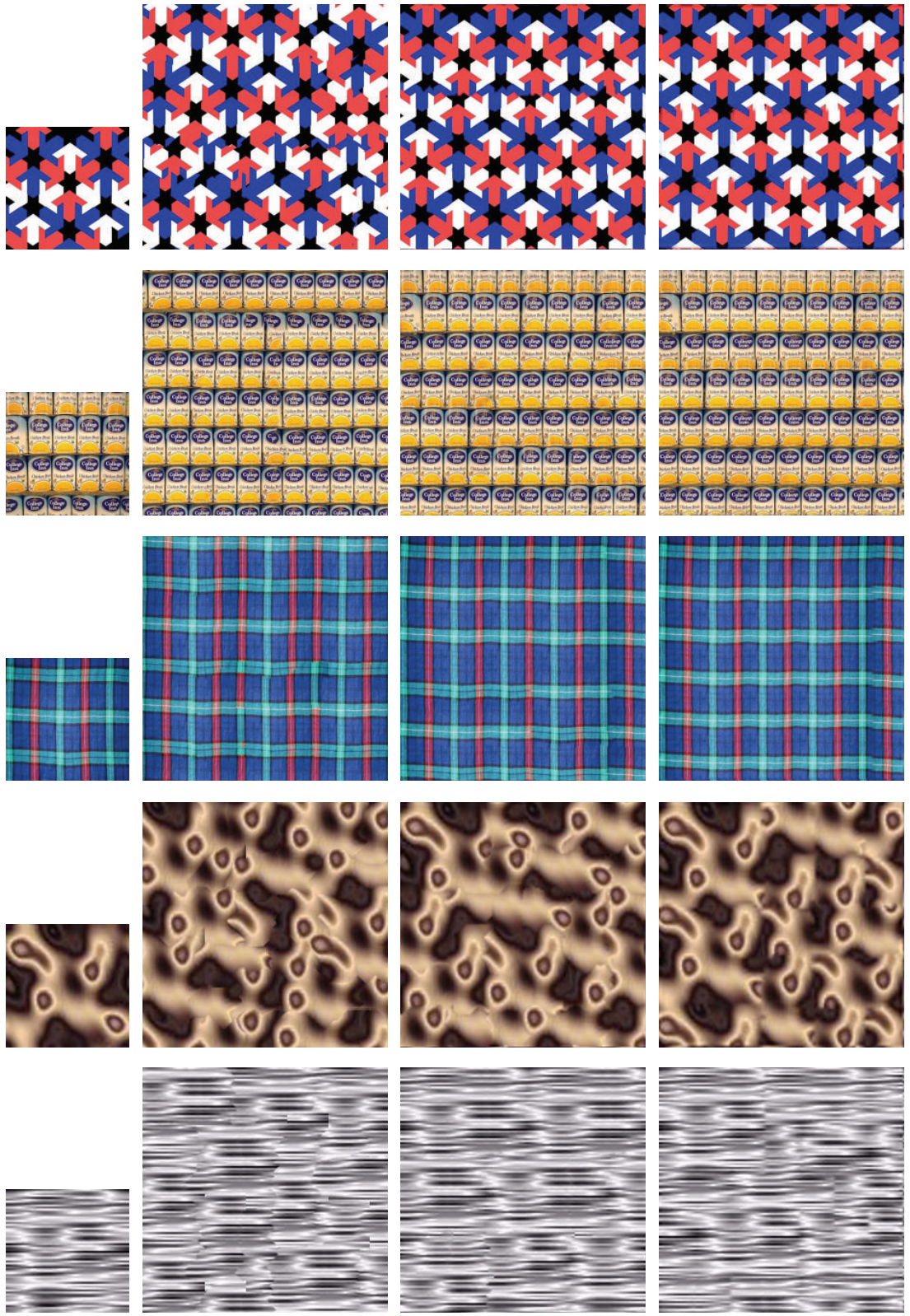


Figure 5: Synthesis results of the input images (1st column) using Efros's algorithm [11] (2nd column), Cui's algorithm [33] (3rd column), and the TSIAWPT algorithm (4th column).

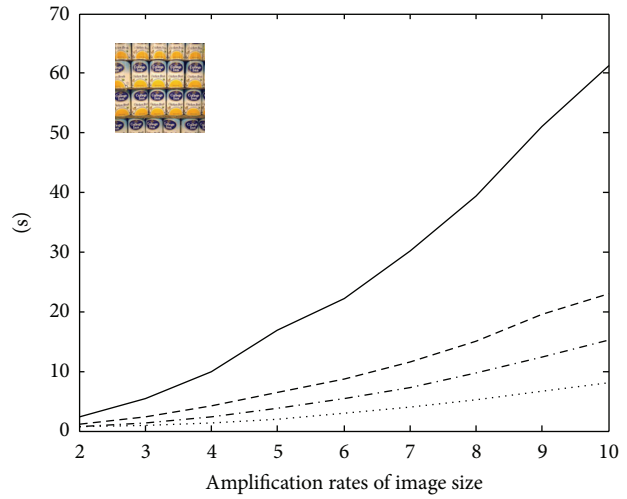


Figure 6: Computation times of Efros's algorithm [11] (solid line), Cui's algorithm [33] (dashed line), TSIWPT [35] (dashdot line), and TSIAWPT (dotted line).

4. Experimental Results

In our experiments, the commonly used 9/7-wavelet filters adopted in the JPEG2000 standard are used to decompose the input image into adaptive wavelet packets; the size of the lowest-frequency subband is 32×32; in other words, the number of decomposition levels is 3 for 256×256 images; rectangular patches with the ACF-based patch sizes are used; the width of overlapping borders is set to one-sixth of the patch size; the k-means algorithm is used to cluster the input AWPT into similarity sets.

In the first experiment, we demonstrate that the color distortion caused by using TSIWPT can be avoided by using TSIAWPT, as shown in Figures 2(b) and 2(c), respectively. More synthesis results using TSIAWPT for various types of textures, namely architectures, regular structures, leaves, raw materials, and man-made materials, are given in Figure 4.

The second experiment is to compare the TSIAWPT algorithm with two other well-known algorithms, Efros's algorithm [11] and Cui's algorithm [33]. The size of synthesis images is twice the size of the input images. Figure 5 shows the synthesis results. It is noted that TSIAWPT outperforms Efros's algorithm and is marginally preferable to the Cui's algorithm. Moreover, there are some blocking defects in the fourth and fifth synthesis images using Efros's algorithm. As the last step of TSIAWPT is inverse adaptive wavelet packet transform, which essentially involves low pass filtering, there is almost no blocking defect using TSIAWPT.

The last experiment is to evaluate the running speed. Figure 6 shows the computation times required to synthesize images with different sizes using Efros's algorithm, the Cui's algorithm, TSIWPT, and TSIAWPT. All of the above were simulated on a PC equipped with a CPU of 1.73 GHz and 4 GB of RAM. It is noted that TSIAWPT is preferable to the others in terms of computation time, especially for synthesizing large textures.

5. Conclusion

Adaptive wavelet packet transform provides a more compact representation for textured images, and moreover they can be efficiently represented in adaptive wavelet packet trees with the same hierarchical structure as the conventional wavelet trees. An efficient algorithm, TSIAWPT, has been proposed to synthesize textures in adaptive wavelet packet trees. It has the advantage of reducing computation time substantially and there is no training process involved. Specifically, the average time required to synthesize an 256×256 image from an 128×128 input texture is in a fraction of a second.

Acknowledgment

The National Science Council of Taiwan, under Grants NSC100-2628-E-239-002-MY2, and NSC100-2410-H-216-003 supported this work.

References

[1] S.-P. Du, S.-M. Hu, and R. R. Martin, "Semi-regular solid texturing from 2D image exemplars," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 3, pp. 460–469, 2012.

[2] M. Petrou, M. H. Jaward, S. Chen, and M. Briers, "Super-resolution in practice: the complete pipeline from image capture to super-resolved subimage creation using a novel frame selection method," *Machine Vision and Applications*, vol. 23, no. 3, pp. 441–459, 2011.

[3] J. Zhang, S. Y. Chen, S. Liu, and Q. Guan, "Normalized weighted shape context and its application in feature-based matching," *Optical Engineering*, vol. 47, no. 9, Article ID 097201, 2008.

[4] M. A. Akhloufi, X. Maldague, and W. B. Larbi, "A new color-texture approach for industrial products inspection," *Journal of Multimedia*, vol. 3, no. 3, pp. 44–51, 2008.

[5] S. Lefebvre and H. Hoppe, "Appearance-space texture synthesis," in *Proceedings of the ACM Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH '06)*, pp. 541–548, Boston, Mass, USA, August 2006.

[6] L. Ritter, W. Li, B. Curless, M. Agrawala, and D. Salesin, "Painting with texture," in *Proceedings of the 17th Eurographics Symposium on Rendering*, pp. 371–376, Nicosa, Cyprus, 2006.

[7] V. Kwatra, D. Adalsteinsson, T. Kim, N. Kwatra, M. Carlson, and M. C. Lin, "Texturing fluids," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 5, pp. 939–952, 2007.

[8] L. Y. Wei, S. Lefebvre, V. Kwatra, and G. Turk, "State of the art in example-based texture synthesis," in *Proceedings of the Eurographics State of the Art Reports (STARs)*, Eurographics, March 2009.

[9] N. Pietroni, P. Cignoni, M. Otaduy, and R. Scopigno, "Solid-texture synthesis: a survey," *IEEE Computer Graphics and Applications*, vol. 30, no. 4, pp. 74–89, 2010.

[10] Y. Xu, B. Guo, and H. Y. Shum, "Chaos mosaic: fast and memory efficient texture synthesis," Tech. Rep. MSR-TR-2000-32, Microsoft Research, 2000.

[11] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*, pp. 341–346, August 2001.

[12] M. F. Cohen, J. Shade, S. Hiller, and O. Deussen, "Wang tiles for image and texture generation," in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '03)*, pp. 287–294, San Diego, Calif, USA, July 2003.

[13] G. Cheng and J. Dong, "Seamless montage of natural texture," in *3rd International Conference on Advanced Computer Control (ICACC '11)*, pp. 48–51, Harbin, China, 2011.

[14] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.

[15] Q. Guan, B. Du, Z. Teng, J. Gillard, and S. Chen, "Bayes clustering and structural support vector machines for segmentation of carotid artery plaques in multi-contrast MRI," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 549102, 6 pages, 2012.

[16] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proceedings of the 7th IEEE International Conference on Computer Vision, (ICCV '99)*, vol. 2, pp. 1033–1038, Kerkyra, Greece.

[17] L. Y. Wei and M. Levoy, "Order independent texture synthesis," Tech. Rep. TR-2002-01, Stanford Computer Science, 2002.

[18] L. Liang, C. Liu, Y. Xu, B. Guo, and H. Y. Shum, "Real-time texture synthesis using patch-based sampling," *ACM Transactions on Graphics*, vol. 20, no. 3, pp. 127–150, 2001.

[19] V. Kwatra, I. Essa, A. Bobick, and N. Kwatra, "Texture optimization for example-based synthesis," in *Proceedings of the ACM Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH '05)*, pp. 795–802, Los Angeles, Calif, USA, August 2005.

[20] J. Han, K. Zhou, L. Y. Wei et al., "Fast example-based surface texture synthesis via discrete optimization," *Visual Computer*, vol. 22, no. 9–11, pp. 918–925, 2006.

[21] F. Dong and X. Ye, "Multiscaled texture synthesis using multisized pixel neighborhoods," *IEEE Computer Graphics and Applications*, vol. 27, no. 3, pp. 41–47, 2007.

[22] C.-W. Fang and J.-J. J. Lien, "Rapid image completion system using multiresolution patch-based directional and nondirectional approaches," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2769–2779, 2009.

[23] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[24] J. S. De Bonet, "Multiresolution sampling procedure for analysis and synthesis of texture images," in *Proceedings of the ACM Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH '06)*, pp. 361–368, Los Angeles, Calif, USA, August 1997.

[25] P. J. Burt, "Fast algorithms for estimating local image properties," *Computer Vision, Graphics, & Image Processing*, vol. 21, no. 3, pp. 368–382, 1983.

[26] L. Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in *Proceedings of the ACM Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH '00)*, pp. 479–488, New Orleans, La, USA, July 2000.

[27] J. C. Goswami and A. K. Chan, *Fundamentals of Wavelets: Theory, Algorithms, and Applications*, Wiley Series in Microwave and Optical Engineering, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2011.

[28] S. C. Lim, C. H. Eab, K. H. Mak, M. Li, and S. Y. Chen, "Solving linear coupled fractional differential equations by direct operational method and some applications," *Mathematical Problems in Engineering*, vol. 2012, Article ID 653939, 28 pages, 2012.

[29] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.

[30] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, vol. 2010, Article ID 408418, 22 pages, 2010.

[31] A. Abbate, C. M. DeCusatis, and P. K. Das, *Wavelets and Subbands: Fundamentals and Applications*, Applied and Numerical Harmonic Analysis, Springer, Boston, Mass, USA, 2002.

[32] Y. Yu, J. Luo, and C. W. Chen, "Multiresolution block sampling based method for texture synthesis," in *Proceedings of the 16th International Conference on Pattern Recognition*, pp. 239–242, 2002.

[33] H. F. Cui, X. Zheng, and T. Ruan, "An efficient texture synthesis algorithm based on WT," in *Proceedings of the 7th International Conference on Machine Learning and Cybernetics (ICMLC '08)*, pp. 3472–3477, Kunming, China, July 2008.

[34] T. Y. Sung and H. C. Hsin, "An efficient rearrangement of wavelet packet coefficients for embedded image coding based on SPIHT algorithm," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E90-A, no. 9, pp. 2014–2020, 2007.

[35] H. C. Hsin, T.-Y. Sung, Y.-S. Shieh, and C. Cattani, "A new texture synthesis algorithm based on wavelet packet tree," *Mathematical Problems in Engineering*, vol. 2012, Article ID 305384, 12 pages, 2012.

[36] H. C. Hsin, T. Y. Sung, and L. T. Ko, "A fast Wavelet packet based algorithm for texture synthesis," in *21st International Conference on Pattern Recognition (ICPR '12)*, pp. 3124–3127, Proceedings of the Tsukuba, Japan, 2012.

[37] N. M. Rajpoot, R. G. Wilson, F. G. Meyer, and R. R. Coifman, "Adaptive wavelet packet basis selection for zerotree image coding," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1460–1472, 2003.

[38] H. C. Hsin and T. Y. Sung, "Adaptive selection and rearrangement of wavelet packets for quad-tree image coding," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E91-A, no. 9, pp. 2655–2662, 2008.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 175769, 3 pages
http://dx.doi.org/10.1155/2013/175769

Research Article

New Inequalities between Information Measures of Network Information Content

Pantelimon-George Popescu, Florin Pop, Alexandru Herişanu, and Nicolae Ţăpuş

Computer Science Department, Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Splaiul Independentei 313, Bucharest 060042, Romania

Correspondence should be addressed to Florin Pop; florin.pop@cs.pub.ro

Received 12 December 2012; Accepted 28 January 2013

Academic Editor: Carlo Cattani

We refine a classical logarithmic inequality using a discrete case of Bernoulli inequality, and then we refine furthermore two information inequalities between information measures for graphs, based on information functionals, presented by Dehmer and Mowshowitz in (2010) as Theorems 4.7 and 4.8. The inequalities refer to entropy-based measures of network information content and have a great impact for information processing in complex networks (a subarea of research in modeling of complex systems).

1. Introduction

We first present the classical Shannon's entropy; see [1, 2]. Let X be a discrete random variable taking values in A, and let $p(x_i) = Pr(X = x_i)$ be the probability mass function of X; then, the Shannon entropy of X is defined by

$$H(X) := -\sum_{x_i \in A} p(x_i) \log(p(x_i)). \quad (1)$$

We continue with some basic theoretical information about graphs; see [3, 4]. We consider $G = (V, E)$, $|V| < \infty$ to be a finite undirected graph, where $E \subseteq \binom{V}{2}$. G is connected if for any arbitrary vertices v_i, v_j exists an undirected path from v_i to v_j. Furthermore, we will consider $\mathcal{G}_{UC}$ to be the set of finite undirected and connected graphs. This is usually used in strategies for improving transport efficiency in Network Information Content, including designing efficient routing strategies and making appropriate adjustments to the underlying network structure [5].

Now, taking into account the work presented in [6, 7], we define the concept of information functional. Let $G = (V, E) \in \mathcal{G}_{UC}$, and let S initially be an abstract set; then, $f : S \to R^+$ is called the information functional of G (we assume that f is monotonous). The abstract set S can be, for example, the vertex sets, a set of paths, certain subgraphs, and so forth. The set S is used to define the functional f that captures the structural information of graph G and has to be defined concretely.

Using different information functionals to measure the entropy of a graph, we obtain different probability distributions; so, the resulting graph entropies are also different. Further, we define, as in [6, 7], the entropy of a graph $G = (V, E) \in \mathcal{G}_{UC}$ with arbitrary vertex labels v_i, using an arbitrary information functional f, to be

$$I^f(G) := -\sum_{i=1}^{|V|} \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \log\left(\frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}\right), \quad (2)$$

and considering a vertex $v_i \in V$, $p^f(v_i) = f(v_i)/\sum_{j=1}^{|V|} f(v_j)$ interpreted as vertex probabilities ($\sum_{i=1}^{|V|} p^f(v_i) = 1$), we conclude that

$$I^f(G) = -\sum_{i=1}^{|V|} p^f(v_i) \log\left(p^f(v_i)\right). \quad (3)$$

2. Refinement of a Classical Logarithmic Inequality

It is well known from the literature that

$$\log_b(x) \le \frac{1}{\ln(b)}(x-1), \quad \forall x > 0. \quad (4)$$

Trying to refine this inequality, we consider the classical discrete Bernoulli inequality, as follows.

Theorem 1 (Bernoulli inequality). *Let $a_i \geq 0$, $x_i > -1$, $i = 1, \ldots, n$ and $\sum_{i=1}^{n} a_i \leq 1$. Then,*

$$\prod_{i=1}^{n}(1+x_i)^{a_i} \leq 1 + \sum_{i=1}^{n} a_i x_i, \tag{5}$$

if $a_i \geq 1$ or $a_i \leq 0$, and if $x_i > 0$, or $-1 < x_i < 0$, $i = 1, \ldots, n$, then

$$\prod_{i=1}^{n}(1+x_i)^{a_i} \geq 1 + \sum_{i=1}^{n} a_i x_i. \tag{6}$$

Now, using a factorization of a real number, one concludes as follows.

Theorem 2. *Let $x > 0$ be a real number; then,*

$$\log_b(x) \leq \frac{1}{\ln(b)} \sum_{i=1}^{n} x_i \leq \frac{1}{\ln(b)}(x-1), \tag{7}$$

where $n \geq 2$ and $x = \prod_{i=1}^{n}(x_i + 1)$, with $x_i > 0$ if $x > 1$, $x_i = 0$ if $x = 1$, and $0 > x_i > -1$ if $1 > x > 0$, $\forall 1 \leq i \leq n$.

Proof. The inequality becomes an identity for $x = 1$; that is, $x_i = 0$, $\forall 1 \leq i \leq n$. We will treat the other two cases, when $x > 1$ and when $1 > x > 0$ in the same manner, as follows:

$$\log_b(x) = \log_b \prod_{i=1}^{n}(x_i + 1) = \sum_{i=1}^{n} \log_b(x_i + 1), \tag{8}$$

for which applying the classical logarithmic inequality for $x_i + 1 > 0$, $\forall 1 \leq i \leq n$ yields

$$\log_b(x) \leq \frac{1}{\ln(b)} \sum_{i=1}^{n} x_i, \tag{9}$$

and making use of the second part of the discrete Bernoulli inequality, previously presented, we obtain

$$\frac{1}{\ln(b)} \sum_{i=1}^{n} x_i \leq \frac{1}{\ln(b)} \left[\prod_{i=1}^{n}(x_i + 1) - 1 \right] = \frac{1}{\ln(b)}(x-1), \tag{10}$$

and we are done. □

3. Refinement of Dehmer and Mowshowitz Inequalities between Information Measures

In the work of Dehmer and Mowshowitz [6], by Theorems 4.7 and 4.8 are presented two information inequalities derived assuming only the characteristic properties of the functions involved. Here, we refine those results by using the previous theorem, as follows.

Theorem 3. *Let f and f^* be information functionals. Then (all logarithms are in base 2),*

$$\begin{aligned} I^f(G) \geq & -\sum_{k=1}^{|V|} p^f(v_k) \log\left(p^{f^*}(v_k)\right) \\ & - \frac{1}{\ln(2)} \sum_{i=1}^{n} p_i^k \sum_{k=1}^{|V|} p^f(v_k) \\ \geq & -\sum_{k=1}^{|V|} p^f(v_k) \log\left(p^{f^*}(v_k)\right) \\ & - \frac{1}{\ln(2)} \sum_{k=1}^{|V|} \frac{\left(p^f(v_k)\right)^2 - p^f(v_k)\, p^{f^*}(v_k)}{p^{f^*}(v_k)}, \end{aligned} \tag{11}$$

$$\begin{aligned} I^{f^*}(G) \leq & \frac{1}{\ln(2)} \sum_{i=1}^{n} p_i^k \sum_{k=1}^{|V|} p^{f^*}(v_k) \\ & - \sum_{k=1}^{|V|} p^{f^*}(v_k) \log\left(p^f(v_k)\right) \\ \leq & \frac{1}{\ln(2)} \sum_{k=1}^{|V|} \Big[p^f(v_k) - p^{f^*}(v_k) \\ & - \ln(2)\, p^{f^*}(v_k) \log\left(p^f(v_k)\right) \Big], \end{aligned}$$

where $n \geq 2$ and $p^f(v_k)/p^{f^*}(v_k) = \prod_{i=1}^{n}(p_i^k + 1)$, $\forall 1 \leq k \leq |V|$, with $\forall 1 \leq k \leq |V|$, $p_i^k > 0$ if $p^f(v_k) > p^{f^*}(v_k)$, $p_i^k = 0$ if $p^f(v_k) = p^{f^*}(v_k)$, and $0 > p_i^k > -1$ if $p^f(v_k) < p^{f^*}(v_k)$, $\forall 1 \leq i \leq n$.

Proof. Applying Theorem 2 for $x = p^f(v_k)/p^{f^*}(v_k) > 0$ yields that

$$\begin{aligned} & \log\left(p^f(v_k)\right) - \log\left(p^{f^*}(v_k)\right) \\ & \leq \frac{1}{\ln(2)} \sum_{i=1}^{n} p_i^k \\ & \leq \frac{1}{\ln(2)} \frac{p^f(v_k) - p^{f^*}(v_k)}{p^{f^*}(v_k)}, \quad \forall 1 \leq k \leq |V|. \end{aligned} \tag{12}$$

By multiplying this inequality first with $-p^f(v_k)$ and then with $p^{f^*}(v_k)$, we obtain

$$\begin{aligned} & -p^f(v_k) \log\left(p^f(v_k)\right) \\ & \geq -\frac{1}{\ln(2)} \sum_{i=1}^{n} p_i^k \cdot p^f(v_k) \\ & \quad - p^f(v_k) \log\left(p^{f^*}(v_k)\right) \end{aligned}$$

$$
\begin{aligned}
&\geq -\frac{1}{\ln(2)}\frac{\left(p^{f}(v_k)\right)^2 - p^{f}(v_k)\,p^{f^*}(v_k)}{p^{f^*}(v_k)} \\
&\quad - p^{f}(v_k)\log\left(p^{f^*}(v_k)\right), \\
&- p^{f^*}(v_k)\log\left(p^{f^*}(v_k)\right) \\
&\quad \leq \frac{1}{\ln(2)}\sum_{i=1}^{n} p_i^k \cdot p^{f^*}(v_k) \\
&\quad\quad - p^{f^*}(v_k)\log\left(p^{f}(v_k)\right) \\
&\quad \leq \frac{1}{\ln(2)}\left(p^{f}(v_k) - p^{f^*}(v_k)\right) \\
&\quad\quad - p^{f^*}(v_k)\log\left(p^{f}(v_k)\right), \\
&\quad\quad \forall 1 \leq k \leq |V|.
\end{aligned}
\tag{13}
$$

Finally, summating after k yields the wanted results. □

4. Conclusions

As a main result of this paper, we improved two inequalities between entropy-based measures, this giving us a way of bounding measures of network properties. A good example for this purpose is presented into [6], where, using concrete information functionals, the corresponding entropies for the graphs in a set of 2265 nonisomorphic chemical graphs, that is, in MS2265, are computed. Here, new stronger lower and, respectively, upper bounds for entropy-based measures are given and thus, by knowing the limitations of such measures, we narrow down the set of expected values. The utility of inequalities between measures of structural information content is found in problem solving through problem transformation, as is also spotted in [6].

Acknowledgment

This work was supported and founded by the project "ERRIC-Empowering Romanian Research on Intelligent Information Technologies/FP7-REGPOT-2010-1", ID: 264207.

References

[1] R. Ash, *Information Theory*, Interscience, New York, NY, USA, 1965.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2006.

[3] F. Buckley and F. Harary, *Distance in Graphs*, Addison-Wesley, 1990.

[4] F. Harary, *Graph Theory*, Addison-Wesley, Reading, Mass, USA, 1969.

[5] S. Chen, W. Huang, C. Cattani, and G. Altieri, "Traffic dynamics on complex networks: a survey," *Mathematical Problems in Engineering*, vol. 2012, Article ID 732698, 23 pages, 2012.

[6] M. Dehmer and A. Mowshowitz, "Inequalities for entropy-based measures of network information content," *Applied Mathematics and Computation*, vol. 215, no. 12, pp. 4263–4271, 2010.

[7] M. Dehmer, "Information processing in complex networks: graph entropy and information functionals," *Applied Mathematics and Computation*, vol. 201, no. 1-2, pp. 82–94, 2008.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 607582, 10 pages
http://dx.doi.org/10.1155/2013/607582

Research Article

A Bilayer Resource Model for Cloud Manufacturing Services

Linan Zhu,[1,2,3] Yanwei Zhao,[1] and Wanliang Wang[2]

[1] *Key Laboratory of Special Purpose Equipment and Advanced Processing Technology, Ministry of Education, Zhejiang University of Technology, Hangzhou 310014, China*
[2] *School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*
[3] *College of Educational Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Yanwei Zhao; zyw@zjut.edu.cn

Received 15 November 2012; Accepted 3 January 2013

Academic Editor: Carlo Cattani

Cloud Manufacturing and Cloud Service is currently one of the main directions of development in the manufacturing industry. Under the Cloud Manufacturing environment, the characteristics of publishing, updating, searching, and accessing manufacturing resources are massive, complex, heterogeneous, and so forth. A bilayer manufacturing resource model with separation of Cloud End and Cloud Manufacturing Platform is proposed in this paper. In Cloud End, manufacturing resources are divided into single resource and complex resource, and a basic data model of manufacturing resources oriented to enterprise interior is established to store the physical characteristics. In Cloud Manufacturing Platform, a resource service attribute model oriented to actual users is established to store the service characteristics. This model is described in detail and realized with stateful Web Service Description Language (WSDL) document. An example is provided for illustrating the implementation of the concept.

1. Introduction

The development and transformation of manufacturing has promoted the sustainable development of human society. At present, along with the generation and development of computer science and Internet, networked manufacturing, which contains production of material products and offering of immaterial services or functionality, has become the main form of manufacturing [1]. Attributed to Cloud Computing theory and application, Cloud Manufacturing and Cloud Service have gradually risen and become the main direction of manufacturing industry.

Presently, there is not a standardized definition of Cloud Manufacturing in academia. Li et al. believe that Cloud Manufacturing is a new networked manufacturing model that provides users with customized manufacturing services by organizing online manufacturing resource (called resource cloud) with the use of Internet and Cloud Manufacturing Service Platform [2]. They also put forward that Cloud Manufacturing is a service-oriented, high efficiency and low consumption, networked, and agile manufacturing model and technology. It enriches and expands Cloud Computing in two aspects of shared resource contents and service models, so it makes the manufacturing model become more agile, servicesation, environment friendly and intelligence [3]. According to Yang's point of view (China Aerospace Science and Technology Corporation) [4], the advantage of Cloud Manufacturing is that we can expand the philosophy of "Software as a Service" to "Manufacturing as a Service," so as to offer products with such services as high value-added, low cost, and global manufacturing under network environment. The authors of [5] believe that, by integration with contemporary technologies such as manufacturing informatization, Cloud Computing, Internet of Things, semantic web, and high-performance computing, Cloud Manufacturing expands and innovates the existing technology of networked manufacturing and service, and makes manufacturing resources and manufacturing capabilities become more virtualization and servicesation. Thus we can manage and operate the resources and service unified, centralized, and intelligently, to provide available, on-demand, reliable, and high-quality and low-cost services in

every phase of manufacturing lifecycle. Cloud Manufacturing reflects the idea of "distributed resources are integrated to be used for one task" and "integrated resources are distributed to be used," and achieves the many-to-many service model, which provides multiple users with services at the same time by aggregating and centralized managing of distributed resources and services. From the perspective of product life cycle, the authors of [6] believe that Cloud Manufacturing, as a kind of manufacturing service, has its own lifecycle with several phases: the definition of manufacturing resource or capability, the provision of manufacturing resource or capability, the ordering of manufacturing task, manufacturing and distribution, and the disposal of manufacturing task.

Along with service-oriented technology, virtualization technology, Cloud Computing, and Internet of Things [7], the rapid development of manufacturing gives birth to a new model called Cloud Manufacturing Service Model (CMSM), which is a kind of service-oriented, high efficiency and low consumption, and knowledge-based networked intelligent manufacturing [2]. The CMSM has been actually accepted by some enterprises worldwide. In 2000 of America, MFG.com provided global manufacturing enterprises with an efficient trading platform, and the total value of average daily inquiry was more than 400 million yuan. Also in America, thanks to a brand new operation model, a micro-factory with the size of dry cleaner Local-Motors.com created a distinctive concept car called Rally Fighter in only 20 months. This is such a short time just equivalent to the time Detroit need to adjust the technical specifications of car door, because the total design tasks of the car were outsourced to community online, and the only task of Local-Motors.com was assembling parts, most of which were available on the market. In 2010 of China, a website for cutlery was built, which was a service system to help enterprises carry out cutlery sale and customer management better. The users of the website are the manufacturing enterprises in the supply chain or manufacturing chain of cutlery industry. Because of the website, the resource allocation involved in many cutlery enterprise supply chain and manufacturing chain is optimized.

The so-called Cloud Service is namely Resource Cloud Service, with some features: huge amounts of data, on-demand scaling, high availability, self-service interface, and flexible use of resources. Through the technologies such as Internet of Things and virtualization, Cloud Service virtually encapsulates manufacturing resources and manufacturing capabilities in different Cloud End based on knowledge, and intelligently accesses to Cloud Manufacturing Platform, thus provide users with manufacturing resources highly virtualized in Cloud End as services in the manufacturing life cycle [8].

2. Cloud Manufacturing System Structure

Showed as Figure 1, the Cloud Manufacturing System (CMS) is composed of Cloud End (CE) and Cloud Manufacturing Platform (CMP), and the CE contains Cloud Provider (CP) and Cloud Demander (CD). The CP and CD are, respectively, the providers and the demanders or users of resource cloud or cloud service. The CP provides corresponding manufacturing resource and service through the CMP; the CD proposes manufacturing demands and gets corresponding resources or services through CMP; according to the CD, the CMP searches suitable resources or services, and provides the CD with demanded resources or services by the use of Cloud technology, Cloud Services management technology, Cloud Manufacturing security technology, and Cloud Manufacturing business management model and technology [2]. The CMP is also composed of many sub-platforms, which can communicate with each other to realize high sharing of resources. Also, the CMP embodies many modules such as resource database and middleware, which have powerful functions of resource scheduling.

3. Networked Manufacturing Resource

Resource description is distinctive in different manufacturing areas and with different modeling aims [9]. Cloud Manufacturing is actually a kind of complex network environment with the characteristics of large-scale, heterogeneous, and high sharing of resources. Under the network environment, the key points of manufacturing resource description are dynamic expression of resource service capabilities, rapid search of resources, optimal allocation of resource, and dynamic planning of the product life cycle.

3.1. Overview on Networked Manufacturing Resource Modeling. In [10], a notion named service domain was presented to help UML analyze and model manufacturing resource in manufacturing grid, and supported hierarchy management of resource model, and the results resolved the disadvantages that modeled huge manufacturing resource in manufacturing grid through only a set of UML class graph. In [11], the authors studied on the multi-dimension manufacturing resource modeling technology: the manufacturing resource life cycle dimension, manufacturing resource application view dimension, and polymerization granularity dimension, and established a multi-dimensional network ontology model of manufacturing resources through semantic web technology. In [12], a combined method of particle size based on ontology and business-oriented needs of internal and external information is proposed. The authors also proposed a model based on Web Process Planning, and expanded the OWL-S for describing the dynamic characteristics of the service. So the information model was set for discovering and matching the manufacturing resource. Aiming at the problem of manufacturing resource functionality similarity under nonlinear process planning environments, meta-resource methodology was introduced in [13], which could accurately express diversity of manufacturing resources and improve the agility of manufacturing system. For distributed networked manufacturing resources, a networked manufacturing resource model based on physical manufacturing unit, as well as an information model and an information integration method of manufacturing resource based on XML, were presented in [14], with the result of realization of heterogeneous

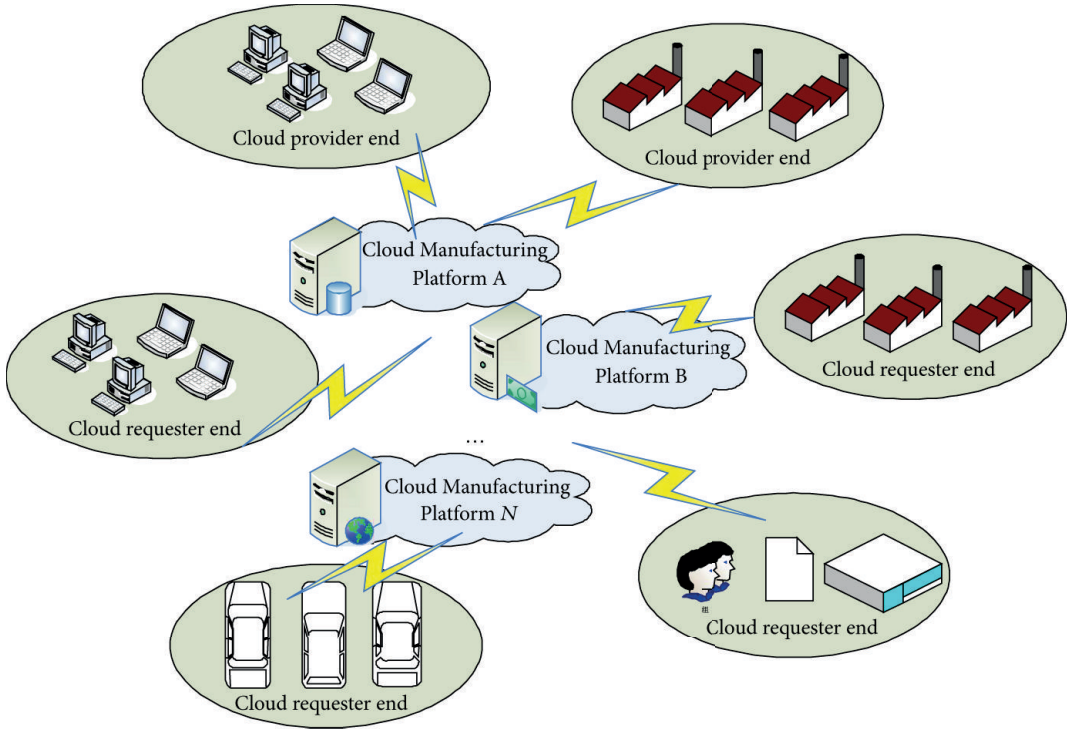


Figure 1: The system of Cloud Manufacturing.

manufacturing resource description. In order to implement the integration and application of legacy hardware/software manufacturing resources in manufacturing enterprises, a scheme and a framework for the manufacturing resource encapsulation and integration based on mobile agent was introduced in [15]. So the function of reconfiguration and encapsulation for legacy manufacturing resources, as well as information interaction and acquirement, could be realized.

3.2. Classification of Manufacturing Resources. With different manufacturing aims and different manufacturing methods, the classification of resource is different. In [12], the authors established resource ontology tree, and divided the resource into six categories: (1) financial resources: the financial elements of business operations including fixed assets, liquidity, and liabilities; (2) technical resources: process technology, industry standards, and so forth; (3) equipment resources: machine tools, tooling, gages, and so forth; (4) human resources: engineering and technical personnel, workers, technicians, management personnel, and so forth; (5) software resources: CAD software, financial software, ERP software, and so forth; (6) logistics resources: truck, train, and so forth. According to the sharing method of manufacturing resource under manufacturing grid environment, the authors of [16] divided resources into six categories: hardware resources, software + hardware resources, software resources, online resources, semionline resources, and offline resources.

In this paper, according to the properties of the resource, the users' needs, usage as well as the role played in manufacturing activities, we divide resources into eight categories (shown in Figure 2): human resources, manufacturing equipment resources, software resources, service resources, material resources, computing resources, manufacturing knowledge resources, and other resources. Human resource means the staff involved in all phases of product life cycle; manufacturing equipment resource means all kinds of hardware in the product design and production processing; software resource means computer software used in product design, manufacturing, enterprise management, and so forth; service resource means service activities related to product design, production marketing, and so forth; material resource means raw materials used in every phases of production; computing resource means computing equipment such as CPUs and memories used in production and enterprise management; knowledge resource means the knowledge or technology related to production.

4. The Bilayer Resource Model

Under Cloud Manufacturing environment, there are a wide range of manufacturing resources in different areas, and the systems in different manufacturing units are heterogeneous. So the resource data has the characteristics of massive, complex and heterogeneous, and there is lack of uniform data standards. For the above reasons, it is difficult to real-time monitor, operate, or update resource data. Aiming at these questions, we propose a bilayer resource model with separation of CE and CMP (shown in Figure 3). In CE, we establish the resource basic data model, which is used to store the basic data of resources such as performance parameters, physical structure, input data type, output data

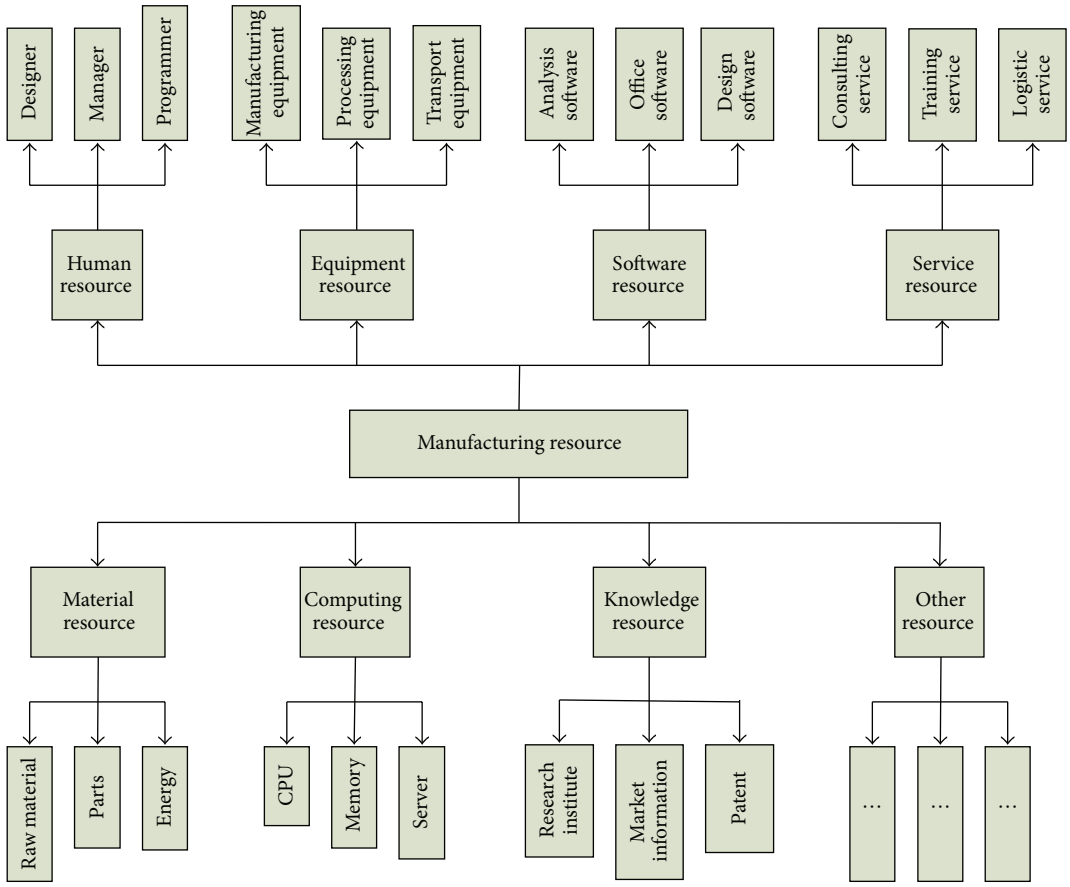


Figure 2: Manufacturing resource classification.

type, geographic information, and manufacturer information, and provides access interface to CMP. In CMP, we establish the resource service attribute model, which is used to store service information such as function parameters, input data template, output structure template, and service quality evaluation. In CMP, there are also some resource service solutions, integrated resource optimization tools, and database for storing data accessed frequently according to usage history.

4.1. Cloud End Model. As described above, resource model in CE is used to store related data of the actual manufacturing resource, which consists of single resource (SR) and complex resource (CR). According to the functionality, the SR is divided into eight categories, as described in Section 3.2; as a resource group, the CR is formed by combination of SR based on functionality and has some function. So, the formalized representation of resource cloud in CE is Re*sCloudEnd* = {*RCEInfo*, *RC*}.

RCEInfo means the basic information of the CE, and its formalized representation is *RCEInfo* = {*RCEID*, *RCE* Pro*vider*, *RCELocation*, *RCEOther*}. *RCEID* no longer changed after registration is the flag of CE, and it's used to uniquely identify CE so as to realize locating and indexing the manufacturing resource providers; *RCE* Pro*vider* is the name and contact information of CE, stored as a string; *RCELocation* is the location of CE, stored as a string; *RCEOther* is the custom information, which can be added or removed according to the actual situation, and the format of data is user-defined, usually as a string.

RC is resource cloud, and its formalized representation is: *RC* ::= (*SRC*, *CRC*). *SRC* is single resource cloud (SRC), and *CRC* is complex resource cloud (CRC).

4.1.1. Single Resource Cloud (SRC). The formalized representation of SRC is *SRC* = {*RC*ID, *SRC*BaseInfo, *SRC*FuncInfo, *SRC*StatusInfo}.

(1) *RCID* is the flag of SRC, used to uniquely identify manufacturing resource cloud to realize locating and indexing manufacturing resource cloud. The uniqueness is within the same CE, while the same kind of manufacturing resource in different CE has the different *RCIDs*.

(2) *SRCBaseInfo* is the basic characteristics set, and its formalized representation is *SRCBaseInfo* = {SRCName, SRCModel, SRCType, SRCInfo, SRCParameter, SRCBInfos}. *SRCName* is name of SRC; SRCModel is the model of SRC; SRCType is the type of SRC, and the formalized representation is *SRCType* = {1, 2, 3, 4, 5, 6, 7, 8}, every element of

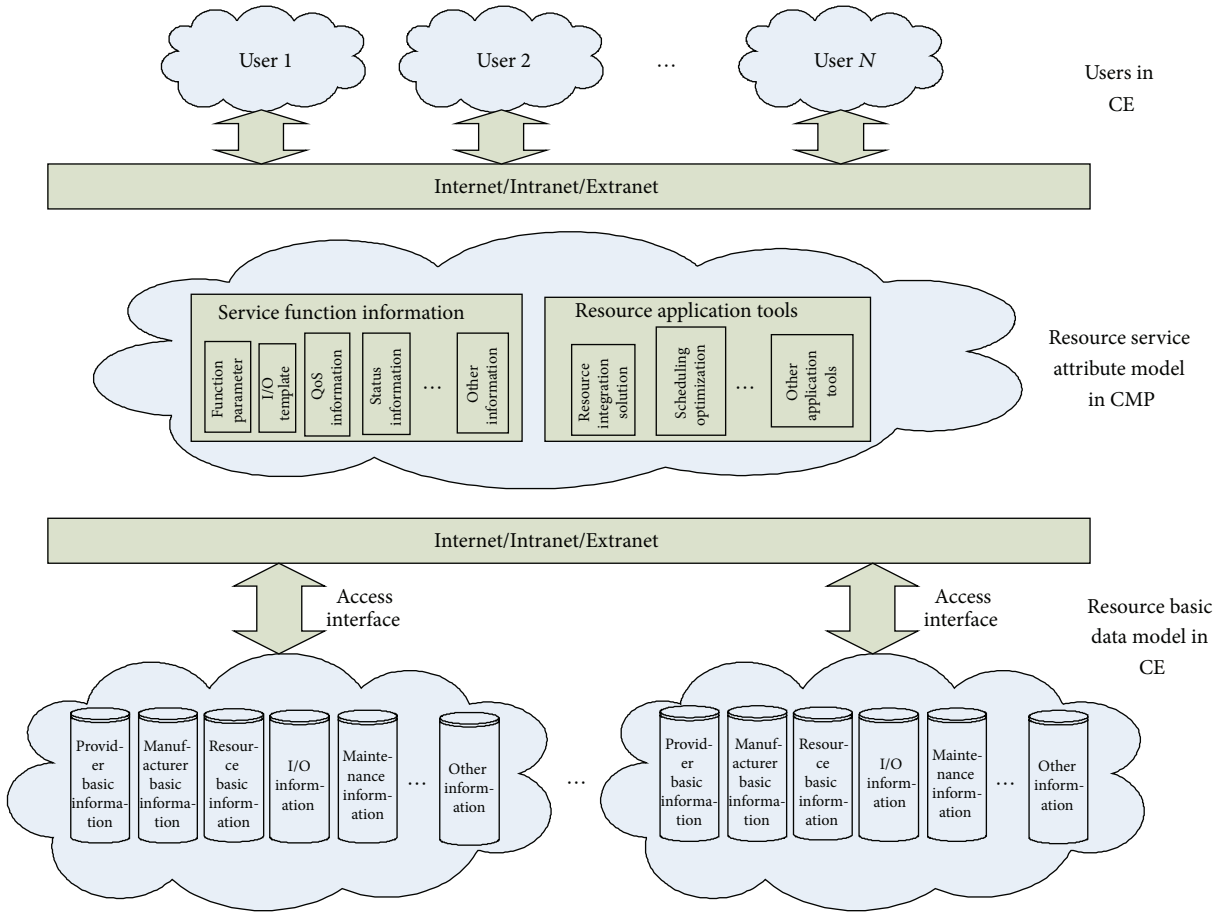


Figure 3: The bilayer resource model in Cloud Manufacturing.

SRCType, respectively, represents 8 categories of resources as described in Section 3.2; *SRCInfo* is the description of SRC, stored as a string; *SRCParameter* is the parameters set of SRC, including size, and cost, defined by the provider according to specification of resources; *SRCBInfos* is custom information, which can be added or removed according to the actual situation, and the format of data is user-defined, usually as a string.

(3) *SRCFuncInfo* is the functionality characteristics attributes set of SRC, and its formalized representation is: *SRCFuncInfo* = {*SRCTaskType*, *SRCCapa*, *SRCTime*, *SRCQua*, *SRCEnviro*, *SRCFInfos*}. *SRCTaskType* is the service type supported by the SRC, such as the type of processing object; *SRCCapa* is the indicators of the processing object, such as input set of object, output set of object, and precision, and its formalized representation is: $SRCCapa = \{IObjSet, OObjSet, \ldots, Capa_n\}$, every el- ement of *SRCCapa* has its own independent formalized representation. For example, $IObjSet = \{IObj_i \mid i = 1, 2, \ldots, n\}$, $OObjSet = \{OObj_i \mid i = 1, 2, \ldots, n\}$, every $IObj_i$ or $OObj_i$ is composed of some characteristic attributes: $IObj_i = \{ID, Name, P_1, P_2, \ldots, P_n\}$, $OObj_i = \{ID, Name, P_1, P_2, \ldots, P_n\}$; *SRCTime* is the minimum time and maximum time to complete the task, and its formalized representation is: $SRCTime = \{STime_{\min}, STime_{\max}\}$; *SRCQua* is the quality standards the resource can reach, and its formalized representation is: $SRCQua = \{N, level_1, level_2, level_3, \ldots, level_n\}$. N represents the sum of quality standards the offering can reach, $level_i$ store the name of quality standards as a string. Take the safety valve as an instance, the formalized representation is: $SRCQua = \{1, "GB150.1 - 2011"\}$; *SRCEnviro* is the service environment requirements of the SRC, such as geographical requirements, software platform requirements, and enterprise-class requirements; *SRCFInfos* is custom information, which can be added or removed according to the actual situation, and the format of data is user-defined, usually as a string.

(4) *SRCStatusInfo* is the status set of SRC, and its formalized representation is: $SRCStatusInfo = \{0, 1, 2, 3, 4, 5\}$. 0 represents that the resource is in maintenance; 1 is idle; 2 is under load; 3 is full load; 4 is over load; 5 is invalid.

4.1.2. Complex Resource Cloud (CRC). The formalized representation of CRC is: *CRC* = {*RC*ID, SRCSet, SRCRelSet, *CRCFunc*Info, *CRC*StatusInfo}.

(1) *RC*ID is the flag of CRC, and has the same characteristics with the flag of SRC.

(2) SRCSet is the SRCs set contained by the CRC, and its formalized representation is: SRCSet =

$\{N, SRC_1, SRC_2, \ldots, SRC_n\}$. N is the sum of SRCs, and SRC_i is the flag *RCID* of SRC.

(3) SRCRelSet is the logical relationships between these SRCs, and its formalization representation is: SRCRelSet ::= {Rand, Ror, Rprev, Rnext}. Rand is the parallel relationship. Rand(A, B) means that the resource A and B have no influence with each other, and can be used in parallel; Ror is the selection relationship. $Ror(A, B)$ means that the resource A and B have the same or similar functionalities, and are alternative in the actual manufacturing; Rprev and Rnext are the precursor relationship and successor relationship. $Rprev(A, B)$ means that A is the precursor of B, and $Rnext(A, B)$ means that A is the successor of B, so the relationship of resource A and B is serial. It's worth nothing that $Rprev(A, B) \neq Rnext(B, A)$, because the precursor relationship or successor relationship may be one to many or many to many.

(4) C*RCFunc*Info is functionality characteristics attributes set, and its formalized representation is: CRCFuncInfo = {CRCTaskType, CRCCapa, CRCTime, CRCQua, CRCEviro, CRCFInfos}. *CRCTaskType*, *CRCCapa*, *CRCQua*, *CRCEviro* and *CRCFInfos* respectively means supported service type, indicator requirements, quality standards, environment requirements and custom contents, with the same meaning and formalized representation with *SRCTaskType*, *SRCCapa*, *SRCQua*, *SRCEviro* and *SRCFInfos*of SRC. *CRCTime* is the minimum time and maximum time to complete the task, and its formalized representation is $CRCTime = \{CTime_{\min}, CTime_{\max}\}$. $CTime_{\min}$ is the sum of $STime_{\min}$ of all serial single resource units (the formula is shown as (1)), taking the minimum value of several parallel single resource (the formula is shown as (2)); $CTime_{\max}$ is the sum of $STime_{\max}$ of all serial single resource units (the formula is shown as (3)), taking the maximum value of several parallel single resource (the formula is shown as (4)):

$$\begin{aligned} CTime_{\min} = {} & SRC_1 \cdot STime_{\min} \\ & + SRC_2 \cdot STime_{\min} + \cdots \\ & + SRC_{i1,i2\ldots,ij} \cdot STime_{\min} + \cdots \\ & + SRC_n \cdot STime_{\min} \end{aligned} \tag{1}$$

$$\begin{aligned} SRC_{i1,i2\ldots,ij} \cdot STime_{\min} = {} & \min \{SRC_{i1} \cdot STime_{\min}, \\ & SRC_{i2} \cdot STime_{\min}, \ldots, \\ & SRC_{ij} \cdot STime_{\min}\} \end{aligned} \tag{2}$$

$$\begin{aligned} CTime_{\max} = {} & SRC_1 \cdot STime_{\max} \\ & + SRC_2 \cdot STime_{\max} + \cdots \\ & + SRC_{i1,i2\ldots,ij} \cdot STime_{\max} + \cdots \\ & + SRC_n \cdot STime_{\max} \end{aligned} \tag{3}$$

$$\begin{aligned} SRC_{i1,i2\ldots,ij} \cdot STime_{\max} = {} & \max \{SRC_{i1} \cdot STime_{\max}, \\ & SRC_{i2} \cdot STime_{\max}, \ldots, \\ & SRC_{ij} \cdot STime_{\max}\}. \end{aligned} \tag{4}$$

(5) *CRC*StatusInfo is the status set of CRC, and its formalized representation is: *CRC*StatusInfo = {0, 1, 2, 3, 4, 5}. 0 represents that the resource is maintenance; 1 is idle; 2 is under load; 3 is full load; 4 is over load; 5 is invalid. Here, we transform SRCSet into SRCSetX, SRCSetX = $\{X, SRC_1, SRC_2, \ldots, SRC_{i1,i2}, \ldots, SRC_{ii,ii+1,\ldots,ij}, \ldots, SRC_x\}$. X is the sum of single resource unit, and $SRC_{i1,i2}$ and $SRC_{ii,ii+1,\ldots,ij}$ both represent a single resource unit composed of several parallel single resources, and the formula of single resource unit resource statues is (5); so the formula of *CRC*StatusInfo is (6):

$$SRC_{ii,ii+1,\ldots,ij} \cdot SRCStatusInfo = \begin{cases} 0, & SRC_j = 0, j = ii, ii+1, \ldots, jj \\ 1, & \exists SRC_j = 1, j = ii, ii+1, \ldots, jj \\ 2, & \exists SRC_j = 2, j = ii, ii+1, \ldots, jj \\ 3, & SRC_j = 3, j = ii, ii+1, \ldots, jj \\ 4, & SRC_j = 4, j = ii, ii+1, \ldots, jj \\ 5, & SRC_j = 5, j = ii, ii+1, \ldots, jj \end{cases} \tag{5}$$

*CRC*StatusInfo

$$= \begin{cases} 0, & \exists SRC_i \cdot SRCStatusInfo = 0, \\ & SRC_i \in SRCSetX \\ 1, & SRC_i = 1, SRC_i \in SRCSetX \\ 2, & SRC_i = 2, SRC_i \in SRCSetX \\ 3, & SRC_i = 3, SRC_i \in SRCSetX \\ 4, & SRC_i = 4, SRC_i \in SRCSetX \\ 5, & \exists SRC_i \cdot SRCStatusInfo = 5, \\ & SRC_i \in SRCSetX. \end{cases} \tag{6}$$

4.2. Cloud Manufacturing Platform Model. The user is the main service object of CMP, so it's needed to establish a service-oriented resource model called resource service attribute model to provide users with information related to resource functionality, usage and service quality, all of which can help users search, select and use resource. The formalized representation of resource model in CMP is Re*sCloudPlatform* = {*ID*, *RCEID*, *BaseInfo*, *ServiceInfo*, *AssessInfo*, *OtherInfo*}.

(1) *ID* is the flag of resource cloud, used to uniquely identify manufacturing resource cloud in the same sub-CMP so as to realize locating and indexing resource. The uniqueness is within the same sub-CMP, while the same *ID* in different sub-CMP may represent different resource cloud, thus avoiding updating the resource list of all the sub-CMPs once a CE publishes or updates resource to a certain sub-CMP. It is more suitable for the massive and complex characteristics of resource cloud data.

```
< ResCloudEnd >
  < ResCloudEnd-info indentifier = "single-resource" >
    < RCEInfo >
      < description name = "RCEID" value = "00888" />
      < description name = "RCEProvider" >
        < description name = "RCEProvider-name" value = "× × × Ltd." />
        < description name = "RCEProvider-tel" value = "86-0571-12345678" />
        ...
      </ description >
      < description name = "RCEAddress" value = "No. 123, Liuhe Road, Zhejiang Province" />
      < description name = "RCEOther" value = null />
    </ RCEInfo >
    < RC >
      < description name = "RCID" value = "CC615" />
      < description name = "SRCBaseInfo" >
        < description name = "SRCName" value = "Single Column Vertical Machine Tool" />
        < description name = "SRCModel" value = "C5116E" />
        < description name = "SRCType" value = "2" />
        < description name = "SRCInfo" value = null />
        < description name = "SRCParameter" >
          < description name = "length" value = "2800" />
          < description name = "Width" value = "3000" />
          < description name = "height" value = "3200" />
          < description name = "power" value = "22/33" />
          ...
        </ description >
        < description name = "SRCBInfos" value = null />
      </ description >
      < description name = "SRCFuncInfo" >
        < description name = "SRCTaskType" value = null />
        < description name = "SRCCapa" >
            ...
            < description name = "workpiece weight" value = "5000 kg" />
            ...
        </ description >
        ...
      </ description >
      < description name = "SRCStatusInfo" value = "2" >
    </ RC >
  </ ResCloudEnd-info >
</ ResCloudEnd >
```

Algorithm 1: WSDL of single resource cloud description in CE.

(2) *RCEID* is the CE flag which provides the resource cloud, used to uniquely identify the provider of resource.

(3) *BaseInfo* is the basic characteristics attributes set of resource cloud, and its formalized representation is *BaseInfo* = {*Name*, *Model*, Re *sInfo*, Pro*vider*}. *Name* is the resource name; *Model* is the resource model; Re *sInfo* is the description information of resource, defined by the provider; Pro*vider* is the description of resource provider, defined by provider-self.

(4) *ServiceInfo* is the service capability information, and its formalized representation is *ServiceInfo* = {*TaskType*, *Capa*, *Time*, *Qua*, *Cost*, *Enviro*, *photo*, *OtherServInfo*}. The elements respectively mean service type, performance indicators, time-consuming, quality standards, price, service environment, other service information, and so forth.

(5) *AssessInfo* is the assessment information of resource, and its formalized representation is AssessInfo = {*TaskID*, *TAss*, *CAss*, *QAss*, *ServAss*, *CreditAss*, *Average*}. *TaskID* is the identification number of a task, automatically generated by sub-CMP; *TAss* is time-consuming evaluation; *CAss* is price evaluation; *QAss* is quality evaluation; *ServAss* is service evaluation; *CreditAss* is credit evaluation; *Average* is average value of every indicator up to now.

(6) *OtherInfo* is other service capability information, added or removed by resource provider, sub-CMPs and users, and its formalized representation is $OtherInfo = \{other \inf o_1, other \inf o_2, \ldots, other \inf o_i, \ldots\}$, $other \inf o_i = \{user_i, \inf o_i\}$, $user_i$ is the provider of other service capability information, $\inf o_i$ is the information content stored as a string.

```
< ResCloudPlatform >
  < ResCloudPlatform-info indentifier = “ResCloudPlatform” >
    < description name = “ID” value = “SB0088” />
    < description name = “RCEID” value = “00888” />
    < description name = “BaseInfo” >
      < description name = “name” value = “Single Column Vertical Machine Tool” />
      < description name = “Model” value = “C5116E” />
      < description name = “ResInfo” value = null />
      < description name = “Provider” value = “× × × Ltd.” />
    </ description >
    < description name = “ServiceInfo” >
      . . .
      < description name = “Price” value = “28 yuan/hour” />
      < description name = “photo” >
        < description name = “photo1” >< a href = CC615_1.jpg > photo1 < /a >< / description >
        < description name = “photo2” >< a href = CC615_2.jpg > photo2 < /a >< / description >
        < description name = “photo3” >< a href = CC615_3.jpg > photo3 < /a >< / description >
        . . .
      </ description >
      . . .
    </ description >
    < description name = “AssessInfo” >
      . . .
      < description name = “IncreditAss-Average” average = “2” >
        < description name = “IncreditAss-001” TaskID = “001” IncreditAss = “1” />
        < description name = “IncreditAss-002” TaskID = “002” IncreditAss = “2” />
        < description name = “IncreditAss-003” TaskID = “003” IncreditAss = “3” />
      </ description >
      . . .
    </ description >
    < description name = “OtherInfo” >
      < description name = “otherinfo1” attribute1 = “John” attribute2 = “value1” />
      < description name = “otherinfo2” attribute2 = “Microsoft” attribute2 = “value2” />
      . . .
    </ description >
  </ ResCloudPlatform-info >
</ ResCloudPlatform >
```

Algorithm 2: WSDL of resource description in CMP.

5. Realization

We build the Cloud Manufacturing Resource Model with stateful Web Service Description Language (WSDL) document, when dynamically publishing and updating resource cloud. Take certain vertical CNC machine tools as an instance, its processing service capability can be simply described as Table 1, the WSDL document of single resource cloud in CE is shown as Algorithm 1, and the WSDL document of resource cloud in CMP is shown as Algorithm 2.

6. Discussion

Cloud Manufacturing is a kind of intelligent networked manufacturing model with the characteristics such as service-oriented, high-efficiency and low-energy, and knowledge-based. Through the integration of some advanced technology [17–19], the manufacturing resources would be virtualization and servicesation, in order to be intelligently, multi-win-win, and efficiently shared and collaborated. In the Cloud Manufacturing environment, enterprises provide convenient, on-demand, safe and reliable, and high-quality and low-cost services. Cloud Manufacturing has the following characteristics: highly fragmented distribution and highly concentrated use of resources, service-oriented and demand-oriented, uncertainty of manufacturing plan, manufacturing with users involved, and use and pay on-demand. Because of all the above, the structure, functionality, operating environment, and basic physical attribute have the following characteristics: diversity, complexity, heterogeneity, being massive, and so forth. The bilayer model of manufacturing resource proposed in this paper can adapt well to such a manufacturing environment. Firstly, the separation of the basic data model and the service attribute model, between which a one-to-one mapping is formed by keywords, enables the bilayer model to adapt to the complex and heterogeneous resource in Cloud Manufacturing. Secondly, the two types of resource

Table 1: Parameters of vertical CNC machine tools.

Attribute	Value	Attribute	Value
ID	CC615	Crosshead trip	760 mm
Name	C5116E Machine	Turret fast moving speed	1800 mm/min
Introduction	.....	Vertical turret slewing angle	$30\pm^{o}$
Enterprise	.....Ltd.	Vertical turret maximum cutting force	25 KN
Address	No. 123 Liuhe Road, Hangzhou, Zhejiang Province	Side turret maximum cutting force	20 KN
Price	28 yuan/hour	Workbench maximum torque	25 KNM
Maximum turning diameter	1600 mm	Tool bar section height	40 mm
Table diameter	1400 mm	Main motor power	22/30 KW
Maximum processing height	1000 mm	Overall dimensions (length × width × height)	$2800 \times 3000 \times 3200\ mm^3$
Maximum workpiece weight	5000 kg	Machine weight	12580 kg
Table speed range (16)	5–160 r/min	Photo gallery	CC615_1.jpg CC615_2.jpg CC615_3.jpg
Turret feed range (12)	0.8–86 mm	Years in use	N years
Vertical turret ram stroke	800 mm	Assess	{001, 1, 672, 4, 1, "comment001"}{002, 1, 336, 5, 2, "commen002"}{003, 0.5, 163, 3, 3, "commen003"} ..... {000, 0.83, 4, 2,}
Side head ram level trip	630 mm	Status	{0, 1, 2, 3, 4, 5}
Side head vertical trip	980 mm	.....	.....
Side head maximum turning diameter	1400 mm	.....	.....

cloud (single resource cloud and complex resource cloud) can describe well the resource basic characteristics with the basic data model, which usually has a huge amount of data and is relatively certainty stored in CE, and thus greatly reduces heavy tasks for data storage, maintenance, and updating in CMP. thirdly, during dynamically publishing and updating resource cloud, we adopt the stateful Web Service Description Language (WSDL) document, which has a high level of flexibility, and thus is suitable for massive data easily accessed by CE in the Cloud Manufacturing environment. Finally, fully taking into account the user participation, user evaluation is included in the resource model, thus eases to updating and improving the manufacturing resources, and fully reflects the Cloud Manufacturing characteristics of user-oriented and service-oriented.

7. Conclusion

With the increasingly fierce competition in the global market and increasingly serious energy and environmental issues, how to provide users with high quality products and services with low energy consumption and environmental friendly, is an urgent problem to manufacturing industry currently [8]. Cloud Manufacturing and Cloud Service is one of the main directions of the current development in the manufacturing industry and results by applying Cloud Computing theory to manufacturing industry, so it has universal characteristics of "Cloud Theory" such as high user participation, high resource sharing, and high process agility. Aiming at making manufacturing agility, servicesation, green, and intelligence, Cloud Manufacturing is a new development of networked manufacturing with the service-oriented manufacturing theory called Manufacturing as a Service (MaaS) [20]. Under Cloud Manufacturing environment, the publishing, updating, searching, and accessing manufacturing resources have the characteristics of being massive, complexity, and heterogeneity, thus putting forward higher flexibility, agility, and versatility on the resource modeling. In this paper, a new bilayer manufacturing resource model with separation of CE and CMP is proposed. The model is comprised of resource basic data model and resource service attribute model: the resource basic data model in CE focuses on the physical characteristics of manufacturing resource, and it faces enterprise interior manufacturing management system. And the two type resource models (single resource and complex resource) can describe well the resource basic characteristics, which usually have a huge amount of data stored relatively certainly in CE; the resource service attribute model, which usually has smaller amount of data stored in CMP, focuses on service characteristics of manufacturing resource, and it faces resource users' actual demands. And fully taking into account the user participation, user evaluation is included in the resource model. Besides, the stateful WSDL documents can describe resource attributes perfectly and flexibly

thus are suitable to the need under Cloud Manufacturing environment. Through realization of the instance, the model can run well in Cloud Manufacturing environment and can store, publish, and update large and complex data. The users' demands on discovering and selecting manufacturing resource and deeply participating in manufacturing process can also be met.

Acknowledgments

The presented research was supported by the National Natural Science Foundation, China (no. 60970021), and the Scientific Research Plan of Zhejiang Provincial Department of Education, China (no. Y201120777).

References

[1] T. S. Baines, H. W. Lightfoot, S. Evans et al., "State-of-the-art in product-service systems," *Proceedings of the Institution of Mechanical Engineers Part B*, vol. 221, no. 10, pp. 1543–1552, 2007.

[2] B. H. Li, L. Zhang, S. L. Wang, F. Tao, and J. W. Cao, "Cloud manufacturing: a new service-oriented networked manufacturing model," *Computer Integrated Manufacturing Systems*, vol. 16, no. 1, pp. 1–7, 2011.

[3] B. H. Li, L. Zhang, L. Ren et al., "Further discussion on cloud manufacturing," *Computer Integrated Manufacturing Systems*, vol. 17, no. 3, pp. 449–457, 2011.

[4] H. W. Yang, *Cloud Manufacturing: A Manufacturing Service*, Manufacture Information Engineering of China, 2010.

[5] B. H. Li, L. Zhang, and X. D. Chai, "Introduction to cloud manufacturing," *ZTE Technology Journal*, vol. 16, no. 4, pp. 5–8, 2010.

[6] G. Breiter and M. Behrendt, "Life cycle and characteristics of services in the world of cloud computing," *IBM Journal of Research and Development*, vol. 53, no. 4, pp. 3:1–3:8, 2009, paper no. 3.

[7] K. Chen and W. M. Zheng, "Cloud computing: system instances and current research," *Journal of Software*, vol. 20, no. 5, pp. 1337–1348, 2009.

[8] F. Tao, L. Zhang, H. Guo, Y. L. Luo, and L. Ren, "Typical characteristics of cloud manufacturing and several key issues of cloud service composition," *Computer Integrated Manufacturing Systems*, vol. 17, no. 3, pp. 477–486, 2011.

[9] M. Lu, X. D. Sun, and G. Wang, "Resource modeling to support manufacturing process optimization under uncertain environment," *Computer Integrated Manufacturing Systems*, vol. 16, no. 12, pp. 2611–2616, 2010.

[10] B. Sheng, Y. Li, Y. Ding, F. Tao, and Z. Zhou, "Modeling and management of manufacturing resource information in manufacturing grid," *China Mechanical Engineering*, vol. 17, no. 13, pp. 1375–1380, 2006.

[11] Y. D. Fang, L. H. Du, W. P. He, H. Chen, and B. Sun, "Multi-dimension manufacturing resource modeling technology research based on networked manufacturing environment," *Application Research of Computers*, vol. 26, no. 2, pp. 559–562, 2009.

[12] L. B. Zheng, J. N. Gu, and Y. R. Dai, "Modeling of manufacturing resources based on ontology," *Machine Design and Research*, vol. 25, no. 5, pp. 61–63, 2009.

[13] F. Liang, Z. B. Jiang, and L. Y. Tao, "Model building approach of manufacturing resource based on meta-resource," *Computer Integrated Manufacturing Systems*, vol. 14, no. 12, pp. 2306–2311, 2008.

[14] C. F. Yao, D. H. Zhang, K. Bu, W. H. Wang, and J. X. Ren, "Networked manufacturing resources modeling and information integrationbased on physical manufacturing unit," *Computer Integrated Manufacturing Systems*, vol. 14, no. 4, pp. 667–674, 2008.

[15] G. H. Zhou and P. Y. Jiang, "Encapsulation and integration for networked manufacturing resources based on mobile agents," *Computer Integrated Manufacturing Systems*, vol. 8, no. 9, pp. 728–732, 2002.

[16] S. Y. Shi, H. B. Liu, H. C. Yang, R. Mo, and Z. F. Chen, "Resource modeling based on web service in manufacturing grid," *Aeronautical Manufacturing Technology*, vol. 12, pp. 80–83, 2008.

[17] S. Y. Chen, Y. F. Li, and N. M. Kwok, "Active vision in robotic systems: a survey of recent developments," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1343–1377, 2011.

[18] S. C. Lim, C. H. Eab, K. H. Mak, M. Li, and S. Y. Chen, "Solving linear coupled fractional differential equations by direct operational method and some applications," *Mathematical Problems in Engineering*, vol. 2012, Article ID 653939, 28 pages, 2012.

[19] W. Huang and S. Y. Chen, "Epidemic metapopulation model with traffic routing in scale-free networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, Article ID P12004, 19 pages, 2011.

[20] L. Zhang, Y. L. Luo, W. H. Fan, F. Tao, and L. Ren, "Analyses of cloud manufacturing and related advanced manufacturing models," *Computer Integrated Manufacturing Systems*, vol. 17, no. 3, pp. 458–468, 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 768658, 7 pages
http://dx.doi.org/10.1155/2013/768658

Research Article

Energy Efficiency of a Greenhouse for the Conservation of Forestry Biodiversity

Alvaro Marucci,[1] Maurizio Carlini,[1] Sonia Castellucci,[2] and Andrea Cappuccini[1]

[1] *DAFNE, University of Tuscia, Via San Camillo de Lellis s.n.c., 01100 Viterbo, Italy*
[2] *CIRDER, University of Tuscia, Via Cavour 23, 01028 Orte, Italy*

Correspondence should be addressed to Alvaro Marucci; marucci@unitus.it

Received 21 August 2012; Accepted 1 November 2012

Academic Editor: Carlo Cattani

Forest biodiversity conservation is one of the most interesting and crucial problems in forestry world. Currently, the conservation methods are based on two phases: the conservation of seeds at low temperatures and the multiplication of vegetable material. This latter operation can be successfully developed in properly designed greenhouses. The aim of this paper is to define a type of greenhouse which is particularly suitable for plant material propagation in order to preserve forest biodiversity in the area of the Central Italy. Some general parameters were first defined for a correct planning of the structure, such as: the shape of the section, volume, cover material, systems for heating and cooling, and those for the control of the internal microclimate parameters (light, air temperature, and relative humidity). Considering the construction characteristics and the climatic conditions of the place, the internal microclimatic conditions have been later determined by the useful implementation in TRNSYS in order to analyse the energy efficiency of the greenhouse.

1. Introduction

Greenhouses crops in Italy are made by using prefabricated structures, leaving out the preliminary study of optical and thermal exchanges between the external environment and the greenhouse, dealing with heating and cooling and the effects of air conditioning needed for plant growth [1].

The design and implementation of greenhouses is done without the study of energy exchanges between the external environment inside the greenhouse. Climatic conditions are evaluated only on the plant species planted without verifying the construction parameters of the greenhouse or efficiency of air conditioning systems [2].

As a result of human actions many forest species have become extinct and many more are at risk of extinction. To oppose this loss of forest biodiversity one of the operations most widely used—especially in those areas where the climatic conditions are not optimal—is the multiplication of plant material.

Conservation strategies currently used are in situ and ex situ [3].

The in situ technique requires that the species are preserved in their natural environment, encouraging natural regeneration of wild species or acting through artificial regeneration from seed or transplant in the same area where the seed was taken [3, 4].

On the contrary, ex situ conservation takes place outside the natural habitat and is practiced where risk of species' extinction occurs because of events difficult to control, such as climate change or natural disasters [5]. This practice necessarily requires human intervention for collection, storage, and reproduction of different vegetative materials.

The production process of plants is an important step for the conservation of forest species; the seeds are put down in the most suitable conditions in order to promote germination and to protect young plants against biotic and abiotic agents. These conditions can be obtained in greenhouses where it is possible to control solar radiation, temperature, and relative

humidity of the air and to adapt these parameters to the specific needs of various plant species [6, 7].

Greenhouse cover, structural elements, soil, and crop contribute to the solar radiation absorption, and consequently, to the generation of sensible or latent heat [8].

Inside a greenhouse are many interactions that can occur between crop species, climate conditions, and design characteristics of the greenhouse itself. During the design phase, it is necessary to verify the climatic conditions that are found inside the greenhouse as a function of climate zone in which they are installed. This verification is required to ensure a correct design of the greenhouse and also a reduction in energy consumption.

Approximately 20–30% of the greenhouse on the Italian territory are equipped with heating and cooling. Has been calculated that the energy consumption on the conditioning of greenhouses is about 140,000 TOE (tons of oil equivalent), approximately 90–95% of global energy demand for the production [9].

From these data it is essential to a detailed study of the energy consumption of the greenhouse. This can be done through the use of TRNSYS 17 dynamic simulator that allows you to verify the energy consumption related to the cultivation of a particular plant species and therefore allows you to change the design of the greenhouse.

The aim of this paper is to identify the most suitable type of greenhouse for the multiplication of plant material in order to preserve forest biodiversity in relation to the specific needs of forest species and climatic conditions in Central Italy. In particular, in this study we created a greenhouse model which allows to check, with TRNSYS, correct climatic conditions inside and the energy consumption related to the cultivation of forest species.

2. Materials and Methods

The study was carried out with considering the type "multispan greenhouse" and located in Central Italy.

After the definition of construction parameters, equipments and climatic conditions of the site, microclimatic conditions inside the greenhouse were determined by simulation software TRNSYS 17. As regards the determination of the thermal needs of each species forestry, the phytoclimatic classification of Pavari was considered: according to it, the Italian territory is divided into five climatic zones, characterized by a Latin name—*Lauretum, Castanetum, Fagetum, Picetum* e *Alpinetum*—and by an altitudinal and thermal interval. The Latin name refers to the cenosis or species, wild or cultivated by man, which characterizes that specific area [10].

2.1. Greenhouse Structural Description. Considering the large size of forest species, the type "multispan greenhouse"—with high height to eaves—was chosen (Figures 1 and 2).

The main dimensional characteristics are the following: length 100 m, width 12.25 m, height to eaves 4.5 m, ridge height 6 m, slope of the roof pitches 22° (slope = 40%), and orientation with the long axis north-south. The cover material

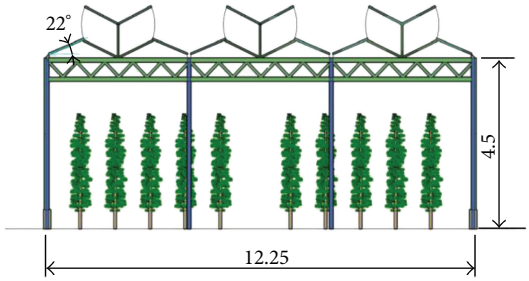


Figure 1: Steel and glass greenhouse: SECTION.

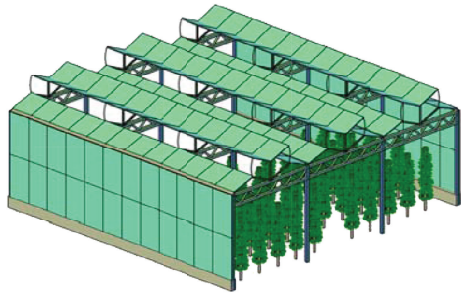


Figure 2: Steel and glass greenhouse: AXONOMETRIC VIEW.

is glass since it is the best material for the greenhouse effect and is strong and durable over time.

In order to allow the multiplication of forest species with very different microclimatic needs, particular attention has been paid to the monitoring and control of the inside microclimate.

Greenhouse is equipped with a system of artificial heating to provide energy in winter. Nevertheless, it is more important, in Central Italy, to contain the excess energy in summer because most forest species are sensitive to high temperatures.

Cooling is ensured by natural and mechanical ventilation. The greenhouse is equipped with large openings on the slopes and on the longitudinal walls and axial ventilators placed at the head of it.

The greenhouse is equipped by a system for automatic regulation and control whose aim is to ensure a correct management of the microclimatic conditions. More precisely, it consists of a series of sensors that detect microclimatic internal and external parameters, and of a control unit that processes the data and transmits the changes to the control system.

2.2. Greenhouse Model. The thermal behavior of the greenhouses was studied using dynamic thermal simulation tool, TRNSYS 17. A model, which includes all greenhouse influential variables on microclimate, has been created.

This model includes

(i) detailed design of structures,

(ii) optimizing resources, and

(iii) verification of climatic parameters in greenhouse.

The simulation of climate parameters of the greenhouse is carried out with a model that calculates the thermal exchanges in nonstationary conditions [11]. Depending on the parameter, one can highlight energy waste and improve the design of the greenhouse. The created model can be applied in any type of greenhouse changing design parameters in the TYPE 56.

The model was built using the simulation program TRNSYS Studio. The first step concerns the construction of a building multizone, which is divided into multiple steps in which the user enters the data relating to the building and its location in space [12]. TRNSYS provides a daily evaluation on the air conditioning system in the greenhouse in response to the climatic conditions of the area in which greenhouse is located.

The climatic parameters affecting the growth of greenhouse plants are [13]

(i) radiation entering the greenhouse in the ultraviolet band (290 to 380 nm), visible (380 to 760 nm) and near infrared (760–3000 nm),

(ii) air, soil and plants temperature,

(iii) air humidity, and

(iv) air composition (particularly the concentration of CO_2).

Solar radiation inside the greenhouse is less than the outer coating material which reflects and absorbs solar radiation; one also needs a ventilation system that ensures the air movement between the plants and facilitates the transpiration; the lower is the air movement, and the smaller is the concentration gradient of CO_2.

It is necessary to assess the detailed exchange of energy between internal and external part to have accurate values of climatic parameters in the greenhouse and to study the temperatures that influence the thermal freight exchange [14–17]:

(i) air temperature,

(ii) soil temperature (or substrate for soilless),

(iii) plants temperature, and

(iv) temperature of the greenhouse cover.

In order to calculate the heat balance in emissions, it is fundamental to examine the individual energy input (Figure 3).

Energy balance equation:

$$R_{\mathrm{i}} + R_{\mathrm{ac}} = RT_{\mathrm{g,sky}} + RT_{\mathrm{g,atm}} + RT_{\mathrm{p,sky}} + RT_{\mathrm{p,atm}} + RT_{\mathrm{c,sky}} + RT_{\mathrm{c,atm}} + T_{\mathrm{c}} + V + \Delta E_{\mathrm{g}} + S_{\mathrm{g}}, \tag{1}$$

where

$$T = KS\left(T_{\mathrm{i}} - T_{\mathrm{e}}\right), \tag{2}$$

K is thermal transmittance ($\mathrm{W\,m^{-2}\,K^{-1}}$); S is wall surface ($\mathrm{m^2}$); T_{i} is internal temperature (°C); and T_{e} is external temperature (°C).

$$E_v = V\left(H_{\mathrm{i}} - H_{\mathrm{e}}\right), \tag{3}$$

E_v = energy lost for ventilation ($\mathrm{KJ\,h^{-1}}$); V = flow rate of ventilation ($\mathrm{Kg\,DryAir\,h^{-1}}$); H_{i} = internal air enthalpy-($\mathrm{KJ\,Kg\,DryAir^{-1}}$); H_{e} = external air enthalpy ($\mathrm{KJ\,Kg\,Dry\text{-}Air^{-1}}$).

The enthalpy H ($\mathrm{KJ\,Kg\,DryAir^{-1}}$) of a Kg of air, at temperature t (°C) and with a water content, in the form of water, equal to x ($\mathrm{Kg\,Kg\,DryAir^{-1}}$) is obtained from

$$H = 1.005t + x\left(2499.5 + 2.005t\right), \tag{4}$$

x ($\mathrm{Kg\,Kg\,DryAir^{-1}}$) represents the water vapor contained in the air and can be determined through the psychrometric diagram or with the following formula:

$$x = 0.6215\frac{P_{\mathrm{H_2O}}}{P - P_{\mathrm{H_2O}}}, \tag{5}$$

P = 1.013 ($\mathrm{Kg\,cm^{-2}}$) atmospheric pressure; $P_{\mathrm{H_2O}}$ = vapor pressure of the air, expressed in $\mathrm{Kg\,cm^{-2}}$ at the temperature T (K) and at relative humidity UR (%);

$$P_{\mathrm{H_2O}} = 1.41 \times 10^{10} e^{(-3928.5/(T-41.5))} \times \mathrm{UR,\,(Pa)}, \tag{6}$$

$$RT_{12} = \sum_{\lambda=2.5\mu\mathrm{m}}^{\lambda=20\mu\mathrm{m}} \varepsilon_{12}(\lambda) \times \tau_a(\lambda) \times \frac{C_1}{\lambda^5} \times \left[\frac{1}{e^{C_2/\lambda\times T_1} - 1} - \frac{1}{e^{C_2/\lambda\times T_2} - 1}\right] \times \Delta\lambda \tag{7}$$

RT_{12} = radiated radiation ($\mathrm{W\,m^{-2}}$); λ = wavelength (μm); T = absolute temperature (K); $C_1 = 3.74 \cdot 10^8$ ($\mathrm{W}\,\mu\mathrm{m}^4\ \mathrm{m}^{-2}$); C_2 = 14385 (K μm); $\tau_a(\lambda)$ = transmittance of the atmosphere at the wavelength λ; $\varepsilon_{(\lambda)}$ = monochromatic emissivity of the object at the wavelength λ.

The TRNSYS is commonly used to simulate transient heat transfer for design and control of power systems exploiting renewable energy sources. Another frequent use of software is now on the energy certification for homes, offices, shops, restaurants, and industries. In this sense the present work is an example of using the software for agricultural systems [18]. The structure is a multispan greenhouse with high height to eaves, used as a greenhouse for growing flowers and plants. It is covered with glass cover horizontal beam pattern and small flat foot north-south. The approach of the model was carried out by using the program TRNSYS Simulation Studio. The first step concerns the construction of a building multizone, which is divided into multiple steps in which the user enters the data relating to the building and its location in space [12]. The data required by the software at this stage will be used for the automatic construction of the project and its connections between the components.

During construction of the project there is also the source of meteorological data that will be used in the simulation. This is indeed a link with the Type 109 (Weather Data Processor) and in the present study, the meteorological station of Rome Fiumicino (Airport) was selected. The outline of the project is in (Figure 4), where

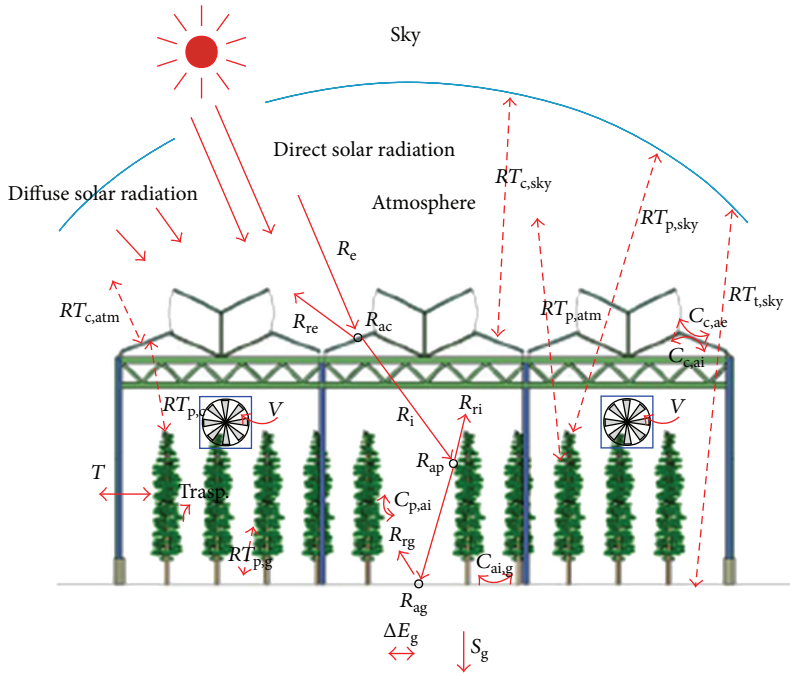


Figure 3: Heat balance in a greenhouse.

weather date: weather generator

psychrometrics: processor psychrometric

sky Temp: CPU temperature sky

radiation: radiation converter

greenhouse: Greenhouse (Type 56)

nat. Vent. 1 and 2: controllers of ventilation

turn: unit converters

charts: Online Plotter.

3. Results

3.1. Simulation Results. The purpose of the simulation is to check the temperatures inside the greenhouse related to external climatic parameters and the type of construction of the greenhouse itself. The required energy for optimal climate conditioning is based on such calculation. These values in fact are calculated in order to maintain a steady temperature of 20 degrees which is considered suitable for the cultivation of a wide range of plant species. The results are shown in the chart (Figures 5, 6, 7, 8 and 9):

The solar radiation arrives in the greenhouse and affects the inside temperature. In particular the characteristic constructive of the greenhouse is 97% of the radiation incident, reaches the ground, and increases on the internal temperature (Figure 7).

In Figure 6 the solar radiation inside the greenhouse is compared with the incident one. It can be seen that the low resistance of the structure results to be an advantage in winter, even if energy consumption for cooling is higher.

During cold months the internal radiation is almost equal to the outside, while in summer transmission of solar radiation falls on the most important reflection of the glass. This is due to the increased presence of direct radiation and different angles of incidence.

During the winter months the trend of the indoor temperature is almost identical to that of the external solar radiation and the structure has a low thermal capacity (infiltration losses of 1/h) (Figure 9), where

$T_{\text{Greenhouse}}$: inside room temperature

Q_{Heat}: required heating, the rate at which energy must be supplied to the device in order to heat the air to its outlet temperature, including the effects of losses and conversion inefficiencies.

Q_{Cool}: rate of energy removed with the cooling system.

Moreover, the amount of heat to be provided or removed by heating and cooling was connected to the printer to obtain the annual energy needs for heating and cooling:

$$Q_{\text{Heat}} = 1,44E + 23\,\text{kJ h}^{-1},$$
$$Q_{\text{Cool}} = 2,25E + 23\,\text{kJ h}^{-1}. \quad (8)$$

It can be seen that the energy demand for cooling is higher than that for heating [14]. This is due to the latitude where the greenhouse is located, characterized by very rigid winters and hot summers. The technology normally used in greenhouses for cooling, as shown in the chart above, is from an energy point of view very expensive. The graphic represents a

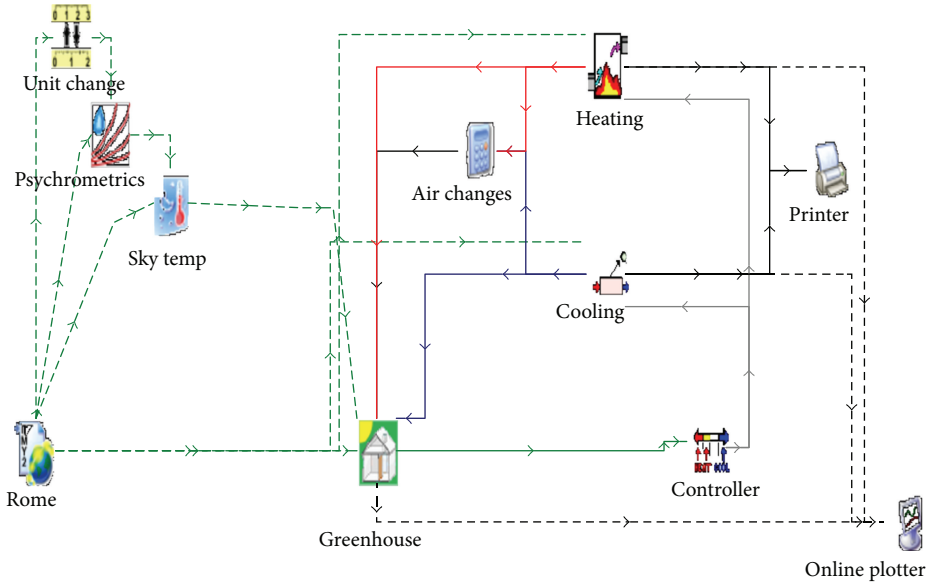


Figure 4: Air conditioning system diagram with TRNSYS.

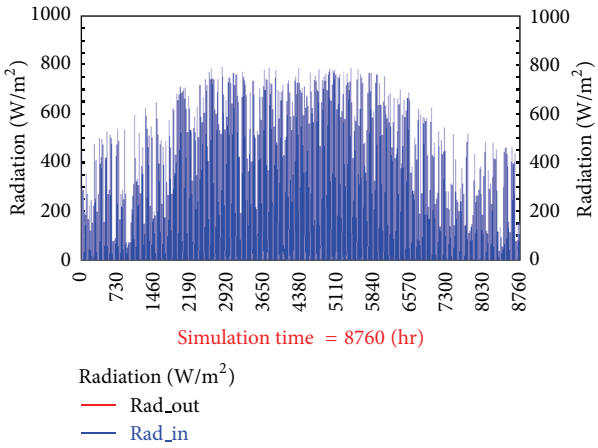


Figure 5: Simulation of annual solar radiation inside the greenhouse (W/m^2).

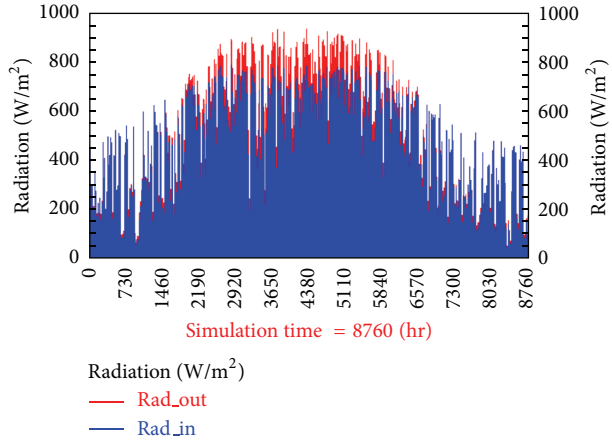


Figure 6: Simulation of annual solar radiation inside the greenhouse (W/m^2) in blue and external red.

constant value of Q_{Cool} for the period July-August, even if the temperature inside the greenhouse undergoes changes.

An analysis of the chart is highlighting the need to rationalize the climate inside the greenhouse. This can be achieved

(i) either through the use of a more precise control systems, such as PID;

(ii) through more sustainable air conditioning systems. The latter using heating and cooling floor system powered by a geothermal heat pump or conventional pump;

(iii) or improving the greenhouse insulation and ventilation systems.

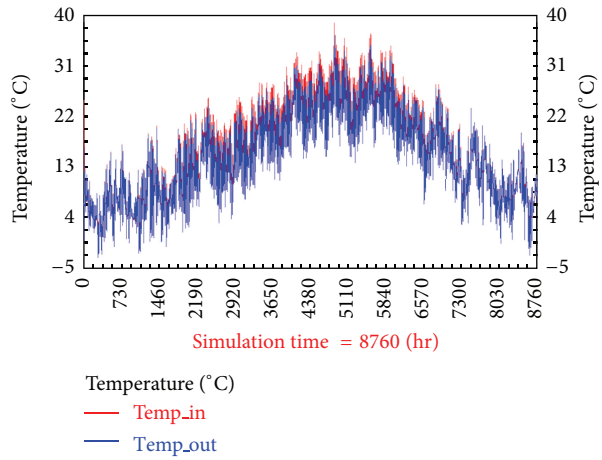


Figure 7: Simulation of annual temperatures.

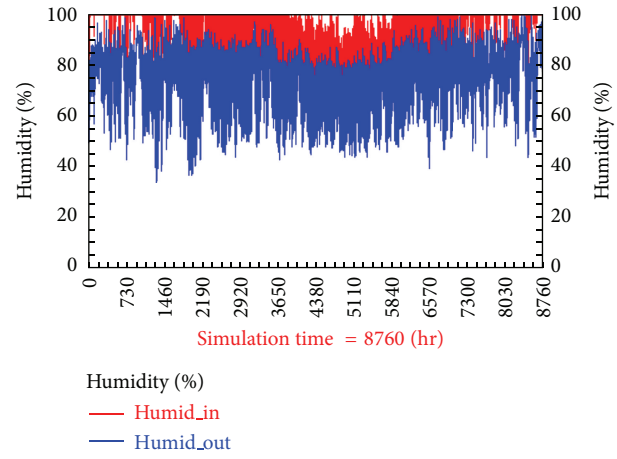


Figure 8: Simulation of annual relative humidity (%).

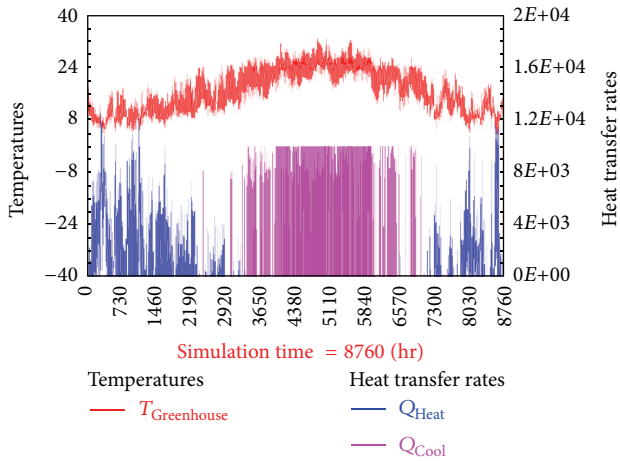


Figure 9: Chart of greenhouse climate simulation.

4. Conclusions

In this research the most suitable type of greenhouse was identified in order to multiply forest plants for biodiversity conservation. It is a multispan greenhouse covered with glass and high height to the eaves. The Energy balance, between the input and output of energy, was carried out for this specific type of greenhouse. This goal was reached by the simulation in TRNSYS and with specific regard to locations in Central Italy. The results obtained in the simulation show the following values for the principal microclimatic parameters:

(i) solar radiation: 300–600 W/m^2 in winter months and 600–800 W/m^2 in summer months;

(ii) air temperature: 10–20°C in winter months and 20–34°C in summer months;

(iii) relative humidity: 85–95% in winter months and 75–85% in summer months.

Furthermore, energy needs in winter (heating) and in summer (cooling) were calculated in order to reduce seasonal fluctuations in the air temperature.

These values are compatible with the needs of forest plants included in the range of phytoclimatic *Lauretum* classification of Pavari.

It has to be underlined that the control systems for heating and cooling should be used—to a greater extent—for other forest species.

The greenhouse model created with the TRNSYS allows use for various applications, that is small changes of parameters allow you to use it for the simulation of different tree species and climatic zones. Turns out to be an important tool for the design of greenhouses. The model created ensures, in fact, a sustainable building, from the point of view of energy, while minimizing energy consumption.

Thanks to a prior verification of heat exchanges achievable between the inside and the outside of the greenhouse is possible to intervene in the construction phase through the modification of the constructive aspects.

Nomenclature

Alphabetic Symbols

R: Global solar radiation [MJ m^{-2} d^{-1}]
RT: Energy lost by radiation [MJ m^{-2} d^{-1}]
C: Energy exchanged by convection [MJ m^{-2} d^{-1}]
T: Energy lost by transmission [MJ m^{-2} d^{-1}]
V: Energy lost by ventilation [MJ m^{-2} d^{-1}]
TR: Plant transpiration [MJ m^{-2} d^{-1}]
S: Energy transferred by conduction in the deep layers [MJ m^{-2} d^{-1}]
E: Thermal storage [MJ m^{-2} d^{-1}].

Subscripts

e: External
re: Reflected external
ac: Absorbed by the cover
i: Internal
ri: Reflected internal
ap: Absorbed by the plants
rg: Reflected from the ground
ag: Absorbed by the ground
c: Cover film
p: Plants
G: Ground
ia: Indoor air
oa: Outdoor air
atm: Atmosphere
sky: Vault of heaven.

References

[1] M. Carlini and S. Castellucci, "Modelling and simulation for energy production parametric dependence in greenhouses," *Mathematical Problems in Engineering*, vol. 2010, Article ID 590943, 28 pages, 2010.

[2] M. Carlini, S. Castellucci, M. Guerrieri, and T. Honorati, "Stability and control for energy production parametric dependence," *Mathematical Problems in Engineering*, vol. 2010, Article ID 842380, 21 pages, 2010.

[3] B. Piotto, V. Giancanelli, and S. Ercoles, *La conservazione ex situ della biodiversità delle specie vegetali spontanee e coltivate in Italia. Stato dell'arte, criticità e azioni da compiere*, Istituto Superiore per la Protezione e la Ricerca (ISPRA '10), Rome, Italy, 2010.

[4] L. Thomson, L. Graudal, and E. Kjær, "Conservation of genetic resources in their natural environment: in managed natural forests and protected areas (in situ)," in *Forest Genetic Resources Conservation and Management*, FAO, DFSC, and IPGRI, Eds., vol. 2, pp. 1–3, International Plant Genetic Resources Institute, Rome, Italy, 2001.

[5] W. Amaral and A. Yanchuk, "Integrated approaches for ex situ conservation and use of forest genetic diversity: in plantations and genebanks (ex situ)," in *Forest Genetic Resources Conservation and Management*, FAO, DFSC, and IPGRI, Eds., vol. 3, pp. 1–3, International Plant Genetic Resources Institute, Rome, Italy, 2004.

[6] A. Marucci and A. Cappuccini, "The characteristics of the greenhouses for the conservation of forest biodiversity in the Mediterranean environment," in *Proceedings of the International Conference on "The Role of Biobanks for Research and Protection of Forest Biodiversity" (MFDB '12)*, University of Tuscia, Viterbo, Italy, April 2012.

[7] P. Piussi, *Selvicoltura Generale*, UTET, 1994.

[8] A. Mistriotis, C. Arcidiacono, P. Picuno, G. P. A. Bot, and G. Scarascia-Mugnozza, "Computational analysis of ventilation in greenhouses at zero- and low-wind-speeds," *Agricultural and Forest Meteorology*, vol. 88, no. 1–4, pp. 121–135, 1997.

[9] M. Carlini, T. Honorati, and S. Castellucci, "Photovoltaic greenhouses: comparison of optical and thermal behaviour for energy savings," *Mathematical Problems in Engineering*, vol. 2012, Article ID 743764, 10 pages, 2012.

[10] Y. Rouphael, M. T. Cardarelli, N. Ajouz, A. Marucci, and G. Colla, "Estimation of eggplant leaf number using thermal time model," *Journal of Food, Agriculture and Environment*, vol. 8, no. 2, pp. 847–850, 2010.

[11] M. Carlini, M. Villarini, S. Esposto, and M. Bernardi, "Performance analysis of greenhouses with integrated photovoltaic modules," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '10)*, vol. 6017 of *Lecture Notes in Computer Science*, pp. 206–214, 2010.

[12] S. Holst, "Heating load of a building model in TRNSYS with different heating systems," ZAE Bayern, Abt. 4, TRNSYS-User Day, Stuttgart, Germany, 1993.

[13] J. E. Braun and J. C. Mitchell, "Solar geometry for fixed and tracking surfaces," *Solar Energy*, vol. 31, no. 5, pp. 439–444, 1983.

[14] D. T. Reindl, W. A. Beckman, and J. A. Duffie, "Diffuse fraction correlations," *Solar Energy*, vol. 45, no. 1, pp. 1–7, 1990.

[15] A. Marucci and B. Pagniello, "Simulation of the growth and the production of the tomato in typical greenhouses of the Mediterranean environment," *Journal of Food, Agriculture and Environment*, vol. 9, no. 3-4, pp. 407–411, 2011.

[16] E. Campiglia, G. Colla, R. Mancinelli, Y. Rouphael, and A. Marucci, "Energy balance of intensive vegetable cropping systems in central Italy," *Acta Horticulturae*, vol. 747, pp. 185–191, 2007.

[17] A. Marucci, E. Campiglia, G. Colla, and B. Pagniello, "Environmental impact of fertilization and pesticide application in vegetable cropping systems under greenhouse and open field conditions," *Journal of Food, Agriculture and Environment*, vol. 9, no. 3-4, pp. 840–846, 2011.

[18] H. Laukamp, "Inverter for photovoltaic systems (in German). User-written TRNSYS source code," Fraunhofer-Institute für Solare Energiesysteme, Freiburg im Breisgau, Germany, 1988.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 249847, 8 pages
http://dx.doi.org/10.1155/2013/249847

Research Article

Differential and Statistical Approach to Partial Model Matching

Kehua Guo, Yongling Liu, and Guihua Duan

School of Information Science and Engineering, Central South University, Changsha 410083, China

Correspondence should be addressed to Kehua Guo; guokehua@csu.edu.cn

Received 2 December 2012; Accepted 23 December 2012

Academic Editor: Sheng-Yong Chen

Partial model matching approaches are important to target recognition. In this paper, aiming at a 3D model, a novel solution utilizing Gaussian curvature and mean curvature to represent the inherent structure of a spatial shape is proposed. Firstly, a Point-Pair Set is constructed by means of filtrating points with a similar inherent characteristic in the partial surface. Secondly, a Triangle-Pair Set is demonstrated after locating the spatial model by asymmetry triangle skeleton. Finally, after searching similar triangles in a Point-Pair Set, optimal transformation is obtained by computing the scoring function in a Triangle-Pair Set, and optimal matching is determined. Experiments show that this algorithm is suitable for partial model matching. Encouraging matching efficiency, speed, and running time complexity to irregular models are indicated in the study.

1. Introduction

Partial model matching has important applications in many cases such as the vision registration for automatic 3D models [1–3], content-based image retrieval from image database [4], assembly and restoration for fractured object [5], docking of proteins in molecular biology [6], and other applications. During the past decade, the problem of finding a partial match between models attracted considerable attention. In many cases, models may be only partially visible because the occlusion or the sensor usually cannot scan all the data, so the goal of partial model matching is to find a transformation of the sample model to overlap a large portion of the object model.

This paper focuses on the research of a 3D partial model matching. In 3D space, the models can be represented as 3D surfaces. Therefore, finding a Euclidean geometry rigid transformation of the sample surface to overlap a large portion of the object surface is the goal of 3D partial model matching. In recent years, there has been lots of research on 3D partial surface matching. The literature [7, 8] proposed a scoring schema for 3D partial surface matching, but this paper did not demonstrate a universal and high-efficiency algorithm for the constructing of a scoring factor. The literature [9] proposed an approach based on the Hausdorff and Frehet distance for geometry structure matching. This approach can be well applied in the matching of points and line segments set, but it is unstable and cannot perform well in the presence of noise, occlusion, and clutter. The literature [10] proposed a scoring schema for matching algorithm to a 3D incomplete object. The literature [11–13] proposed some matching algorithm using differential geometry to 3D partial surface matching. In addition, statistical descriptors [14], probabilistic framework [15], and curve analysis [16] were also applied to the recognition of 3D partial surface; these approaches indicated encouraging matching results for 3D points, line segments set, and mesh model; however, they would suffer from high computing complexity. It is worth mentioning that partial model matching has been successfully applied in video processing [17, 18].

Our work is motivated by the technique of scoring schema and statistical approaches. This technique, which uses the so-called scoring-function method, was originally introduced for 3D partial surface and volumetric matching. This idea has been applied in some pattern recognition problem [19] and achieved good results. The work proposed in this paper has been partly published in an international conference [20]. Here is a brief sketch of our algorithm.

(1) Represent the inherent characteristic of points in sample and object surface utilizing differential geometry method.

(2) Classify point pair with a similar inherent characteristic form sample and object surface.

(3) Classify triangle pair with a similar shape feature and inherent characteristic based on point pairs and compute the geometry rigid transformation for every triangle pair.

(4) Use scoring function to measure the goodness of every transformation through applying it to other triangle pairs.

(5) Determine the optimal transformation with score of every transformation.

Experimental results show that our technique is most successful for the matching of irregular object surface. Encouraging matching efficiency, speed, and running time complexity to irregular models are indicated.

2. Compute and Store the Curvature

2.1. Represent the Spatial Surface Using Gaussian and Mean Curvature. According to differential geometry principle, curvature is the inherent characteristic of a spatial surface. Therefore, Gaussian curvature and mean curvature are employed for the representation of spatial surface in this paper.

In 3D Euclidean space, given a parametric surface defined as

$$S(x, y) = [xyf(x, y)]^T, (x, y) \in D, \tag{1}$$

where X-Y is the reference plane in 3D space, D is projection region of the surface to X-Y plane, and $f(x, y)$ represents the distance from the surface to point (x, y) in X-Y plane.

Gaussian curvature K and mean curvature H can be computed according to the following formulas [18]:

$$K = \frac{f_{xx} f_{yy} - f_{xy}^2}{\left(1 + f_x^2 + f_y^2\right)^2},$$
$$H = \frac{\left(1 + f_x^2\right) f_{yy} + \left(1 + f_y^2\right) f_{xx} - 2 f_x f_y f_{xy}}{2\left(1 + f_x^2 + f_y^2\right)^{3/2}}. \tag{2}$$

For a digital range image surface, approximations can be computed by a local polynomial fitting approach, and $n \times n$ operator is usually utilized to the convolution operation with the original range image:

$$f_x = D_x * f, \quad f_y = D_y * f$$
$$f_{xx} = D_{xx} * f, \quad f_{xy} = D_{xy} * f, \quad f_{yy} = D_{yy} * f, \tag{3}$$

where D is $n \times n$ operator. For $n = 7$, the parameters can be computed as follows:

$$D_x = d_0 d_1^T, \quad D_y = d_1 d_0^T,$$
$$D_{xx} = d_0 d_2^T, \quad D_{yy} = d_2 d_0^T, \quad D_{xy} = d_1 d_1^T$$
$$d_0 = \frac{1}{7}[1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \;\; 1]^T$$
$$d_1 = \frac{1}{28}[-3 \;\; -2 \;\; -1 \;\; 0 \;\; 1 \;\; 2 \;\; 3]^T \tag{4}$$
$$d_2 = \frac{1}{84}[5 \;\; 0 \;\; -3 \;\; -4 \;\; -3 \;\; 0 \;\; 5]^T,$$

where d_0, d_1, and d_2 are column vectors for window operator computing.

2.2. Generate the Point-Pair Set. After presenting the acquisition of Gaussian curvature K and mean curvature H, in our algorithm, point pairs with similar curvature from sample and object surface will be classified, and then the point pairs will be stored into a Point-Pair Set. In order to measure the similarity between two points from sample and object surface, the definition of curvature distance is introduced as follows:

Definition 1. $\forall p_i^1 \in S_1$, $p_j^2 \in S_2$, curvature distance DC between point p_i^1, p_j^2 is

$$\mathrm{DC}\left(p_i^1, p_j^2\right) = \frac{1}{2} \times \left(\frac{\left|K\left(p_i^1\right) - K\left(p_j^2\right)\right|}{\left|K\left(p_i^1\right)\right| + \left|K\left(p_j^2\right)\right|} + \frac{\left|H\left(p_i^1\right) - H\left(p_j^2\right)\right|}{\left|H\left(p_i^1\right)\right| + \left|H\left(p_j^2\right)\right|} \right). \tag{5}$$

However, in many cases, a lot of plane points, whose Gaussian curvature and mean curvature are zero, can exist in the surface. They cannot contribute to the matching but will occupy large computing and reduce the matching efficiency, even lead to an error matching. So that these plane points will be discarded before matching.

Definition 2. A Point-Pair Set named PS is a set defined as

$$\mathrm{PS} = \left\{ \left(p_i^1, p_j^2\right) \mid p_i^1 \in S_1, p_j^2 \in S_2, \; \mathrm{DC}\left(p_i^1, p_j^2\right) \left\langle \varepsilon c, \left|K\left(p_i^1\right)\right| \right\rangle \sigma \right\}$$
$$\varepsilon c = \frac{\max\left(\mathrm{DC}\left(p_i^1, p_j^2\right)\right) + \min\left(DC\left(p_i^1, p_j^2\right)\right)}{2} \tag{6}$$
$$\sigma = \frac{1}{n}\left(\sum_{i=1}^{n} \left|K\left(p_i^1\right)\right| \right).$$

Where εc is an average threshold to guarantee the similarity of p_i^1 and p_j^2, σ is to discard plane points, and N is the number of the points.

The cardinality of PS could be large in special cases. For instance, cardinality of PS will be $|S_1| \times |S_2|$ when the sample and object surface are congruent sphere surface. A large Point-Pair Set will occupy a large amount of the storage space. Considering the quantity of curvature categories cannot exceed the quantity of pixel points in surface, a list can be employed to store the set.

3. Match Algorithm

3.1. Measure Similarity for Spatial Triangles. In order to perform the matching, the further work is to separately seek two asymmetrical triangles with similar shape features from the sample and object surface, and then we will design a set named "Triangle-Pair Set" to store these triangle pairs. A plane can be determined by three spatial points, so the asymmetrical triangle can be employed as the skeleton to uniquely confirm the location of a spatial surface. Therefore, we arbitrarily select three points from the object surface to form an asymmetrical triangle and then seek three corresponding points from sample surface to form another asymmetrical triangle, whose features of corresponding vertices and edges are similar to the previous triangle. The similarity of space triangles is measured by the definition as follows.

Definition 3. Consider $\forall(p_1^1, p_1^2)$, (p_1^1, p_1^2), $(p_1^1, p_1^2) \in$ PS, $|p_1^m p_2^m| < |p_1^m p_3^m| < |p_2^m p_3^m|$, $m \in \{1, 2\}$. The similar distance between $\Delta p_1^1 p_2^1 p_3^1$ and $\Delta p_1^2 p_2^2 p_3^2$ is

$$\mathrm{DT} = \frac{1}{3} \times \sum_{1 \le i < j \le 3} \frac{\left| \left| p_i^1 p_j^1 \right| - \left| p_i^2 p_j^2 \right| \right|}{\left| p_i^1 p_j^1 \right| + \left| p_i^2 p_j^2 \right|}. \tag{7}$$

The absolute distance between $\Delta p_1^1 p_2^1 p_3^1$ and $\Delta p_1^2 p_2^2 p_3^2$ is

$$\mathrm{DA} = 2 \times \frac{\sum_{i=1}^{3} \left| p_i^1 p_i^2 \right|}{\sum_{1 \le i < j \le 3}^{3} \left(\left| p_i^1 p_j^1 \right| + \left| p_i^2 p_j^2 \right| \right)}. \tag{8}$$

3.2. Compute Geometry Rigid Transformation. Suppose that $\Delta p_1^1 p_2^1 p_3^1$ and $\Delta p_1^2 p_2^2 p_3^2$ are two similar asymmetrical triangles, (p_i^1, p_i^2), and $i \in \{1, 2, 3\}$ are corresponding points, we represent the transformation T between $\Delta p_1^1 p_2^1 p_3^1$ and $\Delta p_1^2 p_2^2 p_3^2$ into a translation transformation N and a rotation transformation A:

$$T\left(\Delta p_1 p_2 p_3\right) = A\left[p_1 \quad p_2 \quad p_3\right] + N, \tag{9}$$

where $A = (n_1 n_2 n_3)^T$ is a rotation matrix and N is a translation matrix.

We firstly shift the triangle centers to the origin of coordinate, so the transformation T satisfies

$$\left\lfloor p_1^2 \quad p_2^2 \quad p_3^2 \right\rfloor - N_2 = A\left(\left\lfloor p_1^1 \quad p_2^1 \quad p_3^1 \right\rfloor - N_1\right), \tag{10}$$

where N_1, N_2 is the center location matrix of $\Delta p_1^1 p_2^1 p_3^1$ and $\Delta p_1^2 p_2^2 p_3^2$.

Therefore, rotation matrix can be deduced as in the following formula:

$$A = \left(\left\lfloor p_1^2 \quad p_2^2 \quad p_3^2 \right\rfloor - N_2\right)\left(\left\lfloor p_1^1 \quad p_2^1 \quad p_3^1 \right\rfloor - N_1\right)^{-1}. \tag{11}$$

3.3. Generate Triangle-Pair Set. Similar triangle pairs from two surfaces will be stored in Triangle-Pair Set defined as follows:

Definition 4. A Triangle-Pair Set named TS is a set

$$\mathrm{TS} = \left\{\left(\Delta p_1^1 p_2^1 p_3^1, \Delta p_1^2 p_2^2 p_3^2, T\right) \mid \left(p_i^1 p_i^2\right) \in \mathrm{PS}, \mathrm{DT} < \varepsilon t\right\}$$
$$\varepsilon t = \frac{\max(\mathrm{DT}) + \min(\mathrm{DT})}{2}, \tag{12}$$

where εt is an average threshold to guarantee the similarity of $\Delta p_1^1 p_2^1 p_3^1$ and $\Delta p_1^2 p_2^2 p_3^2$. T is the transformation from $\Delta p_1^2 p_2^2 p_3^2$ to $\Delta p_1^1 p_2^1 p_3^1$. So we can design the logic structure of Triangle-Pair Set as Figure 1.

3.4. Determine Optimal Transformation. The validity of every transformation will be tested through applying it to every other triangle pair in TS. For a specified transformation T_0 in PS, we define the scoring function as follows:

$$\mathrm{Score}\left(T_0\right) = \sum S\left(\Delta p_1^1 p_2^1 p_3^1, p_1^2 p_2^2 p_3^2, T_0\right). \tag{13}$$

Consider $(\Delta p_1^1 p_2^1 p_3^1, \Delta p_1^2 p_2^2 p_3^2)$ is an arbitrary similar triangle pair in PS, and function S is defined as follows:

$$S\left(\Delta p_1^1 p_2^1 p_3^1, \Delta p_1^2 p_2^2 p_3^2, T_0\right) = \begin{cases} 1 & \mathrm{DA}\left(T_0\left(\Delta p_1^1 p_2^1 p_3^1\right), \Delta p_1^2 p_2^2 p_3^2\right) < \varepsilon a \\ 0 & \text{others} \end{cases}$$
$$\varepsilon a = \frac{\max(\mathrm{DA}) + \min(\mathrm{DA})}{2}, \tag{14}$$

where εa is an average threshold to determine whether $\Delta p_1^1 p_2^1 p_3^1$ can be transformed to $\Delta p_1^2 p_2^2 p_3^2$ by T_0. An optimal transformation T_P is the transformation with maximum score satisfying

$$\mathrm{Score}\left(T_P\right) = \max\left(\mathrm{Score}\left(T\right)\right), \tag{15}$$

where T is an arbitrary transformation in TS.

4. Experimental Results and Complexity Analysis

4.1. Detect Partially Occluded Object. The goal of the first experiment is to detect the partially occluded object from a range image using the algorithm proposed in this paper. The image data is established referred to [21].

In Figures 2(a) and 2(b), two original objects (Duck and Venusm) in 3D space are demonstrated, and the surface is formed by pixel points.

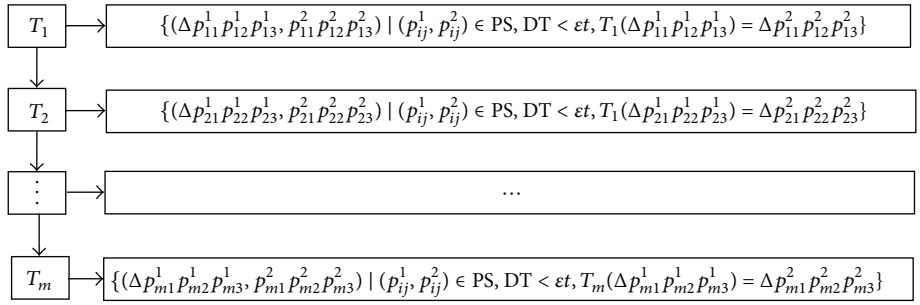
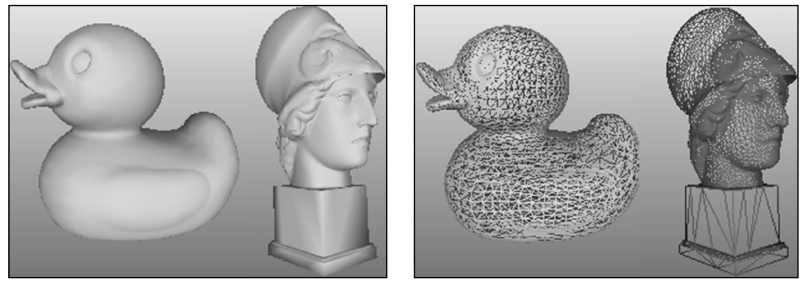
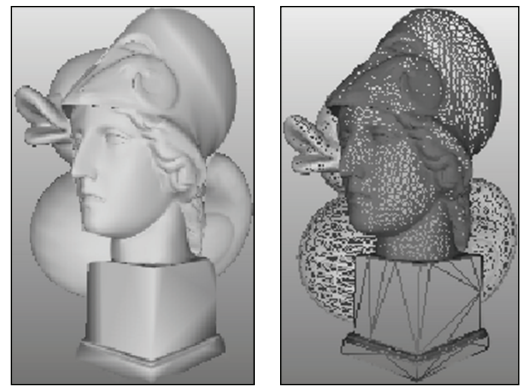


Figure 1: Logic structure of the Triangle-Pair Set.

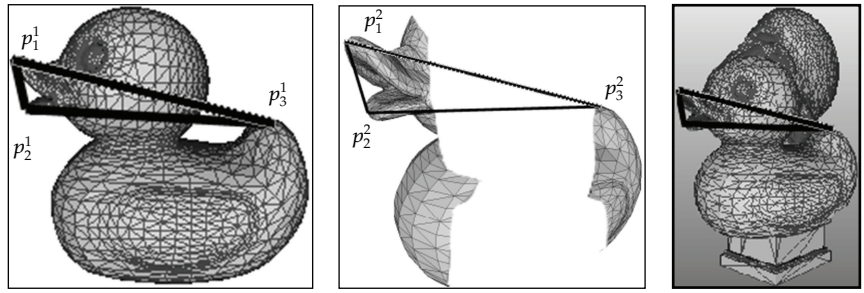


(a) Original object (b) 3D model of original object



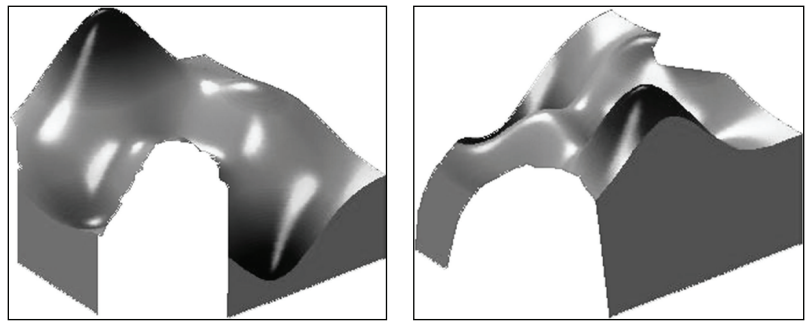
(c) Duck is partially occluded (d) 3D model of the scene

Figure 2: Two 3D original and partially occluded models.



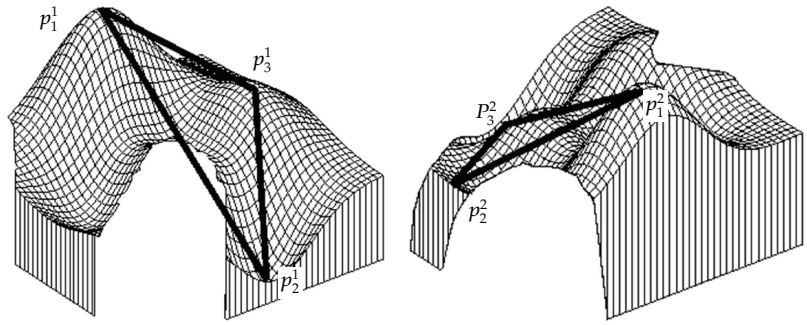
(a) Optimal triangle in original object (b) Optimal triangle in occluded object (c) Reconstruction result to occluded object

Figure 3: Detect partially occluded object.

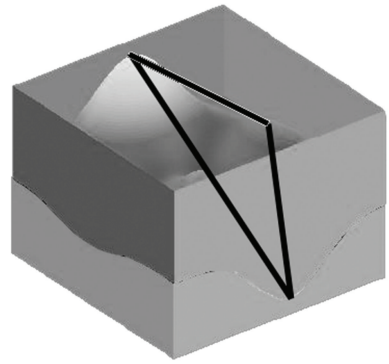


(a) Surface 1 (b) Surface 2

Figure 4: Two partially similar surfaces.



(a) Optimal triangle in surface 1 (b) Optimal triangle in surface 2



(c) Matching and reconstruction result

Figure 5: Partial surface matching of two similar surfaces.

Table 1: Parameters and computing results.

Pixel point quantity	Duck: 9253 Venusm: 27595
Triangle pair for optimal transformation	$p_1^1(125, 23, -15)$ $p_1^2(132, 5, 98)$ $p_2^1(72, 4, -10)$ $p_2^2(83, -18, 108)$ $p_3^1(68, -71, -23)$ $p_3^2(76, -89, 82)$
Transformation	$A = \begin{bmatrix} 1.2984 & -0.6040 & -0.6944 \\ -0.6401 & 1.2802 & -0.6401 \\ -0.6583 & -0.6763 & 1.3345 \end{bmatrix}$

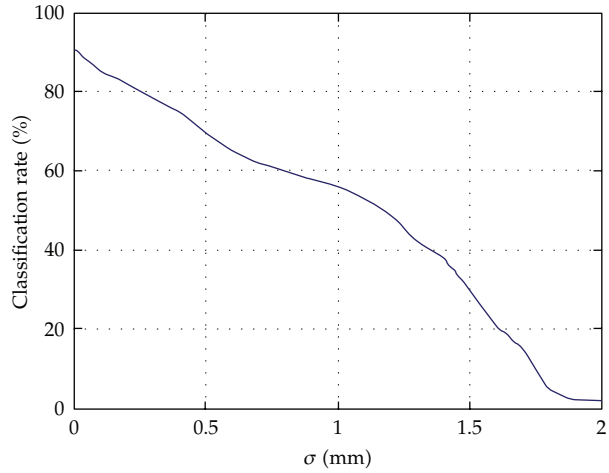


Figure 6: Classification rate when Gaussian noise increases.

Table 2: Parameters and computing result.

Pixel point quantity	Object 1: 1215 Object 2: 1452
Triangle pair for optimal transformation	$p_1^1(10,-5,87)$ $p_1^2(12,-69,18)$ $p_2^1(8,-21,42)$ $p_2^2(25,-31,32)$ $p_3^1(29,-17,32)$ $p_3^2(92,-43,28)$
Transformation	$A = \begin{bmatrix} 0.1028 & 1.7571 & -1.8599 \\ -0.2024 & 1.1216 & -0.9191 \\ 0.0996 & -2.8787 & 2.7791 \end{bmatrix}$

Table 3: Performance comparison of some algorithms.

Algorithm	Running time (ms)	
	Sphere surface	Surface in Figure 3
Algorithm in this paper	2635	758
Algorithm in [9]	1932	2575
Algorithm in [10]	2956	3548
Algorithm in [11]	1562	1652
Algorithm in [14]	1457	1238
Algorithm in [16]	3187	4572

In Figure 2(c), the Duck is partially occluded by the Venusm. In order to find the location of the occluded object (Duck) in the 3D scene, we must compute an optimal transformation from the scene to the occluded object, then apply the transformation to the occluded object and reconstruct it in the 3D scene in Figure 2(d).

In this experimentation, the computing results are demonstrated in Table 1.

Optimal triangle in original object is demonstrated in Figure 3(a), and the optimal triangle in occluded object is illustrated in Figure 3(b). According to the result in Table 1, we apply the transformation to original object. The reconstruction result is shown in Figure 3(c).

4.2. Partial Surface Matching for Two Similar Objects. The second experimentation demonstrates our algorithm applied in a universal case. The two surfaces are partially similar, and the goal is to find a transformation to match the most similar part of the two surfaces.

The two surfaces with missing data in Figures 4(a) and 4(b) are generated by a Matlab function. In order to match the most similar part of the two surfaces, an optimal transformation from surface 1 to surface 2 needs to be computed.

The computing results in this experimentation are demonstrated in Table 2.

Optimal triangle in surface 1 is demonstrated in Figure 5(a), and the optimal triangle in surface 2 is illustrated

in Figure 5(b). According to the results shown in Table 2, the transformation is applied to surface 1 and the matching and reconstruction result is demonstrated in Figure 5(c).

4.3. Complexity Analysis. Suppose that the two input sets have comparable sizes, we measure the computing complexity of the algorithm as a function of N (the cardinality of the input sets); k is the size of Point-Pair Set. Computing the curvature of every point takes $O(N)$ time. Preparing the Point-Pair Set can be executed with expected $O(N \log N)$ running time because we can sort the points based on curvature. In addition, seeking triangle pairs, computing and testing transformation based on Point-Pair Set will cost $O(k^3)$ running time. In total, the running time of the whole algorithm is $O(N + N \log N + k^3)$.

Obviously, the computing efficiency is mainly determined by the size of Point-Pair Set k. The quantity of point pairs will be smaller when the surface processes many different curvatures; in this case, our algorithm will cost less time at this time. So our algorithm is very suitable for the matching of irregular objects.

In order to explicitly describe the performance of our algorithm, we perform experimentations for objects with different shapes. Running times of some algorithms for $N = 900$ are demonstrated in Table 3 (CPU: PIV2.0 GHZ, RAM: 1 GB, Software: MATLAB 7.0).

We can see from Table 3 that our algorithm cannot indicate an encouraging matching efficiency for regular objects such as sphere surfaces.

4.4. Noise Robustness. The proposed work depends crucially on the Gaussian curvature and means curvature computation, which is notoriously sensitive to noise and local perturbation. In order to analyze the affection of noise to the match results, we perform an experimentation to verify the noise robustness of this algorithm. In this experimentation, we generate 100 similar surfaces by Matlab function, and Gaussian noise $(N(0, \sigma))$ is added to each image in the database and σ changes from 0 mm to 2.0 mm. We design the classification rate to represent the percentage of correct matching quantity in the database; the classification rates are shown in Figure 6 for various σ values.

The result shows that the algorithm in this paper does not have good noisy robustness. At present, our approach can get encouraging matching efficiency and running time complexity in case of high signal to noise ratio. Therefore, the future work will concentrate on this problem.

5. Conclusion

In this paper, an approach for solving the 3D partial model matching is proposed. Firstly we utilize Gaussian curvature and mean curvature to represent the two object surfaces. Then curvature distance is defined to classify the point pair with similar curvature, and Point-Pair Set is employed to store point pairs. Secondly, we classify triangle pair with similar shape features from the two surfaces based on Point-Pair Set, similar distance is employed to measure the similarity of two triangles, and then the Triangle-Pair Set is generated. Finally, absolute distance and score function are proposed for measuring the goodness of a given transformation in Triangle-Pair Set, and a successful matching is determined based on the score of optimal transformation between the objects. Experimental results indicated our algorithm can be suitable for the matching of natural objects.

However, the efficiency of this approach would be reduced when the shape is regular and the robustness to noise is not very good. In future work, we will plan to explore several improvements of the algorithm, including the improvement of noise robustness, design of better distance definition, experimentation with other scoring functions, and further study of statistical approaches.

Acknowledgments

This work is supported by Research Fund for the Doctoral Program of Higher Education of China (20090162120069), Science and Technology Plan of Hunan (2009FJ3016), Postdoctoral Fund of Central South University, China Postdoctoral Science Foundation (2012M521552), and NSFC (61202341).

References

[1] L. Ritter, S. D. Reiz, D. Rothamel et al., "Registration accuracy of three-dimensional surface and cone beam computed tomography data for virtual implant planning," *Clinical Oral Implants Research*, vol. 23, no. 4, pp. 447–452, 2012.

[2] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.

[3] S. Chen, Z. Wang, H. Tong et al., "Optimal feature matching for 3D reconstruction by combination of global and local information," *Intelligent Automation and Soft Computing*, vol. 17, no. 7, pp. 957–968, 2011.

[4] K. Bunte, M. Biehl, M. F. Jonkman, and N. Petkov, "Learning effective color features for content based image retrieval in dermatology," *Pattern Recognition*, vol. 44, no. 9, pp. 1892–1902, 2011.

[5] S. M. Tarte, H. Talib, M. A. G. Ballester, and F. Langlotz, "Evaluating partial surface matching for fracture reduction assessment," in *Proceedings of the IEEE 3rd International Symposium on Biomedical Imaging: Nano to Macro*, vol. 4, pp. 514–517, Arlington, Va, USA, April 2006.

[6] Y. Shibberu and A. Holder, "A spectral approach to protein structure alignment," *ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 4, pp. 867–875, 2011.

[7] G. Barequet and M. Sharir, "Partial surface and volume matching in three dimensions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, pp. 929–948, 1997.

[8] G. Barequet and M. Sharir, "Partial surface matching by using directed footprints," *Computational Geometry*, vol. 12, no. 1-2, pp. 45–62, 1999.

[9] H. Alt, P. Bras, M. Godau, C. Knauer, and C. Wenk, "Computing the Hausdorff distance of geometric patterns and shapes," *Discrete and Computational Geometry, Special Issue—The Goodman-Pollack-Festschrift*, vol. 25, pp. 65–76, 2003.

[10] N. J. Mitra, L. J. Guibas, and M. Pauly, "Partial and approximate symmetry detection for 3D geometry," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 560–568, 2006.

[11] R. Gal and D. Cohen-Or, "Salient geometric features for partial shape matching and similarity," *ACM Transactions on Graphics*, vol. 25, no. 1, pp. 130–150, 2006.

[12] S. Wang, Y. Wang, M. Jin, X. D. Gu, and D. Samaras, "Conformal geometry and its applications on 3D shape matching, recognition, and stitching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1209–1220, 2007.

[13] J. Tierny, J. P. Vandeborre, and M. Daoudi, "Partial 3D shape retrieval by Reeb pattern unfolding," *Computer Graphics Forum*, vol. 28, no. 1, pp. 41–55, 2009.

[14] U. Castellani, M. Cristani, S. Fantoni, and V. Murino, "Sparse points matching by combining 3D mesh saliency with statistical descriptors," *Computer Graphics Forum*, vol. 27, no. 2, pp. 643–652, 2008.

[15] A. Itskovich and A. Tal, "Surface partial matching and application to archaeology," *Computers and Graphics*, vol. 35, no. 2, pp. 334–341, 2011.

[16] H. Tabia, M. Daoudi, J. P. Vandeborre, and O. Colot, "A new 3D-matching method of nonrigid and partially similar models using curve analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 852–858, 2011.

[17] S. Y. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.

[18] S. Chen, Y. Wang, and C. Cattani, "Key issues in modeling of complex 3D structures from video sequences," *Mathematical Problems in Engineering*, vol. 2012, Article ID 856523, 17 pages, 2012.

[19] K. Guo, "A novel matching algorithm to 3D incomplete object," *Journal of Computational Information Systems*, vol. 7, no. 1, pp. 73–79, 2011.

[20] K. H. Guo and G. H. Duan, "3D partial surface matching using differential geometry and statistical approaches," *Lecture Notes in Computer Science*, vol. 7751, 2012.

[21] A. S. Mian, M. Bennamoun, and R. Owens, "Three-dimensional model-based object recognition and segmentation in cluttered scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1584–1601, 2006.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 509597, 18 pages
doi:10.1155/2012/509597

Research Article

Adaptive Colour Feature Identification in Image for Object Tracking

Feng Su,[1] Gu Fang,[1] and Ngai Ming Kwok[2]

[1] *School of Computing, Engineering, and Mathematics, University of Western Sydney, Locked Bag 1797, Penrith, NSW 2751, Australia*
[2] *School of Mechanical and Manufacturing Engineering, The University of New South Wales, Sydney, NSW 2052, Australia*

Correspondence should be addressed to Gu Fang, g.fang@uws.edu.au

Received 8 October 2012; Accepted 14 November 2012

Academic Editor: Sheng-yong Chen

Identification and tracking of a moving object using computer vision techniques is important in robotic surveillance. In this paper, an adaptive colour filtering method is introduced for identifying and tracking a moving object appearing in image sequences. This filter is capable of automatically identifying the most salient colour feature of the moving object in the image and using this for a robot to track the object. The method enables the selected colour feature to adapt to surrounding condition when it is changed. A method of determining the region of interest of the moving target is also developed for the adaptive colour filter to extract colour information. Experimental results show that by using a camera mounted on a robot, the proposed methods can perform robustly in tracking a randomly moving object using adaptively selected colour features in a crowded environment.

1. Introduction

Surveillance is the task of monitoring the behaviours and/or activities of people from a distance. Security cameras are considered to be the most commonly used equipment. These cameras are used for applications such as, industrial process control, traffic monitoring, and crime prevention. However, despite their wide usages, security cameras still have many weaknesses. One of the weaknesses is its blind spot [1], since they are mounted on mechanical hinges, cameras are only able to monitor at certain angles, and the security system can be infiltrated through those unseen areas. Another weakness rests on the involvement of human operators [2], who usually monitor a large number of inputs from cameras. Because these operators could be subjected to boredom, fatigue, and distractions, it is possible that they

fail to detect criminal or other unwanted behaviours. Therefore, a mobile robot could be used to overcome these potential problems. A robot would be able to travel throughout the monitoring areas autonomously and continuously, making its own decisions while identifying the unwanted behaviours or activities, and respond accordingly such as sending alerts.

Object tracking using computer vision is a crucial component in achieving robotic surveillance. The goal of object tracking is to track the position of the moving objects in a video sequence. This can be achieved by identifying and tracking a specific feature such as colour that belongs to the moving object. The trajectories of the moving object can then be traced through the process over time.

Most existing colour tracking methods are designed to track a fixed salient colour feature. However, if the camera is moving then the tracked colour feature may be no longer salient due to the changing environment. In this case, the tracking may take place to follow a wrong object. Therefore, new methods are required so that the colour feature can be determined adaptively according to the environment the camera is operating in.

The main contribution of this paper is the introduction of a colour filtering method that is capable of adaptively identify the most salient colour feature that belongs to the moving object and using this colour feature for tracking. If the saliency of the chosen colour feature changes due to the operating environment and lighting conditions, the filter will automatically determine a different colour to track. A method of determining the region of interest (ROI) of the moving target is also introduced for the adaptive colour filter to extract colour information.

This paper is organised into six sections. The related research in the area is reviewed in Section 2. The method of determining the ROI is provided in Section 3. In Section 4, an adaptive colour filtering method is introduced. Experimental results are presented in Section 5, and conclusions are given in Section 6.

2. Related Work

2.1. Moving Object Detection

Moving object detection and tracking is an important and fundamental topic in computer vision. Its applications can be found in a large number of engineering fields including: traffic monitoring [3], video surveillance [4], autonomous navigation [5], and robotics [6–10]. One of the most commonly used methods to detect moving objects is background subtraction [11, 12].

Background subtraction involves the separation of moving foreground objects from the static background. The fundamental assumption of the algorithm is that the background is relatively static compared to the foreground. When objects move, the regions in a set of video frames that differ significantly from the background model can be considered to be the foreground (moving objects). A vast amount of research in moving object detection has been done with many algorithms proposed. The most fundamental method uses the Gaussian Mixture Model (GMM) [13]. This method models the evolution of each background pixel intensity by a mixture of K (a small number usually from 3 to 5) Gaussian distributions. There have also been many revised and improved methods based on GMM. One of them is the Improved Adaptive Gaussian Mixture Model (AGMM) [14]. In this model, both the parameters and the number of components of the mixture are constantly adapted. Another

enhanced version is the Improved Adaptive Background Mixture Model (IABMM) [15]. In the work reported, the likelihood factor is removed from the GMM, because it causes slow adaptations in the means and the covariance matrices which can result in failure of the tracker. The IABMM also contains an online Expectation Maximization algorithm which provides a good initial estimate by expected sufficient statistics update equations before enough samples can be collected. Other background subtraction methods include the Codebook [16]. In this method, each pixel is represented by a Codebook which is a compressed form of background model for a long image sequence. This allows the method to capture background variation over a long time with a low memory requirement.

2.2. Moving Object Tracking

Although the background subtraction-based method can robustly identify moving objects with a stationary camera, it cannot provide satisfactory results with a moving camera. This is because the background subtraction methods extract the foreground by distinguishing the differences between the moving objects and a "stationary" background. This difference finding mechanism is built on the assumption that background stays longer and is more static when comparing to the foreground. If a moving camera is used, for example, a camera mounted on a mobile robot, background subtraction will face the problem that the background of the image is constantly changing due to camera movement. This will lead to false classification of the majority of an image to be foreground. This false classification will cause the moving camera system to lose track of the target object. Therefore, existing object tracking methods using mobile robots (moving cameras) usually rely on certain features belong to the tracked objects, such as colour.

A moving object tracking method using mobile robot is performed with background subtraction and colour probability distribution [17]. This is done by stopping the robot when background subtraction is performed and using the colour probability distribution information to track the target. This method assumes that the colour of the tracked object never changes which is not always the case. Also, the locomotion of the robot was remotely controlled in the experiment rather than fully autonomous.

Another object tracking approach is developed based on scale invariant feature transform (SIFT) and mean shift [18]; SIFT is used to find features corresponding to the region of interests, while mean shift is used to find similarities in the colour histograms. This method combines the advantages of both SIFT and mean shift to achieve more accurate tracking results; however, due to high computation costs, it also has the slowest processing speed (1.1 fps) when compared to SIFT or mean shift alone. This level of computation complexity has imposed difficulties in real-time applications.

Object tracking method can also be used for skin segmentation and the tracking of sign language recognition [19]. This method could track the face and the hand accurately using a colour model with a stationary camera. However, the testing background is fairly simple, the distance between the testing subject and the camera remains very close as the subject always occupies large part, sometimes more than half, of the image.

Other colour tracking methods include tracking by transductive learning which requires a high computational cost [20], colour tracking specifically designed for illumination variations in an uncontrolled operating environment [21], multicamera colour tracking which relies on accurate target identification between different cameras [22], kernel-based object tracking using colour and boundary cues [23], object tracking using mean shift with

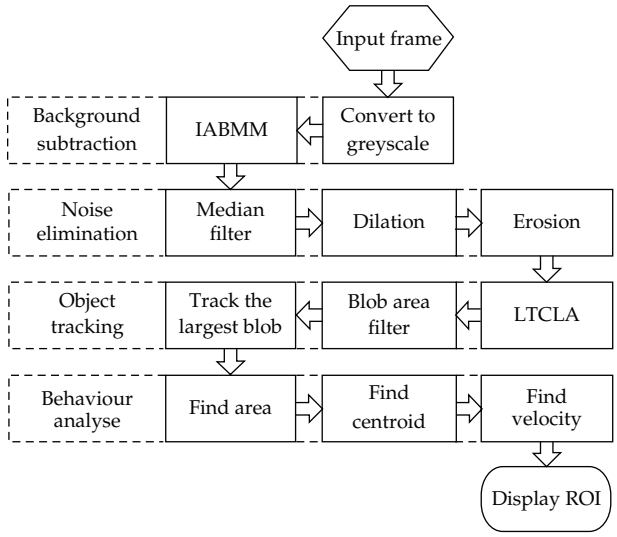


Figure 1: Flowchart of the ROI determination algorithm.

a clustering algorithm and colour model [24], selecting reliable features from colour and shape-texture cues [25], or using area weighted mean of the centroids [26].

The majority of existing colour tracking methods is designed for stationary cameras without the use of mobile robots. Furthermore, the environments presented within these methods are generally stable with little to no variances. To the best of our knowledge, there is no reported method on using GMM-based approach with adaptive colour feature selection in moving object tracking using a robot-mounted camera.

3. Region of Interest (ROI) Determination

As illustrated by the flowchart in Figure 1, the ROI determination algorithm starts by converting the RGB input to both Hue-Saturation-Value (HSV) and greyscale images. It includes four major stages: background subtraction, noise elimination, object tracking, and behaviour analysis. More details of each step are given in the following sections.

3.1. Background Subtraction

To perform background subtraction, the live RGB videos are firstly converted into greyscale images. Greyscale images are used as inputs for background subtraction process because they require less memory to operate and produce faster processing speed than colour images [27, 28].

Then, the IABMM [15] method is used to identify the moving objects and present them in a binary image as foreground objects using white pixels, while allocating all stationary objects as the background using black pixels. The IABMM method is used because it has a faster learning rate and a lower computation requirement than GMM [15, 29]. Thus, it is very efficient to detect motions of objects especially in indoor environments.

The IABMM method used is an improved version of the GMM [13], it begins with the original GMM equation to determine the probability of a given pixel with value X_t at time t using

$$P(X_t) = \sum_{i=1}^{K} \omega_i \eta(X_t; \mu_i, \Sigma_i), \tag{3.1}$$

where K is the number of Gaussians used for the mixture, ω_i is the weight parameter of the ith Gaussian component, and $\eta(X_t; \mu_i, \Sigma_i)$ is the normal distribution of the ith component with the mean μ_i and the covariance Σ_i.

The major improvement of the IABMM [15] method is the inclusion of the online Expectation Maximization algorithm. This is done by updating the model differently at different phases. Initially when the number of frames t is smaller than L, the model is updated according to the expected sufficient statistics update equations, shown in (3.2), (3.3), and (3.4), then switch to L-recent window update equations when the first L samples (frames) are processed, see (3.5), (3.6), and (3.7). The expected sufficient statistics update equations increase the performance in the beginning while providing a good estimate by allowing fast convergence on a stable background model. The tracker can also adapt to changes in the environment because the L-recent window update equations gives priority over recent data;

$$\omega_{k,t+1} = \omega_{k,t} + \frac{1}{t+1}(P(\omega_k \mid X_{t+1}) - \omega_{k,t}), \tag{3.2}$$

$$\mu_{k,t+1} = \mu_{k,t} + \frac{P(\omega_k \mid X_{t+1})}{\sum_{i=1}^{t+1} P(\omega_k \mid X_i)}(X_{t+1} - \mu_{k,t}), \tag{3.3}$$

$$\Sigma_{k,t+1} = \Sigma_{k,t} + \frac{P(\omega_k \mid X_{t+1})}{\sum_{i=1}^{t+1} P(\omega_k \mid X_i)}\left((X_{t+1} - \mu_{k,t})(X_{t+1} - \mu_{k,t})^T - \Sigma_{k,t}\right), \tag{3.4}$$

$$\omega_{k,t+1} = \omega_{k,t} + \frac{1}{L}(P(\omega_k \mid X_{t+1}) - \omega_{k,t}), \tag{3.5}$$

$$\mu_{k,t+1} = \mu_{k,t} + \frac{1}{L}\left(\frac{P(\omega_k \mid X_{t+1})X_{t+1}}{\omega_{k,t+1}} - \mu_{k,t}\right), \tag{3.6}$$

$$\Sigma_{k,t+1} = \Sigma_{k,t} + \frac{1}{L}\left(\frac{P(\omega_k \mid X_{t+1})(X_{t+1} - \mu_{k,t})(X_{t+1} - \mu_{k,t})^T}{\omega_{k,t+1}} - \Sigma_{k,t}\right), \tag{3.7}$$

where $\omega_{k,t+1}$, $\mu_{k,t+1}$, and $\Sigma_{k,t+1}$ are the estimates of weight, mean, and covariance of the kth Gaussian component at time $t+1$, respectively. $P(\omega_k \mid X_{t+1})$ is the posterior probability that X_{t+1} is generated from the kth Gaussian component. Note that $P(\omega_k \mid X_{t+1}) = 1$ for the matched model and 0 for the remaining models.

3.2. Noise Elimination

After the background subtraction, noise elimination is performed to filter possible noises caused by reflections or motion blurs. The noise elimination consists of median filtering and binary morphological operations.

Median filter [30] is used to remove the so-called "salt and pepper" noise and to restore foreground pixels while preserving useful details.

Noises caused by a changing background or illumination condition may misidentify some of background pixels to be the foreground objects or produce gaps or holes within the foreground objects and separate them into different regions.

Morphological operations, that is, dilation and erosion [31], are used to reduce noise by connecting possible foreground regions and removing any false ones. Dilation is done by first computing the maximal pixel value overlapped by the kernel and then replaces the image pixel under the anchor point with that maximum value. The kernel used is a 3 by 3 kernel with the anchor at its centre. Moreover, erosion is the converse function that operates with minimum value instead of maximum. By combining dilation and erosion, this results in morphological closing which causes the bright regions to join together to form blobs and therefore improves the detection of foreground (represented by white blobs).

3.3. Object Tracking

Object tracking stage begins with the tracking of the blobs (if any) from the output binary image of noise elimination, and they are tracked using linear time component labelling algorithm (LTCLA) [32], which is a fast labelling technique that labels connected components and their contours simultaneously.

The major component of this algorithm involves a contour tracing technique with a tracer to detect the external contour and internal contours of each component. Once a contour point is identified, the tracer works by searching other contour points among its eight neighbours in a clockwise direction. If the initial point is the starting point of an external contour, the search begins at top right, while if the initial points is the starting point of an internal contour, then the search begins at bottom left. Moreover, when the initial point is not the starting point of a contour, the search begins at a point located at 90 degrees clockwise from the position of the previous contour point. This process also marks surrounding pixels (background represented with black pixels) when tracing the contour of the component (foreground represented with white pixels). By marking surrounding background pixels, this ensures no overtracing of the contour occurs. Unless the initial point is identified as an isolated point, the tracer is used continuously in this contour tracing procedure to output the contour point following the initial point until a full loop is done which the entire external or internal contour is traced. LTCLA is used because of its high efficiency. Once the labelling stage is completed, blobs can then be selected for further analysis through a series of filtering processes.

Blobs identified using the LTCLA need to be filtered to eliminate unnecessarily small blobs in order to reduce computational costs. This can be done by using a blob area filter. In our experiments, any blob that contains the pixel number below 0.5% of the total number of pixels in the image is eliminated.

The next step in the object tracking stage is the determination of which blob to track. In this paper, it is decided that the tracking object is the largest moving blob identified. Although the algorithm can be used to track multiple targets, it has been found the speed of the algorithm is heavily influenced by the number of objects tracked. Therefore, in this method, only the largest blob is selected and tracked. This simplification is also justified by the fact that one robot can only track one moving object.

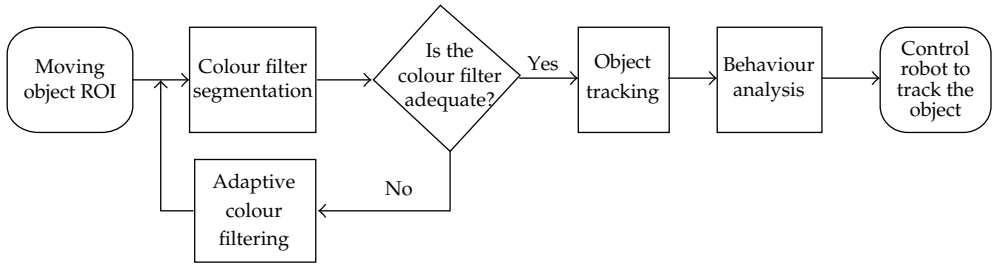


Figure 2: The control concept of the ACF.

3.4. Behaviour Analysis

After the identification of the largest blob, the final stage of the ROI determination algorithm is the behaviour analysis of the blob. The area, centroid, and velocity are obtained and can be used to determine the behaviour characteristics of the object.

To calculate these behaviour characteristics, a ROI is established by a bounding box that encloses the target object and is determined by using the maximum width and height of that object. The area of the object is calculated by counting the number of pixels existed in the tracked blob.

The centroid coordinates X_C and Y_C for the blob and can be found using the centre-of-mass;

$$X_C = \frac{\sum_{i=1}^{K} X_i}{K},$$
$$Y_C = \frac{\sum_{i=1}^{K} Y_i}{K}, \tag{3.8}$$

where X_i and Y_i are the x and y location of the ith pixel in the image plane, while K is the total number of pixels belong to the blob.

Once the coordinate of the centroid is found, the velocity can be obtained by comparing the centroid's locations between video frames obtained at different time steps. In this method, previous four centroid coordinates are stored to indicate the path and moving direction of the tracked target, while the velocity is calculated from the difference in pixel locations of the centroid in the current and that of the immediate last image.

4. Adaptive Colour Filter for Moving Object Tracking

The adaptive colour filter (ACF) is a colour tracking method developed in this paper for using a robot mounted camera to track a moving object. The control concept of the ACF is shown in Figure 2.

Initially, the robot and the camera are stationary. Moving objects are detected using the IABMM method. Once the ROI of the moving object is determined, the colour information of the object and the background is filtered using a colour filter to find the most salient feature in the object for tracking. This feature is then used to track the object. When the selected colour feature becomes no longer salient due to changed environment, an adaptive method

Table 1: List of properties defined for the 15 colour filter segments.

Filter segment	Colour	HSV range					
		Hue (°)		Saturation		Value	
		Min	Max	Min	Max	Min	Max
1	White	0	360	0	0.25	0.76	1
2	Grey	0	360	0	0.25	0.26	0.75
3	Black	0	360	0	1	0	0.25
4	Red	0 346	15 360	0.26	1	0.26	1
5	Orange	16	45	0.26	1	0.26	1
6	Yellow	46	75	0.26	1	0.26	1
7	Chartreuse	76	105	0.26	1	0.26	1
8	Green	106	135	0.26	1	0.26	1
9	Spring green	136	165	0.26	1	0.26	1
10	Cyan	166	195	0.26	1	0.26	1
11	Azure	196	225	0.26	1	0.26	1
12	Blue	226	255	0.26	1	0.26	1
13	Violet	256	285	0.26	1	0.26	1
14	Magenta	286	315	0.26	1	0.26	1
15	Rose	316	345	0.26	1	0.26	1

is introduced to update the colour selection. The following sections provide details of this method.

4.1. The Colour Filter

After the ROI is established, the HSV colour space of both the ROI and the entire image are analysed using the proposed colour filter. HSV is selected over RGB because HSV performs better in identifying objects under different lighting conditions such as in shadow, shade, and highlights. This allows the filter to have fewer segments than in RGB [33, 34]. Also, RGB colour space tends to merge neighbouring objects with different colour together and produce blurred results, while since HSV colour space separates out the intensity from the colour information, the results it produced tend to distinguish neighbouring objects with different colour by sharpening the boundaries and retaining the colour information of each pixel [35].

The proposed colour filter is designed to cover the entire HSV colour space. The number of segments of the HSV space can be defined by the user with consideration of hardware limits such as the resolution of the camera and the processing speed required for the algorithm. In this paper, 15 colour filter segments are chosen with their HSV component properties determined empirically, see Table 1.

To apply the colour filtering method, every pixel is checked against the H (hue), S (saturation), and V (value) values for a specific colour filter segment and is regarded as belong to the segment if all H, S, and V values of the pixel fall into the segment range. A uniqueness measure, U, is introduced, shown in (4.1), to measure the ratio of number of

filtered pixels that belong to the foreground (ROI), over the filtered pixels that belong to the background (outside the ROI) in that colour segment;

$$U_i = \frac{F_i}{B_i}, \tag{4.1}$$

where U_i is the uniqueness level of ith colour segment, and F_i and B_i are the number of foreground and background pixels existed in the ith colour segment.

The colour segment with the highest uniqueness value U is picked and applied to the HSV space of the same image. The HSV space is then filtered according to HSV range values of that colour segment (the 15 segments are listed in Table 1).

This filtering process will result in a binary image that has only the pixels with the colour segment of the highest U in the original image to be white while all other pixels to be black. Then, by using object tracking and behaviour analysis, same as in Sections 3.3 and 3.4, the properties required for tracking can be obtained. The tracking process can then be carried out by sending commands to the mobile robot to follow the foreground object.

4.2. Adaptive Colour Filtering (ACF)

An adaptive filtering method is introduced to allow the "picked" colour segment to be reassigned when the saliency of the tracked colour is no longer prominent or when the target object changed colour appearance possibly due to changing lighting conditions.

To invoke the application of the ACF, one of the following three conditions has to be met.

(1) When a sudden increase of pixels that share the same colour as the tracked target object is detected and exceeds a certain threshold as

$$\frac{B_i}{B_{i-1}} > T_1, \tag{4.2}$$

where B_i is the total number of the detected pixels presented in the ith image frame, T_1 is a control threshold, and the value used is $T_1 = 2$. This threshold value is chosen as it is expected that a 100% increase in the pixel number that share the same colour as the identified colour would make the chosen colour no longer unique or salient. This threshold value, T_1, was chosen by running the experiments under different conditions in different environments. It was found that by making the T_1 smaller it would increase use of ACF therefore increase the computational load. However, if T_1 is set too large, then the wrong tracking will occur. It was found that by setting the current value of 2 a balanced outcome could be achieved.

(2) When the blob is too large and has covered a significant part of the image, as determined by

$$\frac{(x_{\max} - x_{\min})(y_{\max} - y_{\min})}{X_d Y_d} > T_2, \tag{4.3}$$

where $x_{\min}$, $x_{\max}$, $y_{\min}$, and $y_{\max}$ are the minimum and maximum x and y values of the ROI of the tracked target object, X_d and Y_d are the width and height of the input frame determined

by the camera resolution, T_2 is a threshold, and the value used is $T_2 = 0.5$. This threshold value is chosen as it is expected that when the blob covers a significant part of the image, a reanalysis is needed as the chosen colour may no longer be unique or salient. This threshold selection is similar to the selection of T_1; that is, a larger T_2 will result in a potential wrong tracking while a smaller T_2 will result in unnecessary invoking of the ACF. We found that the T_2 can be chosen in the range from 0.45 to 0.55 without tracking errors, therefore a round value of 0.5 was chosen.

(3) When the speed of the blob between frames exceeds a certain threshold (sudden jump) as

$$\sqrt{\frac{(X_i - X_{i-1})^2 + (Y_i - Y_{i-1})^2}{X_d^2 + Y_d^2}} > T_3, \tag{4.4}$$

where X_i and Y_i are the x and y coordinates of the centroid of the blob in the ith frame, X_d and Y_d are the same as in (4.3), T_3 is a threshold, and the value used is $T_3 = 0.2$. This threshold number should vary depend on the moving velocity of the robot—as the robot in the experiment was moving in relative slow speed (~0.1 metre/s), and the frame rate is around 15 frames per second—this 20% movement limit of the image size is chosen. If the robot speed is increased then the value of T_3 should be increased proportionally. However, if the frame rate increases, this threshold should be decreased proportionally.

Once the ACF is invoked, the ROI is redefined according to the size of the moving object from the previous image frame. To accommodate the change in speed and position of the target object between the current and the previous image frame, the ROI's width and height for the moving object are evenly increased by N if it covers less than half of the total area of the frame, otherwise its width and height are evenly decrease by the same percentage, N. The value used is $N = 50\%$. The image is then reanalysed with the new ROI to find a new colour segment to apply. The threshold values used are based on the robot speed and frame rate and are obtained empirically.

4.3. Reducing Unneccessary Adaptation

During tracking, the colour features of the entire image are constantly updated. The largest blob that contained the selected colour feature is deemed to be the moving object. However, due to possible changes in the surrounding environment, there may be large objects in the background that has similar colour to the tracked object. To avoid the unnecessary use of the ACF (particularly the wrongly use of condition 3, (4.4), in Section 4.2), a method of "chopping" background into smaller regions has been introduced.

During the "chopping" process, a number of division lines with black pixels are introduced to cut background objects into smaller pieces. This will cut possible background objects that have similar colour to the moving object into smaller white blobs. This process is

intended to reduce the size of the blobs in the selected colour segments outside the ROI. The division lines are either horizontal or vertical, and the locations of these lines are

$$\begin{aligned} X_{k,\text{left}} &= \frac{2x_{\min} - x_{\max}}{k}, \\ X_{k,\text{right}} &= X_d - \frac{X_d - 2x_{\max} + x_{\min}}{k}, \\ Y_{k,\text{upper}} &= \frac{2y_{\min} - y_{\max}}{k}, \\ Y_{k,\text{lower}} &= Y_d - \frac{Y_d - 2y_{\max} + y_{\min}}{k}, \end{aligned} \tag{4.5}$$

where $x_{\min}$, $x_{\max}$, $y_{\min}$, $y_{\max}$, X_d, and Y_d are the same as in (4.3), while $X_{k,\text{left}}$, $X_{k,\text{right}}$ are the locations of the left and right vertical "chopping" lines of the kth cut ($k = 1, \ldots, K$), $Y_{k,\text{upper}}$, $Y_{k,\text{lower}}$ are the upper and lower "chopping" lines of the kth cut. The K value is used to determine the number of "cuts" to the larger objects outside the ROI; in this paper $K = 8$.

As division lines (black pixels) cut through objects that share the same colour (white pixels) outside the ROI, this results in the segmentation of potential background objects in same colours (noises) into smaller sizes and avoids the possibility of unnecessarily invoking the ACF process. Since the cutting is more concentrated as it is further away from the ROI, this decreases the impact of potential noises when they are further away to the centroid of the tracked object.

To avoid "chopping" the tracked object, the ROI needs to be redefined for analysis so that the change in speed and position of the target object between the current and the previous image frame can be accommodated for the cutting process. The ROI's width and height for the moving object are evenly increased or decreased by a certain percentage (similar to that mentioned in Section 4.2).

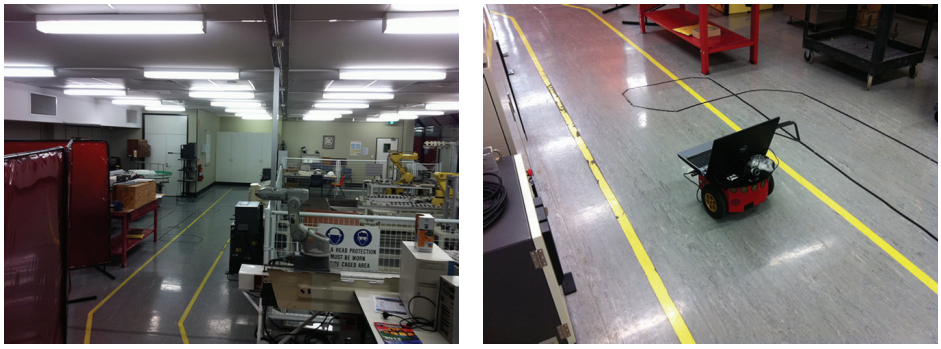
5. Experimental Results

A Pioneer 3 mobile robot is used to demonstrate the effectiveness of the proposed real-time adaptive colour feature identification and tracking method. A laptop with a 2.53 GHz duo Pentium 4 processor and a low cost Universal Serial Bus (USB) web camera with a maximum of 30 fps and 240 × 320 resolution are used to control the robot platform. The experiment is performed in the robotics laboratory that can be regarded as a crowded and noisy environment with objects of different shapes and colours presented, shown in Figure 3(a). The Pioneer 3 mobile robot is shown in Figure 3(b).

The robot is programmed using Microsoft Visual Studio C++ with Open CV, and the control of the robot is performed using the ARIA library. Although the Pioneer 3 mobile robot has many sensors, only the vision sensor provided by the web camera is used in this experiment. The robot is used to follow a randomly moving human object. The control concept of following the target object (human) is based on the area and centroid coordinates of the detected blob, and the commands and their trigger conditions are listed in Table 2. Note that the origin of the image x-y plane lies at the top left corner of the image frame as set by the camera. The movement speed of the robot and the threshold values form the trigger conditions are obtained empirically to accommodate experimental needs.

Table 2: List of commands and their trigger conditions used by the pioneer 3 mobile robot.

Commands	Trigger conditions
Drive forward at 100 mm/sec	Area falls below 4% of the entire screen
Drive backward at 100 mm/sec	Area reaches above 20% of the entire screen
Turn left 4°	Centroid lies between 21% and 40% of the maximum width
Turn left 12°	Centroid lies below 20% of the maximum width
Turn right 4°	Centroid lies between 60% and 79% of the maximum width
Turn right 12°	Centroid lies above 80% of the maximum width



(a) The robotics laboratory (b) The Pioneer 3 mobile robot

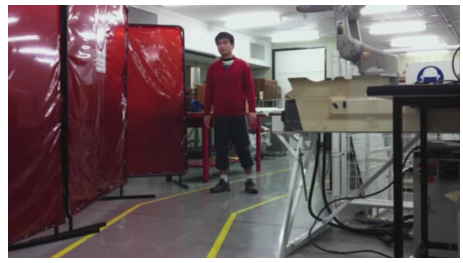
Figure 3: The robotics laboratory and the Pioneer 3 mobile robot.

Since the focus of this research is on the adaptive colour feature identification, during experiments it is assumed that the tracked person is not to walk in narrow aisles. Therefore, the robot has no obstacle avoidance method implemented. The video frames are recorded during the operation using the USB web camera attached to the laptop computer with a rate of 10 fps during the adaptive colour filtering stage and a rate of 15 fps during the moving object tracking stage with the selected colour segment. Two different cases are considered in the experiments: simple colour (Figure 4(a)) and complex colour (Figure 4(b)) situations. In the simple colour case, the colour saliency of the target remains strong throughout the tracking, represented by the azure jumper shown in Figure 4(a). While in the complex colour case, the colour saliency of the target changes as colour of background interferes with the colour of the target continually, represented by the red jumper and black pants shown in Figure 4(b).

Key video frames of the tracking process for both cases are shown in Figures 5–8, in which the simple colour case is shown in Figures 5 and 6, while the complex colour case is shown in Figures 7 and 8. In these figures, cross represents the centroid position of the moving object, lines linked to the centroid show the paths of the moving target, the lengths of these lines indicate the moving speed of the target, the colour rendered area indicates the ROI of the moving object, and division lines for the "chopping" process are presented by the bounding boxes.

In the simple colour situation, the initial setting of the environment is shown in Figure 5(a), the background generation using IABMM is completed, and the robot is stationary. As a person walks into the scene, shown in Figure 5(b), object tracking using

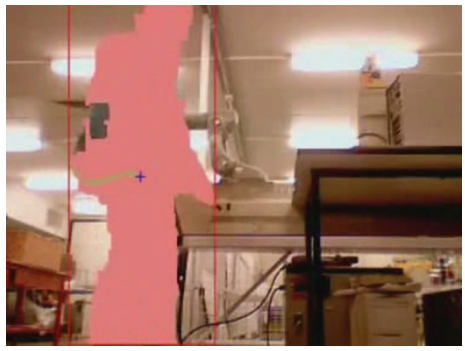
(a) Simple colour situation (azure)



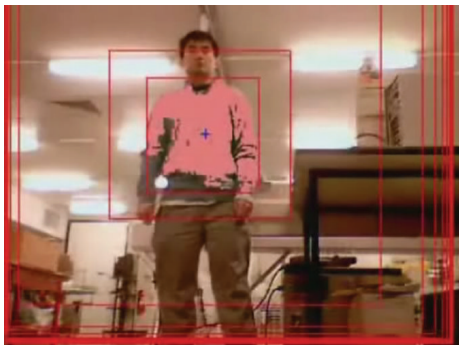
(b) Complex colour situation (red and black)

Figure 4: Two different cases used in the experiments.

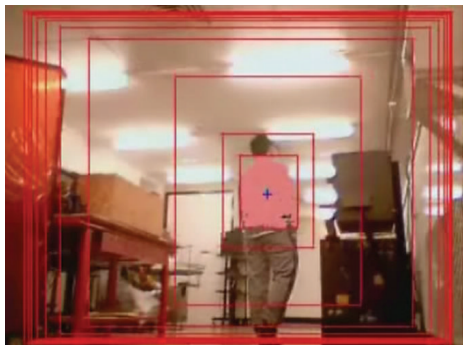
(a) Initial setting



(b) Background subtraction



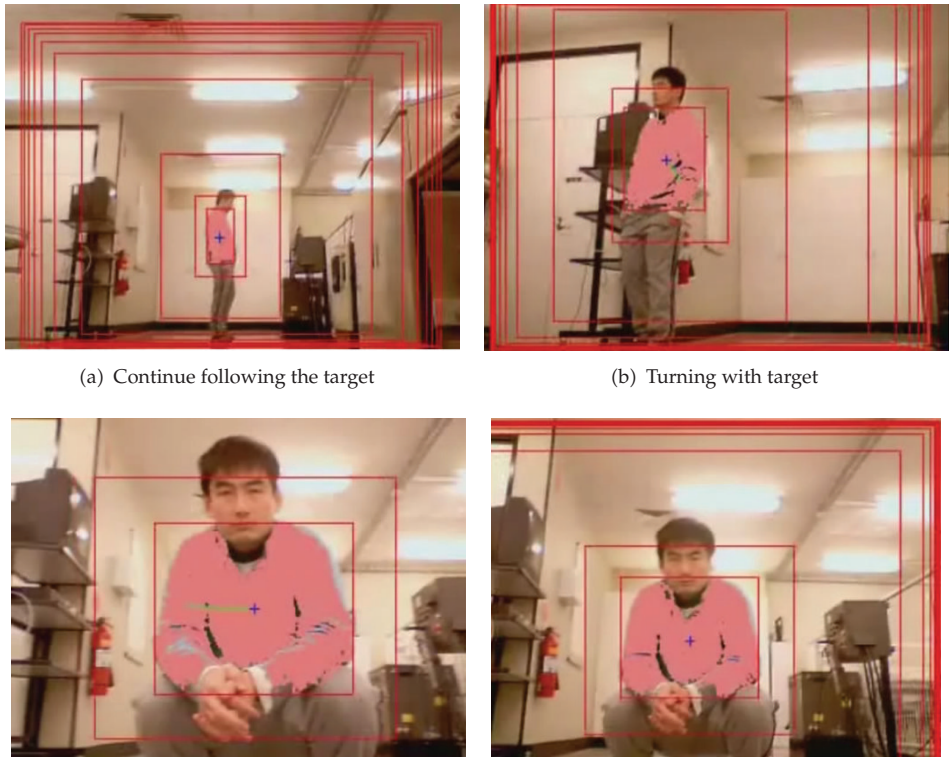
(c) Selecting and tracking the colour filter segment azure



(d) Following the target

Figure 5: Key video frames of moving object tracking in the simple colour situation part 1.

the background subtraction starts to track the person. By using the ROI identified from the background subtraction, the image is filtered with each of the 15 colour filter segments, and "azure" has the highest uniqueness value U. Therefore, it is chosen to be the most suitable colour feature for tracking shown in Figure 5(c). This is because the azure jumper of the person is the most salient colour feature. The robot then tracks and follows the target person as he moves randomly around in the laboratory, shown in Figures 5(d), 6(a), and 6(b). It can be seen that the robot is adjusting the distance between itself and the tracked person,



(a) Continue following the target

(b) Turning with target

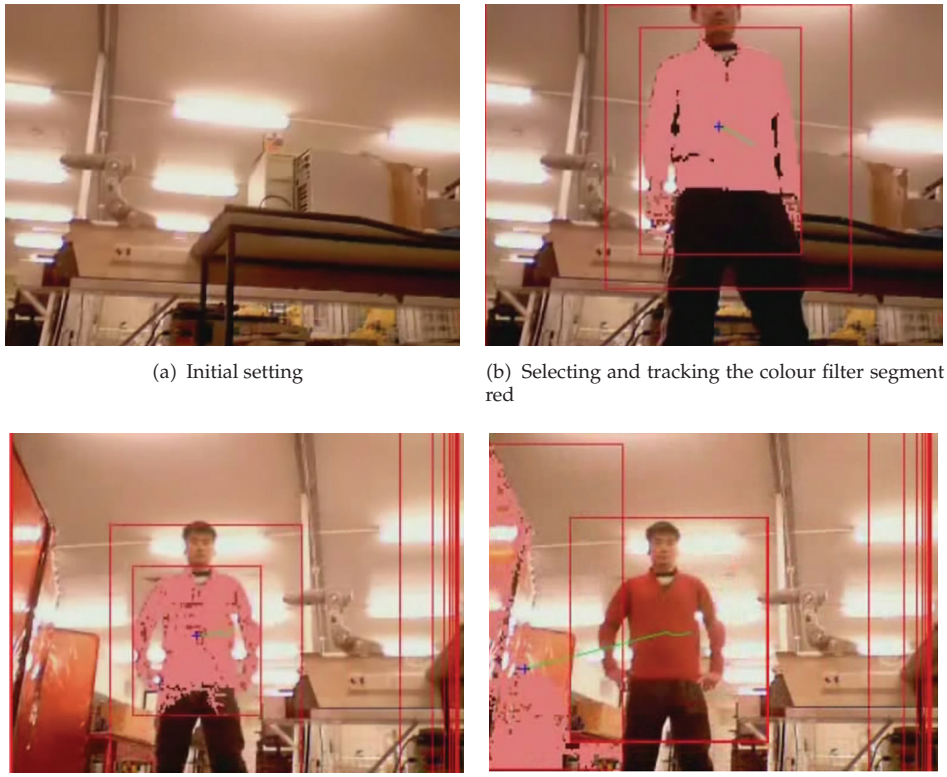
(c) Start moving backward

(d) Stop moving backward

Figure 6: Key video frames of moving object tracking in the simple colour situation part 2.

shown in Figures 6(c) and 6(d). During the tracking process, there is no reanalysis using the colour adaptation method introduced in Section 4. This is because the level of interference is relatively low in the azure colour segment as the colour saliency of the target remains high throughout the experiment.

In the complex colour situation, the initial setting of the environment is shown in Figure 7(a). As a person walks into the scene, by using the ROI identified from the background subtraction, the image is filtered and "red" is chosen to be the most suitable colour feature for tracking, as the red jumper of the person is the most salient colour feature in Figure 7(b). The robot then tracks and follows the target person as he moves around in the laboratory. As the robot follows the subject and drives into a background that shares the same colour, shown in Figure 7(c), a sudden jump in number of pixels sharing the same colour is detected, and the robot starts to reanalyse due to the adaptation method of the adaptive colour filter. This result is shown in Figure 7(d). The colour filter is updated and "black" is chosen to be the most salient colour segment for tracking. Thus, the black pants of the subject are tracked, shown in Figure 8(a). When the target person moves further down the laboratory, background interference occurs due to the television and its shelf that share the same colour, shown in Figure 8(b). Thus, another jump in number of pixels sharing the same colour occurred, shown in Figure 8(c). Due to the adaptation method introduced in (4.4), the robot starts to reanalyse. The colour filter is updated and "red" is once again chosen to be the most suitable colour feature for tracking, shown in Figure 8(d).



(a) Initial setting

(b) Selecting and tracking the colour filter segment red

(c) Following the target

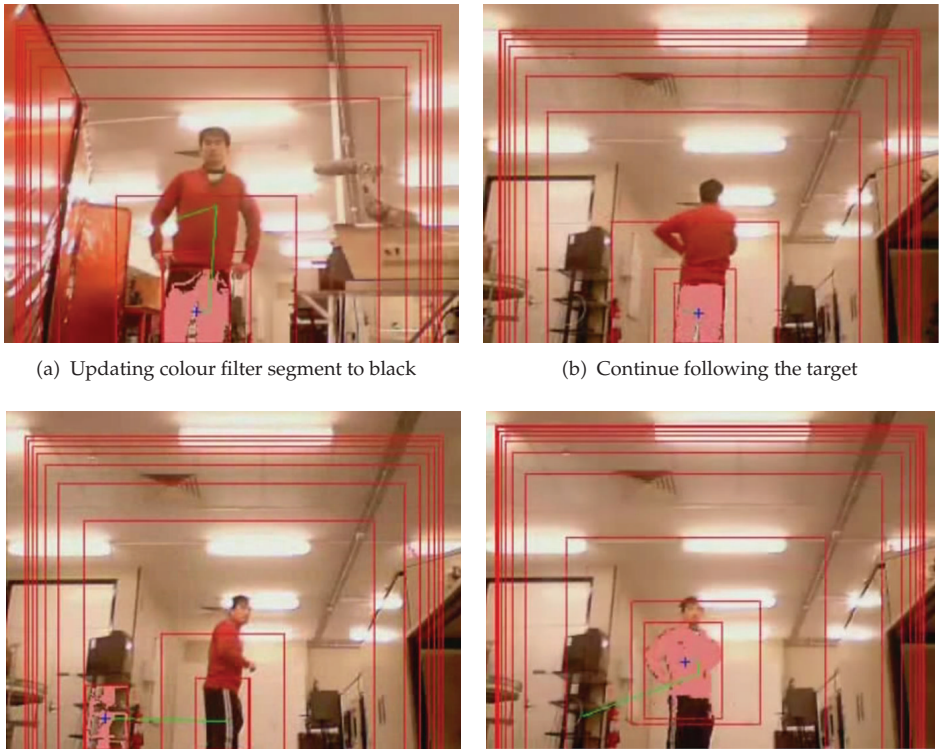
(d) Red interference detected

Figure 7: Key video frames of moving object tracking in the complex colour situation part 1.

These two cases (simple colour and complex colour cases) demonstrate that the proposed method can robustly and adaptively identify a salient colour in a moving object to track. It also shows, in the second case, that the proposed thresholds can adequately cope with the changing environment and keep tracking the correct object. In comparison with existing colour-based tracking methods, the advantage of the proposed method is its capability in identifying the most salient colour feature of the moving object with a moving camera mounted on a mobile robot. This method has the ability to adapt when the identified colour lost its saliency due to changing environment. However, although not common, there are situations where colour saliency is hard to identify, this can be the results of background and foreground objects sharing one uniform colour, such as a person wearing black clothing walking in a dark room. A possible approach to solve this is by introducing a secondary feature such as shape or corner.

6. Conclusions

In this research, a new adaptive colour feature identification method for real-time moving object tracking is presented. This adaptive colour tracking method is achieved by automatically determining a unique colour of the target object by comparing to the background and updating it when required. A method of determining the region of interest of the moving



(a) Updating colour filter segment to black

(b) Continue following the target

(c) Black interference detected

(d) Updating colour filter segment to red

Figure 8: Key video frames of moving object tracking in the complex colour situation part 2.

target is also introduced for the adaptive colour filter to extract objects of unique colour. Experimental results show that the methods developed are reliable of motion detection and moving object tracking in a busy/crowded indoor environment.

Acknowledgment

This work is partially supported by the Australian Research Council under project ID LP099-1108 and the Lincoln Electric Company Australia.

References

[1] V. Jupp, P. Davies, and P. Francis, *Doing Criminological Research*, SAGE, Thousand Oaks, Calif, USA, 1st edition, 2007.

[2] C. J. Cohen, K. A. Scott, M. J. Huber, S. C. Rowe, and F. Morelli, "Behavior recognition architecture for surveillance applications," in *Proceedings of the 37th IEEE Applied Imagery Pattern Recognition Workshop (AIPR '08)*, pp. 1–8, Washinton, DC, USA, October 2008.

[3] B. Liu, O. Jesorsky, and R. Kompe, "Robust real-time multiple object tracking in traffic scenes using an optical matrix range sensor," in *Proceedings of the 10th International IEEE Conference on Intelligent Transportation Systems (ITSC '07)*, pp. 742–747, Seattle, Wash, USA, October 2007.

[4] J. Zhu, Y. W. Lao, and Y. F. Zheng, "Object tracking in structured environments for video surveillance applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 2, pp. 223–235, 2010.

[5] X. G. Guo, C. H. Wang, and Z. S. Qu, "Object tracking for autonomous mobile robot based on feedback of monocular-vision," in *Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications (ICIEA '07)*, pp. 467–470, Harbin, China, May 2007.

[6] S. Y. Chen, J. W. Zhang, H. X. Zhang, N. M. Kwok, and Y. F. Li, "Intelligent lighting control for vision-based robotic manipulation," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 8, pp. 3254–3263, 2012.

[7] S. Y. Chen, Y. F. Li, and N. M. Kwok, "Active vision in robotic systems: a survey of recent developments," *International Journal of Robotics Research*, vol. 30, no. 11, pp. 1343–1377, 2011.

[8] S. Y. Chen, "Kalman filter for robot vision: a survey," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4409–4420, 2012.

[9] S. Y. Chen, "Active vision for robotic manipulation," *Industrial Robot*, vol. 39, no. 2, pp. 111–112, 2012.

[10] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.

[11] R. Yumiba, M. Miyoshi, and H. Fujiyoshi, "Moving object detection with background model based on spatio-temporal texture," in *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV '11)*, pp. 352–359, Kona, Hawaii, USA, January 2011.

[12] Z. Liu, K. Q. Huang, and T. N. Tian, "Cast shadow removal in a hierarchical manner using MRF," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 1, pp. 55–56, 2012.

[13] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, vol. 2, pp. 246–252, Fort Collins, Colo, USA, June 1999.

[14] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 2, pp. 28–31, Cambridge, UK, August 2004.

[15] P. KaewTraKulPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Proceedings of the European Workshop on Advanced Video Based Surveillance Systems*, pp. 1–5, London, UK, September 2001.

[16] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, pp. 172–185, 2005.

[17] C. H. Chen, C. Cheng, D. Page, A. Koschan, and M. Abidi, "A moving object tracked by a mobile robot with real-time obstacles avoidance capacity," in *Prceedings of the18th International Conference on Pattern Recognition (ICPR '06)*, pp. 1091–1094, Hong Kong, China, August 2006.

[18] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Computer Vision and Image Understanding*, vol. 113, no. 3, pp. 345–352, 2009.

[19] J. Han, G. Awad, and A. Sutherland, "Automatic skin segmentation and tracking in sign language recognition," *IET Computer Vision*, vol. 3, no. 1, pp. 24–35, 2009.

[20] Y. Wu and T. S. Huang, "Color tracking by transductive learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 133–138, Hilton Head, SC, USA, June 2000.

[21] T. Goncalves and A. I. Comport, "Real-time direct tracking of color images in the presence of illumination variation," in *Proceedings of the IEEE International Conference on Automation Science and Engineering*, pp. 4417–4422, Shanghai, China, May, 2011.

[22] J. Orwell, P. Remagnino, and G. A. Jones, "Multi-camera colour tracking," in *Proceedings of the IEEE Workshop on Visual Surveillance*, pp. 14–21, Fort Collins, Colo, USA, July 1999.

[23] I. Leichter, M. Lindenbaum, and E. Rivlin, "Tracking by affine kernel transformations using color and boundary cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 164–171, 2009.

[24] P. Li, "An adaptive binning color model for mean shift tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 9, pp. 1293–1299, 2008.

[25] J. Wang and Y. Yagi, "Integrating color and shape-texture features for adaptive real-time object tracking," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 235–240, 2008.

[26] S.-H. Lee and M. G. Kang, "Motion tracking based on area and level set weighted centroid shifting," *IET Computer Vision*, vol. 4, no. 2, pp. 73–84, 2010.

[27] T. Ojala, M. Pietikäinen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[28] T. Randen and J. H. Husøy, "Filtering for texture classification: a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291–310, 1999.

[29] P. Gorur and B. Amrutur, "Speeded up Gaussian mixture model algorithm for background subtraction," in *Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 386–391, Klagenfurt, Austria, August 2011.

[30] T. S. Huang, G. J. Yang, and G. Y. Tang, "A fast two-dimensional median filtering algorithm," *IEEE Trans Acoust Speech Signal Process*, vol. 27, no. 1, pp. 13–18, 1979.

[31] H. Sugano and R. Miyamoto, "Parallel implementation of morphological processing on cell/BE with OpenCV interface," in *Proceedings of the 3rd International Symposium on Communications, Control, and Signal Processing (ISCCSP '08)*, pp. 578–583, St. Julian's, Malta, March 2008.

[32] F. Chang, C. J. Chen, and C. J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Computer Vision and Image Understanding*, vol. 93, no. 2, pp. 206–220, 2004.

[33] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern Recognition*, vol. 34, no. 12, pp. 2259–2281, 2001.

[34] G. Fang and N. M. Kwok, "Image segmentation using adaptively selected color space," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '09)*, pp. 1838–1843, Guilin, China, December 2009.

[35] S. Sural, G. Qian, and S. Pramanik, "Segmentation and histogram generation using the HSV color space for image retrieval," in *Proceedings of the International Conference on Image Processing (ICIP '02)*, pp. 589–592, New York, NY, USA, September 2002.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 409856, 11 pages
doi:10.1155/2012/409856

Research Article

Modeling Transitions in Complex Systems by Multiplicative Effect of Temporal Patterns Extracted from Signal Flows

Ezzat G. Bakhoum[1] and Cristian Toma[2]

[1] *Department of Electrical and Computer Engineering, University of West Florida, 11000 University Parkway, Pensacola, FL 32514, USA*
[2] *Faculty of Applied Sciences, Politechnica University, Hagi-Ghita 81, 060032 Bucharest, Romania*

Correspondence should be addressed to Ezzat G. Bakhoum, ebakhoum@uwf.edu

Received 2 December 2012; Accepted 10 December 2012

Academic Editor: Carlo Cattani

This study presents a mathematical model based on Fourier decomposition of a sequence of internal signals generated in a complex system by a sequence of external pulses (time series) for characterizing suddenly emerging phenomena as nonlinear transitions. Newly created temporal patterns extracted from internal signal flow (mathematically represented as oscillations with long period) interact as new entities in a multiplicative manner with subsequent pulses from the external time series (already existing entities) in order to generate nonlinear transitions within the system. Such effects are enhanced when the period of external pulses creating new patterns is similar to the settling time of the complex system (this being the condition for an efficient external action). For complex systems where both classical and quantum phenomena generated by external time series are involved, this mathematical model can correctly explain the transition from classical to quantum behaviour (corresponding to a more ordered structure) avoiding typical contradictions generated by analysis performed on transient time intervals or by wave superposition.

1. Introduction

Nonlinear phenomena generated by an external time series represented by a sequence of pulses applied upon a complex system can be noticed not just for physical structures, but for biological and human systems also. For example, a set of medium-power shocks applied as transverse force upon a crystalline material fixed at both ends can generate significant deformations (possible breaking effect) which can not be always explained as a superposition of individual effects of each pulse. In a similar manner, transitions generated by periodical phenomena in biological systems cannot be analyzed using a linear model, since genetic mutations are often involved (for this reason genetic algorithms were developed). The human behavior is also driven by basic concepts created by repetitive aspects from the environment.

Some practical aspects were studied in [1], where it was mentioned that workers using traditional tools apply some medium-power shocks (External Time Series) at certain intervals upon a beam fixed at both ends, before the application of a final great-power shock for breaking the material. Each medium-power shock generates specific damping vibrations (Signal Flow) inside the material medium, and the subsequent shock has to be applied right before the annihilation of these damping vibrations by the fluctuations of the external medium (the Noise). Thus, a certain degree of Coherence for the effects of external pulses could be achieved, for the maximum possible value of the amplitude of the fundamental harmonic corresponding to the envelope of the generated vibrations. The mathematical model in this study uses the fundamental harmonic as a factor in a Coherence Function acting as a free term upon linear systems on extended time intervals. However, the practical aspects regarding breaking phenomena require a mathematical model able to generate a certain high intensity effect within a very short time interval, corresponding to just two or three external pulses.

Similar aspects can be noticed at High-Ordered Low Scale phenomena, by analyzing the wave functions with certain frequencies associated to particles represented as phonons (representing the quanta of the harmonic lattice vibrations for high-ordered complex materials). These wave functions can be represented as a sequence of pulses (a Time Series—each pulse corresponding to an oscillation); through interaction they generate transitions after a certain number of oscillations for the wave-trains of particles involved in interaction, as a Suddenly-Emerging Effect. However, quantum transitions cannot be considered as instantaneous phenomena, because the frequency (an important physical quantity in quantum dynamics theory) requires a certain time interval for an estimation. So it is important to determine some expected values for their frequency and for their corresponding energy not just in a stationary regime, but in a transient regime also (when supplementary vibrations are just generated).

Linear differential equations are not suitable for modelling such aspects. Better qualitative results were obtained using dynamical equations able to generate practical test functions [2] (similar to wavelets [3]) for generating significant delayed pulses (when a free term which corresponds to an external action is added) [4] and for justifying Spatial Patterns appearing in a certain material medium [5]. However, this model cannot explain the effect of a sequence of external pulses when the time interval between these pulses is large enough so as the final effect not be considered a superposition of individual effects of each pulse. For this reason, some specific differential equations based on the Coherence Function between the generated deformation and an alternating cosine input have been taken into consideration [1]. Since this coherence function vanishes if the output is equal to zero, the initial condition should be set at a small nonzero value (the choice of this initial value being not justified). This model has given good qualitative results for modelling the generation of oscillations with different local maximum/minimum values (for second-order differential equations) similar to wavelets analysis of solitary waves [6] and multiscale behaviour (for higher-order differential equations) similar to multi-scale behavior of waves in materials [7]. However, different initial values could lead to different temporal behavior, this being in contradiction with practical aspects of Suddenly-Emerging Phenomena. Moreover, the $\pi/2$ phase of the external alternating function $\sin(2\pi t + \phi) = \cos 2\pi t$ used in that mathematical model is not justified.

For the above reasons, an improved mathematical model based on the Fourier decomposition of internal signals generated in a system by a sequence of external pulses is introduced in this paper. The model is suited for characterizing Suddenly-Emerging

Phenomena as non-linear transitions in a newly created Quantum Environment, providing conditions for an effcient external action. This model should be also be able to explain specific features of quantum transitions on a very short (transient) time interval.

2. Carrier-Envelope Temporal Patterns Created by External Time Series

For a better understanding of basic features of the proposed mathematical model, we analyse the breaking effect of a set of medium-power shocks (External Time Series) applied as transverse force upon a beam (a crystalline structure-High-Ordered System) fixed at both ends when the time interval between these pulses is large enough cannot be explained as a superposition of individual effects of each pulse.

The effect of each pulse in the middle of the beam can be usually represented as

$$y(t) = \exp(-\alpha t)\sin(\omega_0 t + \phi) \tag{2.1}$$

which corresponds to the output of a damped second-order system (both α and ω_0 are considered to be positive quantities).

The envelope generated by a single pulse is represented by

$$f(t) = \exp(-\alpha t). \tag{2.2}$$

Since the effect of external shocks acting similar to Dirac functions upon the beam at certain time intervals consists in a specific deformation generated inside the material within a very short time interval, we will consider that the effect of different pulses cannot be superposed (the effect of each pulse is cancelled by next one). For a preliminary analysis, the temporal patterns generated by the long period envelope will be analyzed for a sequence of pulses with period set to unity for $\alpha = 5, \alpha = 1$ and $\alpha = 10$, respectively.

If the envelope is represented by

$$f(t) = \exp(-5t) \tag{2.3}$$

then the time interval beteween two succesive pulses will be approximately equal to the settling time of this function. The long period envelope will be represented by a periodical signal with a certain constant component and an alternating component, as in Figure 1.

If the time constant of the exponential function corresponding to the envelope is increased, so as the envelope generated by a single pulse to be represented by

$$f(t) = \exp(-t) \tag{2.4}$$

then the same time interval beteween two succesive pulses will be smaller than the settling time of the function. Thus the long period envelope will be represented by a sum of a greater constant component and a smaller alternating component as in previous case, as in Figure 2.

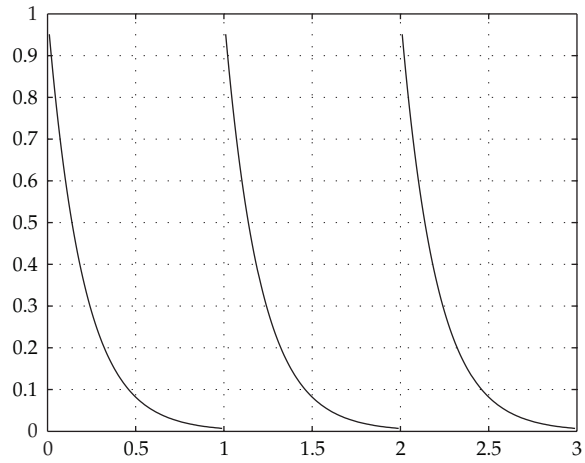


Figure 1: Long period envelope with Period ≈ Settling Time.

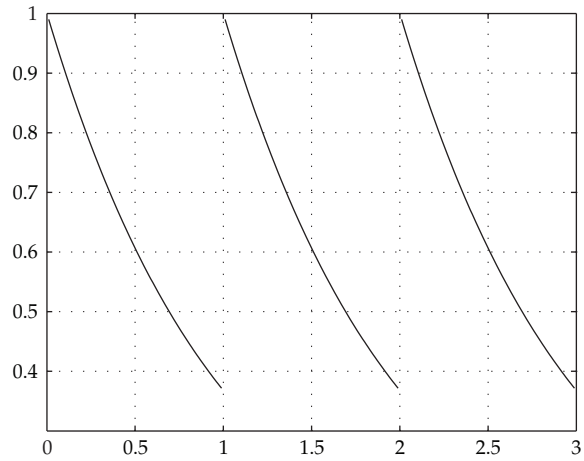


Figure 2: Long period envelope with Period ≪ Settling Time.

If the time constant of the exponential function corresponding to the envelope is decreased, so as the envelope generated by a single pulse to be represented by

$$f(t) = \exp(-10t) \tag{2.5}$$

then the same time interval beteween two succesive pulses will be greater than the settling time of the function. Thus the long period envelope will be represented by a sum of a smaller constant component and a smaller alternating component as in previous case since the function is approximately zero on large time intervals, as can be noticed in Figure 3.

It can be noticed from the very beginning that a greater alternating component can be obtained for the case when the settling time of the system is comparable to the period between two succesive external pulses.

A more rigorous analysis can be performed using the Fourier decomposition for the long period envelope function. For this purpose, we denote by T the period of

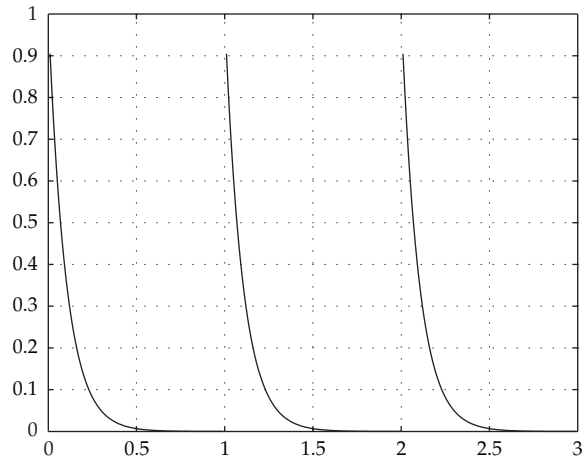


Figure 3: Long period envelope with Period ≫ Settling Time.

the external pulses. The analysis will be performed on the time interval $(0, 2\pi/T)$. The constant component does not represent any interest for modellimg non-linear effects. More important is the cosine component of angular frequency $\beta = 2\pi/T$, since transitions are expected to occur when another external shock (usually a great-power shock, as was shown) is applied—this means when the cosine function presents a maximum. The amplitude $A_{\cos}$ of this cosine component will be determined as

$$A_{\cos} = \frac{1}{T}\int_0^T \exp(-\alpha t)\cos(\beta t)\,dt = \frac{\beta}{2\pi}\,\mathrm{Re}\int_0^{2\pi/\beta} \exp\left[(-\alpha + i\beta)t\right]dt. \tag{2.6}$$

It results

$$A_{\cos} = \frac{\beta}{2\pi}\,\mathrm{Re}\left\{\frac{1}{-\alpha + i\beta}\exp\left[(-\alpha + i\beta)t\right]\Big|_0^{2\pi/\beta}\right\} \tag{2.7}$$

then

$$A_{\cos} = \frac{\beta}{2\pi}\,\mathrm{Re}\left\{\frac{-\alpha - i\beta}{\alpha^2 + \beta^2}\left(\exp\left[(-\alpha + i\beta)\frac{2\pi}{\beta}\right] - 1\right)\right\} \tag{2.8}$$

and finally

$$A_{\cos} = \frac{\alpha\beta}{2\pi(\alpha^2 + \beta^2)}\left(1 - \exp\left[\frac{-2\pi\alpha}{\beta}\right]\right). \tag{2.9}$$

By denoting the ratio between β and α as k and substituting $\beta = k\alpha$ in previous equation, it results after simplifying certain factors

$$A_{\cos} = \frac{k}{2\pi(1 + k^2)}\left(1 - \exp\left[\frac{-2\pi}{k}\right]\right). \tag{2.10}$$

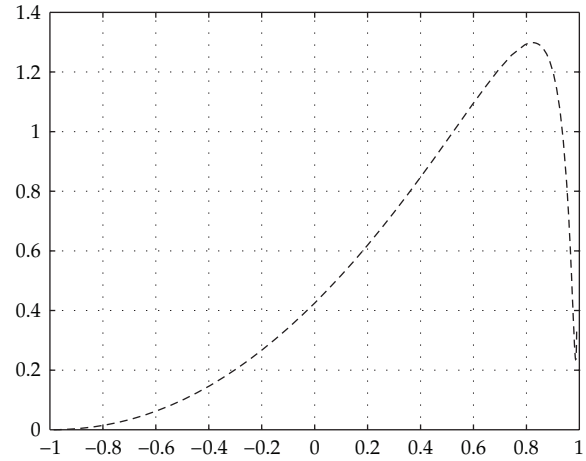


Figure 4: Amplitude of cosine function versus k.

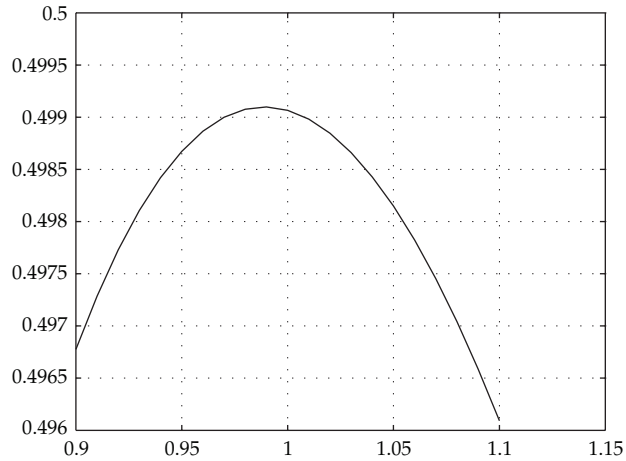


Figure 5: Amplitude of cosine function versus k-Details around $k = 1$.

This function has zero limit when k trends to zero and to ∞, presenting a maximum value for $k \approx 1$. The graph of this function is represented in Figure 4, for $k \in (0.5, 1.5)$ (the 2π factor at denominator was not taken into consideration).

It can be noticed that the maximum value is reached for $k \approx 1$. The value of $A_{\cos}$ varies with less than 10% for a wide range of $k(k \in (0.6, 1.5))$. A more refined graph is presented in Figure 5. It can be noticed that the value of $A_{\cos}$ varies with less than 0.6% for $(k \in (0.9, 1.1))$.

The maximum value of $A_{\cos} \approx 0.499/2\pi$ corresponds to $k \approx 0.98$ (the 2π factor has been omitted in previous figures). This result was expected, since the factor $k/(1 + k^2)$ (with a maximum for $k = 1$) is multiplied by $1 - \exp(-2\pi/k)$ (a function increasing from 0 to 1 with a small slope for a wide range of values). The physical meaning of this result consists in the fact that a high-amplitude cosine component is obtained when the time period T between two subsequent pulses (depending on β as $T = 2\pi\beta \approx 6.28\beta$) is approximately equal to the transient (settling) time T_{tr} of the damped second order system upon these pulses act

(T_{tr} depending on α as $T_{tr} = 4 \ldots 6\alpha$, according to basic dynamical aspects of second order linear systems). This means that short-period oscillations generated inside the system by an external pulse are not completely damped when another external pulse is applied (usually these short period oscillations are damped by friction phenomena or thermal interaction with the environment).

A slightly different result would have been obtained if the ratio k is adjusted for a maximum amplitude $A_{\sin}$ for the sine component of period T for the long period envelope. In a similar manner, it is determined as

$$A_{\sin} = \frac{1}{T}\int_0^T \exp(-\alpha t)\sin(\beta t)\,dt = \frac{\beta}{2\pi i}\,\mathrm{Im}\int_0^{2\pi/\beta} \exp\left[(-\alpha + i\beta)t\right]dt. \tag{2.11}$$

It results

$$A_{\sin} = \frac{\beta}{2\pi i}\,\mathrm{Im}\left\{\frac{1}{-\alpha + i\beta}\exp\left[(-\alpha + i\beta)t\right]\Big|_0^{2\pi/\beta}\right\} \tag{2.12}$$

then

$$A_{\sin} = \frac{\beta}{2\pi i}\,\mathrm{Im}\left\{\frac{-\alpha - i\beta}{\alpha^2 + \beta^2}\left(\exp\left[(-\alpha + i\beta)\frac{2\pi}{\beta}\right] - 1\right)\right\} \tag{2.13}$$

and finally

$$A_{\sin} = \frac{\beta^2}{2\pi(\alpha^2 + \beta^2)}\left(1 - \exp\left[\frac{-2\pi\alpha}{\beta}\right]\right). \tag{2.14}$$

Substituting $\beta = k\alpha$ in previous equation, it results

$$A_{\sin} = \frac{k^2}{2\pi(1 + k^2)}\left(1 - \exp\left[\frac{-2\pi}{k}\right]\right). \tag{2.15}$$

The ratio between $A_{\cos}$ and $A_{\sin}$ equals $1/k$. It corresponds to the phase Φ of the alternating compoment of period T as

$$\Phi = \arctan\frac{A_{\cos}}{A_{\sin}} = \frac{1}{k}. \tag{2.16}$$

For $k \approx 1$ (as previously determined) the ratio will be around unity, so $A_{\cos} \approx A_{\sin}$ and $\Phi \approx \pi/4$. The modulus A of the alternating component of period T can be determined as

$$A = \sqrt{A_{\cos}^2 + A_{\sin}^2} = \frac{k}{2\pi\sqrt{1 + k^2}}\left(1 - \exp\left[\frac{-2\pi}{k}\right]\right). \tag{2.17}$$

However, the transition created by this long period Carrier-Envelope Function is determined (according to our previous assumption) by cosine function, since it is generated around

zero moment of time (with time origin considered when an external pulse is applied). For this reason the value of k corresponding to the maximum for $A_{\cos}$ ($k \approx 1$) and an angular frequency $\beta \approx \alpha$ ($1/\alpha$ being the time constant of the system subject to external pulses) should be taken into consideration for modelling non-linear transitions. The cosine component of this angular frequency could be considered as a Multiplicative Factor for the intensity I_{pul} of an external pulse, generating a supplementary non-linear effect $E_{\text{non-lin}}$

$$E_{\text{non-lin}} = A_{\cos} I_{\text{pul}} \tag{2.18}$$

which is added to the linear effect E_{lin} of a single (isolated) pulse

$$E_{\text{lin}} = k I_{\text{pul}} \tag{2.19}$$

so as to result the global effect E of a subsequent pulse in a certain sequence as

$$E = E_{\text{lin}} + E_{\text{non-lin}} = k I_{\text{pul}} + A_{\cos} I_{\text{pul}} = (k + A_{\cos}) I_{\text{pul}}. \tag{2.20}$$

It can be noticed that this long period Carried-Envelope Function enhances a certain external pulse generating the transition, for example, the third pulse in a sequence is Multiplied by the envelope cosine component of period $T = 2\pi/\beta$ generated by the effect of two previous pulses. It results that an efficient external action consists in a set of external pulses (a time series) with period similar to the transient time of the complex system where transitions should be generated.

For more complex dynamical models based on Coherence Function where delayed effects and multiscale phenomena appear [1, 8], the Fourier decomposition presented in this paragraph justifies the amplitude of the alternating components generated by external pulses taken into consideration for determining the Coherence Function.

3. Temporal Patterns Created and Self-Organization by Fast-Varying External Signals in Low Scale Physical Systems: Modelling Transition to Quantum Behaviour

The multiplicative model of long period Carrier-Envelope Function is suitable for modelling and explaining a large class of phenomena in complex physical systems subject to fast-varying external signals when the selection of classical or quantum approach is not obvious for determining expected values through measurement procedures for certain physical quantities.

First we analyze some aspects connected with field emission for charged particles moving on repetitive closed-loop trajectories. When a beam of charged particles moves first time along such a trajectory, then for the first part of the movement (before the particle to return around its initial position) the trajectory should be considered as an open-loop trajectory. According to basic electromagnetic aspects, the particle should radiate energy. After a certain transient time, it can be considered that the particle is in a steady-state which should be described by quantum physics laws, and thus no signal will be emitted anymore. The multiplicative model of Carrier-Envelope Pattern is recommended for modelling the particle behaviour on this transient time, determining expected values for measurement

procedures in a more rigorous manner. It requires just a few periodical movements in order for a certain amplitude of an alternating component to be determined—then it is multiplied by the potential energy so as a specific term to be added in Schroedinger equation (the need of external fields generating potential energy has been shown in [9] due to the paradox of phase loss for a quantum particle moving in a quantified electrostatic potential). Thus the system achieves a high-ordered structure by itself- through Self Organization based on hiding dynamics and internal memory (see [10, 11] for more details about learning hiding dynamics and [12] for systems with memory at biological level of organization). In a certain point of space M_S, we can consider that after a few periodical movements (when classical laws for observable physical quantities are valid) the set of waves emitted by the particle from different points of space S_i along the steady-state trajectory will create also a very low-amplitude Temporal Pattern (a low-amplitude Carrier-Envelope Function), since the phase of received waves presents a circular uniform distribution (practically the coherent waves cancel each other). Thus no emitted field can be detected when the particle is in a steady-state (when the transient time has passed), according to quantum physics laws. This aspect differs to reception of continuous signals based on well-known detection of Carrier-Envelope Functions, where non-linear signal processing is achieved trough additional devices (rectifier diodes, threshold circuits, memories). The use of advanced mathematical aspects as multifractional formula [13], cyber-physical laws [14] and stochastic bound [15] imply also the use of such additional devices for memory and decision.

In a similar manner, when the particle performs a transition from a steady-state to another, a wave train corresponding to a photon is emitted. The wave-train corresponding to this emitted photon is detected by a certain material medium as a non-linear phenomenon, some alternances corresponding to the photon creating a Temporal Pattern enhancing the effect of subsequent alternances and generating the quantum interaction with the receiving material medium.

Some basic features of this model can be extended to analysis of light signals propagating in fluctuating weak gravitational fields. As was shown in [16], in every point in a space, at a certain moment of time should be defined a certain material reference system which acts upon a received light wave-train through Lorentz transformation (thus determinig expected values for space-time measurements based on light wave-trains). In case of vacuum, the last material medium that light has passed through should be taken into consideration. When the light passes near great material bodies, it interacts with the significant gravitational field of that body and thus the reference system of this material body should be taken into consideration. Yet in the limit case, when the gravitational field is approximately zero, some fluctuations created by different material bodies still exist. So the propagating wave can be considered as subject to a great number of Lorentz transformations which could scatter the wave, and the wave should vanish within a short space-time interval (considered at the scale of cosmic systems). Applying the Carrier-Envelope Model, it results that the stochastic low-amplitude variations of the gravitational field around the null value creates a null temporal pattern which does not allow an influence upon a propagating wave-train (low-amplitude vectors with different directions cancel each other). This is in accordance with the fact that, according to Conformal Field Theory, all quantum field theories are only effective over some range of length and energy scales, and those used in particle physics are no more fundamental than in condensed matter.

Dynamical relativistic aspects can be also justified by a Multiplicative Model. The factor $1-v^2/c^2$ appearing at denominator in mass formula can be obtained just by multiplying $(c-v)/c$ and $(c+v)/c$—as two functions are multiplied. For a movement along Ox axis, it

can be easily understand that when v equals c no interaction between wave and particle occur and thus the mass seems to be infinite (no result of the wave action upon the particle being noticed, so the particle seems to be out of the range of any possible external command). However, according to well-known mass formula the same result is valid when $v = -c$, this means when the particle is moving with light speed c in the opposite direction as related to the received wave. This aspect cannot be explained as lack of interaction between wave and particle (or, generally speaking, as lack of interaction between an external command and the system upon this command acts), and thus a model based on a multiplication of certain functions is necessary. In an intutive manner, these two functions can be put in correspondence with a direct and a reverse wave which interact in the point of space where the particle with mass m is situated (the system and the external command interacts in a more complex manner, as in a sequence action-reaction defined on a very short time interval).

4. Conclusions

This study has presented a mathematical model based on Fourier decomposition of internal signals created in a complex environment by a sequence of external pulses for justifying suddenly-emerging phenomena as non-linear transitions. Newly created temporal patterns (mathematically represented as long period oscillations) interact in a multiplicative manner (as a Carrier-Envelope Function) with subsequent pulses, generating supplementary effects which can be noticed both at macroscopic and quantum (microscopic) scale, for physical, biological and human systems. Such effects are enhanced when the period of external pulses is similar to the settling time for the environment, justifiying the amplitude of alternating components used in more elaborated mathematical models based on Coherence Function for simulating suddenly-emerging phenomena. Adding phase aspects, this mathematical model can explain transitions between classical and quantum behaviour in a complex material environment in a correct manner, avoiding contradictions generated by analysis performed on transient time intervals or by wave superposition.

References

[1] E. G. Bakhoum and C. Toma, "Dynamical aspects of macroscopic and quantum transitions due to coherence function and time series events," *Mathematical Problems in Engineering*, vol. 2010, Article ID 428903, 13 pages, 2010.

[2] G. Toma, "Practical test-functions generated by computer algorithms," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '05)*, vol. 3482 of *Lecture Notes in Computer Science*, pp. 576–584, May 2005.

[3] C. Cattani, "Harmonic wavelets towards the solution of nonlinear PDE," *Computers and Mathematics with Applications*, vol. 50, no. 8-9, pp. 1191–1210, 2005.

[4] B. Lazar, A. Sterian, St. Pusca et al., "Simulating delayed pulses in organic materials," in *Proceedings of the Computational Science and Its Applications (ICCSA '06)*, vol. 3980 of *Lecture Notes in Computer Science*, pp. 779–785, 2006.

[5] F. Doboga, "Different structural patterns created by short range variations of internal parameters," in *Proceedings of the 7th international conference on Computational Science (ICCS '07)*, vol. 4488 of *Lecture Notes in Computer Science*, pp. 1060–1066, 2007.

[6] J. J. Rushchitsky, C. Cattani, and E. V. Terletskaya, "Wavelet analysis of the evolution of a solitary wave in a composite material," *International Applied Mechanics*, vol. 40, no. 3, pp. 311–318, 2004.

[7] C. Cattani, "Multiscale analysis of wave propagation in composite materials," *Mathematical Modelling and Analysis*, vol. 8, no. 4, pp. 267–282, 2003.

[8] E. G. Bakhoum and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.

[9] E. G. Bakhoum and C. Toma, "Mathematical transform of traveling-wave equations and phase aspects of quantum interaction," *Mathematical Problems in Engineering*, vol. 2010, Article ID 695208, 15 pages, 2010.

[10] C. Cattani and A. Ciancio, "Hybrid two scales mathematical tools for active particles modelling complex systems with learning hiding dynamics," *Mathematical Models and Methods in Applied Sciences*, vol. 17, no. 2, pp. 171–187, 2007.

[11] C. Cattani, A. Ciancio, and A. d'Onofrio, "Metamodeling the learning-hiding competition between tumours and the immune system: a kinematic approach," *Mathematical and Computer Modelling*, vol. 52, no. 1-2, pp. 62–69, 2010.

[12] A. Ciancio, "An approximate evaluation of the phenomenological and state coefficients for visco-anelastic media with memory," *U.P.B. Scientific Bulletin A*, vol. 73, no. 4, pp. 3–14, 2011.

[13] M. Li and W. Zhao, "Quantitatively investigating locally weak stationarity of modified multifractional Gaussian noise," *Physica A*, vol. 391, no. 24, pp. 6268–6278, 2012.

[14] M. Li and W. Zhao, "Visiting power laws in cyber-physical networking systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 302786, 13 pages, 2012.

[15] M. Li and W. Zhao, "Representation of a stochastic traffic bound," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 9, pp. 1368–1372, 2010.

[16] E. G. Bakhoum and C. Toma, "Relativistic short range phenomena and space-time aspects of pulse measurements," *Mathematical Problems in Engineering*, vol. 2008, Article ID 410156, 20 pages, 2008.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 671397, 13 pages
doi:10.1155/2012/671397

Research Article

Contextual Hierarchical Part-Driven Conditional Random Field Model for Object Category Detection

Lizhen Wu, Yifeng Niu, and Lincheng Shen

College of Mechatronics Engineering and Automation, National University of Defense Technology, Changsha 410073, China

Correspondence should be addressed to Lizhen Wu, lzwu@nudt.edu.cn

Received 25 October 2012; Accepted 11 November 2012

Academic Editor: Sheng-yong Chen

Even though several promising approaches have been proposed in the literature, generic category-level object detection is still challenging due to high intraclass variability and ambiguity in the appearance among different object instances. From the view of constructing object models, the balance between flexibility and discrimination must be taken into consideration. Motivated by these demands, we propose a novel contextual hierarchical part-driven conditional random field (CRF) model, which is based on not only individual object part appearance but also model contextual interactions of the parts simultaneously. By using a latent two-layer hierarchical formulation of labels and a weighted neighborhood structure, the model can effectively encode the dependencies among object parts. Meanwhile, beta-stable local features are introduced as observed data to ensure the discriminative and robustness of part description. The object category detection problem can be solved in a probabilistic framework using a supervised learning method based on maximum a posteriori (MAP) estimation. The benefits of the proposed model are demonstrated on the standard dataset and satellite images.

1. Introduction

Object category detection is one of the most important problems in computer vision and is still full of challenges because of various factors such as object deformation, occlusion, and viewpoint change. To address these challenges, successful object detection methods need to strike the balance between being flexible enough to model intraclass variability and being discriminative enough to find objects with ambiguity appearance in complicate scenes [1–3].

Part-based object model, firstly proposed by Fischler and Elschlager [4] in 1973, has been proved as a powerful paradigm for object category detection and recognition in numerous researches [5–10], due to its advantages of intuitive interpretation and semantic expression. In such models, each part is generally represented by small templates or local

image feature information, and the whole object is modeled as a collection of parts with or without geometric and cooccurrence constraints. The final discriminate of object is achieved by solving the probability density function or using a Hough vote mechanism. In the early researches on part-based approaches, parts are learned purely on the basis of their appearance by clustering visually similar image patches in the training images and do not exploit any spatial layout of the parts. Obviously, since the part appearance only reflects local image characteristics, these models cannot get enough spatial information support. The neglected contextual interactions that are used to capture geometric relationships between parts of an object should play a more crucial role in the part-based model to enhance the representational power of model.

On the other hand, most current part-based approaches can be roughly divided into two separate groups: generative and discriminative. Generative part-based models [5–9] have shown high flexibility because of their advantage of handling missing data (i.e., the correspondence between local features and parts) in a principled manner. So, each part can be interpreted in a semantically meaningful way. The most popular generative approach for part-based object detection was proposed by Fergus et al. [7] in 2003, in which objects are modeled as flexible constellations of parts and the appearance, spatial relations, and cooccurrence of local parts are learned in an unsupervised manner. Felzenszwalb and Huttenlocher [8] proposed a pictorial structure model, in which deformable configuration is represented by spring-like connections between pairs of parts. By integrating spatial relationships with "bag of features," Sudderth et al. [9] developed a hierarchical probabilistic model to capture the complex structure in multiple object scenes. However, generative approaches often cannot compete with discriminative manner in the field of object category detection. The generative framework has natural drawback that it has to assume the independence of the observed data to make the model computationally. In contrast to the discriminative model, the generative model may be quite complex even though the class posterior is simple. Moreover, learning the class density models may become even harder when the training data is limited [10].

In this paper, we focus on the discriminative random field model, called conditional random field (CRF), which is originally proposed by Lafferty et al. [11] in 2001. Kumar and Herbert [12, 13] first introduced the extension of 1D CRFs to 2D graphs over image and applied it to object detection. By treating object detection problem as a labeling problem, CRF model cannot only flexibly utilize various heuristic image features, but also get the contextual interactions among image parts through its classic graphical structure. In order to deal with multiple labels for object parts, Kumar and Hebert presented a multiclass extension of CRF [14], and utilized fully labeled data where each object part is assigned a part label to train the model. By contrast, Quattoni et al. presented an expansion graph structure of CRF framework [15] that uses hidden variables, which are not observed during training, to represent the assignment of parts. Moreover, located CRF model [16], proposed by Kapoor and Winn, introduces global positions to the hidden variables and can model the long-range spatial configuration and local interactions simultaneously.

The goal of this paper is to introduce a novel contextual hierarchical part-driven CRF model for object category detection. The main novelty of our approaches lies in the use of a latent two-layer hierarchical formulation of labels and a weighted minimum spanning tree neighborhood structure. The model can effectively encode latent label-level context, as well as observation-level context. Meanwhile, beta-stable local features are also introduced as observed data to ensure the discriminative and robustness of part description. Such features

provide a sparse and repeatable manner to express object parts and actually reduce the computation complexity of the model.

The remainder of this paper is organized as follows. Section 2 gives detailed introduction on the proposed contextual hierarchical part-driven CRF Model. The parameter learning and inference algorithms are introduced in Section 3. Experimental results are presented in Section 4. Finally, in Section 5 we draw the conclusions.

2. Contextual Hierarchical Part-Driven CRF Model

The conditional random field is simply a Markov random field (MRF) [17] globally conditioned on the observation. It is a discriminative model that relaxes conditional independence assumption by directly estimating the conditional probability of labels [18, 19].

In other words, let y be the observed data from an input image, where $y = \{y_i\},\ i \in S$, y_i is the data from the ith site, and S is the set of sites. The corresponding labels at image sites are given by $x = \{x_i\},\ i \in S$. For labeling problems, the general form of a CRF can be written as

$$\begin{aligned} P(x \mid y, \theta) &= \frac{1}{Z(\theta)} \exp\{\Phi(x, y, \theta)\} \\ &= \frac{1}{Z(\theta)} \prod_{i \in S} \varphi_i(x_i, y, \theta) \prod_{(i,j) \in E} \varphi_{ij}(x_i, x_j, y, \theta), \end{aligned} \tag{2.1}$$

where partition function $Z(\theta)$ is a constant normalization with respect to all possible values of x with parameters θ, E denotes the set of edges, and φ_i and φ_{ij} are the unary and pairwise potentials, respectively. Here, φ_i encodes compatibility of the label x_i with the observed image y and φ_{ij} encodes the pairwise label compatibility for all $(i, j) \in E$ that $j \in N_i$ conditioned on y.

2.1. Problem Formulation

For our object category detection problem, assume that we are given a training set of N images $Y = (y^1, \ldots, y^N)$, which contains objects from a particular class and background images. The corresponding labels can be denoted as $X = (x^1, \ldots, x^N)$, each x^n is a member of a set of possible image labels. Since in object detection we only focus on presence or absence of objects, the possible labels should be limited to binary data, that is, $x^n \in \{0, 1\}$ or $x^n \in \{\text{background, object}\}$. Now, our task is to learn a mapping from images Y to labels X. For simplicity of notation, we drop the superscript n indicating training instance.

According to the theory of part-based model, assume that each image y can be seen as a collection of parts $y = (y_1, \ldots, y_m)$, each part y_i corresponds to a local observation or local feature. In order to describe the relationship between these parts, similar to hidden random field approach [15], we introduce latent labels $h = (h_1, \ldots, h_m),\ h_i \in H$, where h_i corresponds to the "part-label" of part y_i, and H corresponds to the actually object parts, for example, $H = \{\text{nose, tail}, \ldots, \text{wing}\}$ for airplane objects.

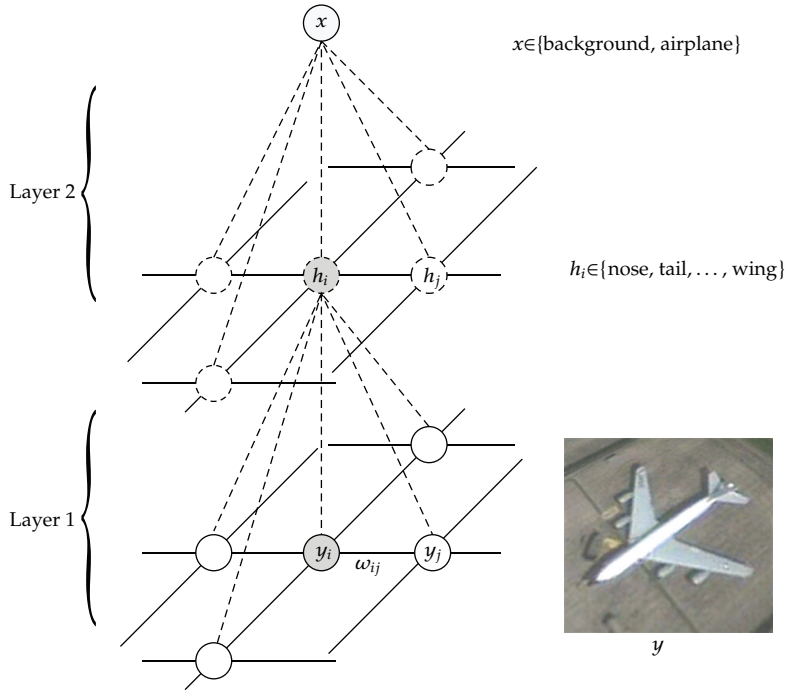


Figure 1: The hierarchical graphical structure of proposed contextual hierarchical part-driven CRF model.

Now, we can model the posterior directly by marginalizing out the latent labels h, and the model can be defined as

$$P(x \mid y, \theta) = \sum_{h} P(x, h \mid y, \theta) = \sum_{h} \underbrace{P(x \mid h, \kappa)}_{\text{Layer 2}} \underbrace{P(h \mid y, \lambda)}_{\text{Layer 1}}, \tag{2.2}$$

where $\theta = \{\kappa, \lambda\}$ is the set of parameters.

Here, we assume that $P(x \mid h, \theta)$ is conditional independence of y given h. This means that the final object label only relies on the latent middle-level labels rather than on the original observations. The hypothesis makes sense because it is theoretically possible that we estimate the object or background occurrence by the spatial distribution of meaningful object parts in real world. Also, by doing this, we can build distinct two-layer structure of our proposed model. The hierarchical graphical structure of our contextual hierarchical part-driven CRF model is shown as Figure 1.

Obviously, both $P(x \mid h, \kappa)$ and $P(h \mid y, \lambda)$ can be modeled as CRFs. Thus, the whole model can be seen as the combination of two single-layer CRFs. By modeling the contextual interactions of these two layers, respectively, our model may have a high level of ability to describe different levels of context. Note that the latent label h cannot be observed during training (i.e., unlabelled), so we must learn the models in a unified framework to avoid the direct use of h. Detailed modeling approach and potential definitions for the two layers will be described below.

2.2. Model of Layer 1

Without considering the object label x, the distribution over the latent part labels h given the observations y may be modeled as a multiclass CRF. In such a model, observations are linked to local features located at certain spatial positions of the image. Therefore, the distribution of y may be arbitrary and disorganized due to the uncertainty of local feature extraction. Meanwhile, different observations may be associated with the same part label, which corresponded to the meaningful object component. Due to the fact that adjacent or relevant observations are more likely to have the same label, we should consider label-level context in our model in addition to observation-level context.

Furthermore, although we cannot use the part labels explicitly, we can theoretically use them to define the posterior to capture the context structure of layer 1. Considering only unary and pairwise potentials, the posterior distribution $P(h \mid y, \lambda)$ can be modeled as

$$P(h \mid y, \lambda) = \frac{1}{Z(\lambda)} \prod_{i \in S} \varphi_i^{(1)}(h_i, y, \mu) \prod_{(i,j) \in E} \varphi_{ij}^{(1)}(h_i, h_j, y, \nu), \tag{2.3}$$

where the set of parameters is given by $\lambda = \{\mu, \nu\}$, as shown in Figures 2(a) and 2(b), $\varphi_i^{(1)}(h_i, y, \mu)$ denote the unary potentials and are responsible for modeling part occurrences based on a single image feature, $\varphi_{ij}^{(1)}(h_i, h_j, y, \nu)$ denote the pairwise potentials and are responsible for modeling the cooccurrences of h_i and h_j based on the corresponding pairwise image feature. The connectivity of nodes (i, j), that is, neighborhood structure of the observations, is defined in Section 2.4.

Note that, different from multiclass CRF [14], the definitions of potentials here must consider the missing label data h. By using the parameter vector, the potentials can be denoted as

$$\varphi_i^{(1)}(h_i, y, \mu) = \exp\left[\mu(h_i)^T f_i(y)\right], \tag{2.4}$$

$$\varphi_{ij}^{(1)}(h_i, h_j, y, \nu) = \exp\left[\nu(h_i, h_j)^T g_{ij}(y)\right], \tag{2.5}$$

where $f_i(y)$ and $g_{ij}(y)$ refer to the unary feature vector and the pairwise feature vector, respectively. Parameter vectors of $\mu(h_i) \in \Re^d$ and $\nu(h_i, h_j) \in \Re^B$ have the same dimensions with the corresponding feature vectors.

In comparison with hidden CRF [15], our approach introduces the pairwise potentials $\varphi_{ij}^{(1)}$ to the model and can effectively capture the part label-level context by measuring the compatibility of different part labels.

2.3. Model of Layer 2

According to the conditional independence assumption mentioned in Section 2.1, the final image label x only depends on part labels h. In other words, the occurrence of an object can be estimated by the spatial distribution of object parts.

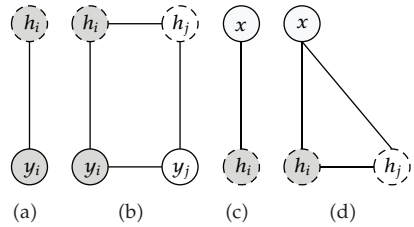


Figure 2: (a) Part evidence from single observation, (b) cooccurrence of connected parts, (c) compatibility between image label and single part label, and (d) compatibility between image label and connected part labels.

Particularly, part labels h should be regarded as observations in this layer, and the posterior distribution $P(x \mid h, \kappa)$ can be easily defined as

$$P(x \mid h, \kappa) = \frac{1}{Z(\kappa)} \prod_{i \in S} \varphi_i^{(2)}(h_i, x, \alpha) \prod_{(i,j) \in E} \varphi_{ij}^{(2)}(h_i, h_j, x, \gamma), \tag{2.6}$$

where $\kappa = \{\alpha, \gamma\}$ is the set of parameters, unary potentials $\varphi_i^{(2)}(h_i, x, \alpha)$ describe the compatibility between image label x and part label h_i, pairwise potentials $\varphi_{ij}^{(2)}(h_i, h_j, x, \gamma)$ describe the compatibility between image label x, part label h_i, and part label h_j, as in Figures 2(c) and 2(d).

Note that there is only one image label for an instance, so we do not need to model label-level context like in layer 1, and the potentials can be defined as

$$\varphi_i^{(2)}(h_i, x, \alpha) = \exp[\alpha(h_i, x)], \tag{2.7}$$

$$\varphi_{ij}^{(2)}(h_i, h_j, x, \gamma) = \exp[\gamma(h_i, h_j, x)], \tag{2.8}$$

where parameter vectors $\alpha(h_i, x) \in \Re$ and $\gamma(h_i, h_j, x) \in \Re$.

Now, we can give the complete expression of our part-driven CRF model with the specific potentials, which can be denoted as

$$\begin{aligned} P(x \mid y, \theta) &= \sum_h P(x \mid h, \kappa) P(h \mid y, \lambda) \\ &= \sum_h \left\{ \frac{1}{Z(\theta)} \prod_{i \in S} \varphi_i^{(1)}(h_i, y, \mu) \varphi_i^{(2)}(h_i, x, \alpha) \prod_{(i,j) \in E} \varphi_{ij}^{(1)}(h_i, h_j, y, \nu) \varphi_{ij}^{(2)}(h_i, h_j, x, \gamma) \right\}, \end{aligned} \tag{2.9}$$

where the set of parameters is given by $\theta = \{\mu, \alpha, \nu, \gamma\}$.

2.4. Neighborhood Structure

For probabilistic graphical models, the neighborhood structure is an important factor affecting the model capability. Moreover, as mentioned above, the observations y in our model are distributed over the image plane in an arbitrary layout. So, how to define the neighborhood structure becomes a question we have to consider during the model design phase.

In [15], Quattoni et al. evaluated a range of different neighborhood structure and come to the conclusion that the minimum spanning tree (MST) shows better performances than many other complex connected graph structures. Following this, we adopt MST as the basic structure and extend it to a novel weighted neighborhood structure (WNS). The basic idea is to exploit the edge cost, which is discarded in previous work, as heuristic information to reflect the degree of correlation between two nodes.

In other words, different edges should have different weights during calculating pairwise potentials. We denote the edge cost between node i and node j as ω_{ij} and modify (2.5) to

$$\varphi_{ij}^{(1)}(h_i, h_j, y, v) = \exp\left[\omega_{ij} \cdot v(h_i, h_j)^T g_{ij}(y)\right]. \tag{2.10}$$

Similarly, (2.8) is changed by

$$\varphi_{ij}^{(2)}(h_i, h_j, x, \gamma) = \exp\left[\omega_{ij} \cdot \gamma(h_i, h_j, x)\right]. \tag{2.11}$$

By doing this, the weighted neighborhood structure can not only describe the connectivity of nodes, but also encode the assumption that parts that are spatially close are more likely to be dependent.

3. Parameter Learning and Inference

Given N-labeled training images, the parameters $\theta = \{\mu, \alpha, v, \gamma\}$ can be learnt by using maximum A posteriori (MAP) estimation. Gaussian prior $p(\theta) \sim \exp(\|\theta\|^2/2\sigma^2)$ is introduced to prevent overfitting. So, parameter learning can be achieved by maximizing the following objective function:

$$L(\theta) = \sum_{n=1}^{N} \log P\left(x^{n'} \mid y^n, \theta\right) - \frac{1}{2\sigma^2}\|\theta\|_2. \tag{3.1}$$

We use gradient ascent to search for the optimal parameter values $\theta^* = \arg\max_\theta L(\theta)$. In our model, the derivatives of the log-likelihood $L(\theta)$ with respect to the model parameters

$\theta = \{\mu, \alpha, \nu, \gamma\}$ can be written in terms of local feature vectors, marginal distributions over individual part label h_i, and marginal distributions over pairwise labels h_i and h_j:

$$\begin{aligned}
\frac{\delta L(\theta)}{\delta \mu(h')} &= \sum_{i \in s} f_i(y) \cdot \left[p(h_i = h' \mid x, y, \theta) - p(h_i = h' \mid y, \theta)\right], \\
\frac{\delta L(\theta)}{\delta \alpha(h', x')} &= \sum_{i \in s} p(h_i = h' \mid x, y, \theta) - p(h_i = h', x' \mid y, \theta), \\
\frac{\delta L(\theta)}{\delta \nu(h', h'')} &= \sum_{(i,j) \in E} \omega_{ij} \cdot g_{ij}(y) \cdot \left[p(h_i = h', h_j = h'' \mid x, y, \theta) - p(h_i = h', h_j = h'' \mid y, \theta)\right], \\
\frac{\delta L(\theta)}{\delta \gamma(h', h'', x')} &= \sum_{(i,j) \in E} \omega_{ij} \cdot \left[p(h_i = h', h_j = h'' \mid x, y, \theta) - p(h_i = h', h_j = h'', x' \mid y, \theta)\right].
\end{aligned} \tag{3.2}$$

Note that, all the terms in the derivatives can be calculated using Belief Propagation (BP) algorithm [17], provided the graphical structure does not contain cycles. Otherwise, approximate methods, such as loopy BP could be considered. Here, BP is suitable for our case due to the tree-like neighborhood structure.

For the final class inference, we need to find the image label $\hat{x}$ that maximizes the conditional distribution $(x \mid y, \theta)$, given parameters θ^*. For this work, we can also use the max-product version of BP to find the MAP estimate $\hat{x} = \arg\max_x P(x \mid y, \theta^*)$.

4. Experiments

In this section, we demonstrate the capability of the proposed model on two different datasets: Caltech-4 standard dataset and airplane images collected from Google Earth. The aim of these experiments is to illustrate the performance of this detection framework using contextual hierarchical part-driven CRF model and compare with state-of-the-art models.

4.1. Image Features

For the object category detection task, the robustness and distinctiveness are basic requirements for local features in order to provide powerful expression ability for objects. On the other hand, the quantity of features corresponds to the quantity of observations and has an enormous influence on computational complexity. Taking these aspects into consideration, we should try to use the local features which have the characteristics of sparse, robust, and discriminative simultaneously.

In our experiments, we use the beta-stable feature extracting method [20] to locate local features and SIFT descriptor [21] to construct feature vectors. Rather than selecting features that persist over a wide interval of scales, beta-stable features are chosen at a scale so that the number of convex and concave regions of the image brightness function remains constant within a scale interval of length beta. As a result, the beta-stable features have stronger robustness than SIFT-like features and are better anchored to visually significant parts. The comparative feature-points detecting results are shown in Figures 3(a) and 3(b).

(a) (b) (c)

Figure 3: (a) SIFT feature detecting, (b) beta-stable feature detecting, (c) beta-stable features connected by MST.

The unary feature vector, $f_i(y)$, used in this work is represented by the combination of SIFT descriptor and relative location features. The pairwise feature vector, $g_{ij}(y)$, is just the joint of unary feature vector $f_i(y)$ and $f_j(y)$.

Given the locations of local observations, we construct graphical models using MST, as shown in Figure 3(c). The edge cost used in MST construction between two observations was computed by

$$\cos t_{ij} = \varepsilon_1 \times \text{2D distance } (i,j) + \varepsilon_2 \times \text{Distance of color histograms } (i,j), \tag{4.1}$$

where ε_1 and ε_2 are balance factors depending on the actual object, and $\varepsilon_1 + \varepsilon_2 = 1$. If the object has richer shape information than appearance information, we think that the 2D distance might be more useful for discrimination, so we will take a bigger ε_1 than ε_2.

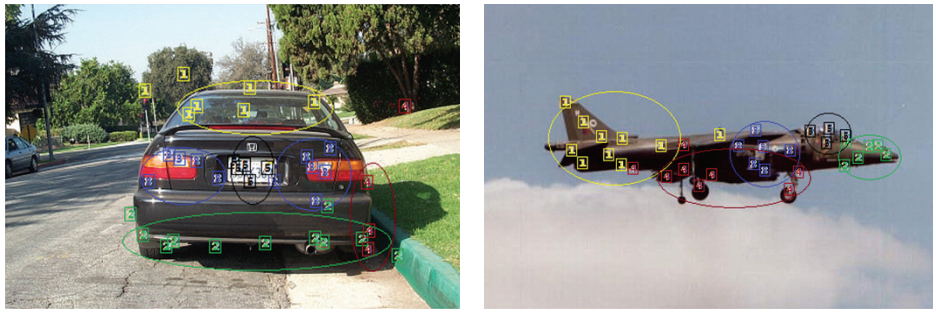
4.2. Object Detection on Standard Database

The first dataset that we used to test our model is a subset of the Caltech-4 standard dataset, which contains images for two object categories, car (rear view) and airplane (side view), and one background category. Each image contains at most a single instance of the objects in diverse natural background and, therefore, is suitable for our 2-class detection task. We randomly split the images into two equal separate subsets for training and testing.

Figure 4 shows the examples of the assignment of parts to local features for two object categories. It is apparent that the proposed model can effectively associate the mass of scattered and unordered observations with their corresponding object parts. Note that multiple observations may be assigned to the same part label with the premise that they physically belong to the same part. The number of parts can be empirically set according to the complexity of objects.

4.3. Airplane Detection in Satellite Images

In this section, we verify our model by 170 airplane (top view) satellite images taken from Google Earth. To gather a sufficiently large learning dataset, we acquire images from different heights and different directions. Furthermore, a few synthetic images with simulation



(a) (b)

Figure 4: Examples of the assignment of parts to local features for car object category (a) and airplane object category (b), which are labeled by different number and colors. The number of parts is set to 5.

Figure 5: Examples of successful detections on satellite images and synthetic images. Note that simulation airplane models in the last two synthetic images are also correctly detected.

airplane models are also used for testing. All images are resized to $150 * 100$ pixels. The balance factor ε_1 used in the weighted neighborhood structure is set to 0.7 to encourage the use of shape information.

Due to space constraints, we provide a few examples of the detection results (as shown in Figure 5). We use a simple bounding box located at the center of the efficient observations to roughly label the detected objects in the test images.

Table 1: Comparisons of detection performance (EER).

Models	Car (rear)	Airplane (side)	Airplane (top)
Hidden CRF	91.0%	94.1%	93.4%
Located HRF	92.1%	95.6%	**97.0**%
Multiclass CRF	90.6%	93.8%	92.1%
Our part-driven CRF	**94.2**%	**97.3**%	**96.5**%
Part-driven CRF without WNS	93.4%	94.9%	93.7%

4.4. Performance Comparison

We compare the detection performance of our model with those of three existing models: hidden CRF [15] model, located HRF [16] model, and multiclass CRF [14] model. For fairness of comparison, the local features in these three models are also computed by SIFT descriptors. In order to measure the influence of neighborhood structures, we also investigate the performance of an equivalent model without weighted neighborhood structure. The object categories are car (rear), airplane (side), and airplane (top), which have been mentioned in previous sections. The equal error rate (EER) defined in [7] is adopted as evaluation criterion, in which higher EER values means better classification performance. The comparative results are summarized in Table 1.

As can be seen, our model consistently gives the best results for these three object categories for the car rear dataset and airplane side dataset. Note that the airplanes (side) are easier to be discriminated than cars due to their distinct shape structure. On the airplane top dataset, our model is exceeded slightly in accuracy only by the located HRF model. This may be caused by overfitting since our model has to use more parameters to encode more contextual dependencies.

From the results in the last row of Table 1, we can see that incorporating the weights of neighborhood structures is important since the performance of such a model dropped. Rather than hypothesizing that all the edges in MST are equally important, the weighted neighborhood structure uses weights to measure the degree of correlation between connected nodes and inherently have higher representational power.

Note that, since the local features are extracted automatically during training and testing, the quantity and structure (constructed by MST) of observations should be unpredictable. As a result, the computing time of our model is influenced by the object complexity and image quality. In this experiment, we use about 3 hours for training, and 1.2 second per image on the average for testing on a 2.8 GHz computer.

5. Conclusion

In this paper we presented a contextual hierarchical part-driven CRF model for object category detection. By incorporating two single-level models, the proposed model can effectively represent latent label-level context and observation-level context simultaneously. A weighted neighborhood structure is also introduced to capture the degree of correlation between connected nodes. Experimental results on challenging datasets with high intraclass variability have demonstrated that the proposed model can effectively represent multiple context information and give competitive detection performance. Our future researches will focus on the following directions: introducing more sparse and robust local features to reduce

the computational complexity and utilizing high-order clique potentials to investigate more contextual dependencies in images.

Acknowledgments

This work is supported by a special financial grant from the China Postdoctoral Science Foundation (Grant no. 20100481512, 201104765). The authors would like to thank Kevin Murphy for publishing CRF2D toolbox, which is very helpful for this work. They also thank X. J. Xiang for data collection.

References

[1] P. Schnitzspan, S. Roth, and B. Schiele, "Automatic discovery of meaningful object parts with latent CRFs," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 121–128, San Diego, Calif, USA, June 2010.

[2] J. H. Zhang, J. W. Zhang, S. Y. Chen et al., "Constructing dynamic category hierarchies for novel visual category discovery," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '12)*, Vilamoura, Portugal, October 2012.

[3] S. Y. Chen, J. H. Zhang, Y. F. Li, and J. W. Zhang, "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 118–127, 2012.

[4] M. Fischler and R. A. Elschlager, "The representation and matching of pictorial structures," *IEEE Transactions on Computers*, vol. 22, no. 1, pp. 67–92, 1973.

[5] S. Lazebnik, C. Schmid, and J. Ponce, "A maximum entropy framework for part-based texture and object recognition," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 832–838, Beijing, China, October 2005.

[6] D. Crandall, P. Felzenszwalb, and D. Huttenlocher, "Spatial priors for part-based recognition using statistical models," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 10–17, San Diego, Calif, USA, June 2005.

[7] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, pp. II/264–II/271, San Diego, Calif, USA, June 2003.

[8] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[9] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky, "Describing visual scenes using transformed objects and parts," *International Journal of Computer Vision*, vol. 77, no. 1–3, pp. 291–330, 2008.

[10] S. Kumar, *Models for learning spatial interactions in natural images for context-based classification [Ph.D. thesis]*, The Robotics Institute, Carnegie Mellon University, 2005.

[11] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Proceedings of the 18th International Conference on Machine Learning (ICML '01)*, pp. 282–289, 2001.

[12] S. Kumar and M. Hebert, "Discriminative fields for modeling spatial dependencies in natural images," *Advances in Neural Information Processing Systems*, pp. 1351–1358, 2004.

[13] S. Kumar and M. Hebert, "Discriminative random fields: A discriminative framework for contextual interaction in classification," in *Proceedings of the International Conference on Computer Vision (ICCV '03)*, vol. 2, pp. 1150–1157, Nice, France, October 2003.

[14] S. Kumar and M. Hebert, "Multiclass discriminative fields for parts-based object detection," in *Snowbird Learning Workshop*, March 2004.

[15] A. Quattoni, S. Wang, L. P. Morency, M. Collins, and T. Darrell, "Hidden conditional random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1848–1853, 2007.

[16] A. Kapoor and J. Winn, "Located hidden random fields: Learning discriminative parts for object detection," in *Proceedings of the European Conference on Computer Vision (ECCV '06)*, pp. 302–315, Vienna, Austria, May 2006.

[17] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
[18] C. Cattani, S. Y. Chen, and G. Aldashev, "Information and modeling in complexity," *Mathematical Problems in Engineering*, vol. 2012, Article ID 868413, 3 pages, 2012.
[19] S. Y. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.
[20] S. Gu, Y. Zheng, and C. Tomasi, "Critical nets and beta-stable features for image matching," in *Proceedings of the 11th European Conference on Computer Vision (ECCV '10)*, vol. 6313 of *Lecture Notes in Computer Science*, pp. 663–676, 2010.
[21] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV'99)*, pp. 1150–1157, Kerkyra, Greece, September 1999.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 673648, 23 pages
doi:10.1155/2012/673648

Review Article

On $1/f$ Noise

Ming Li[1,2] and Wei Zhao[1]

[1] *Department of Computer and Information Science, University of Macau, Avenida Padre Tomas Pereira, Taipa, Macau*
[2] *School of Information Science & Technology, East China Normal University, Shanghai 200062, China*

Correspondence should be addressed to Ming Li, ming_lihk@yahoo.com

Received 11 October 2012; Accepted 23 October 2012

Academic Editor: Carlo Cattani

Due to the fact that $1/f$ noise gains the increasing interests in the field of biomedical signal processing and living systems, we present this introductive survey that may suffice to exhibit the elementary and the particularities of $1/f$ noise in comparison with conventional random functions. Three theorems are given for highlighting the particularities of $1/f$ noise. The first says that a random function with long-range dependence (LRD) is a $1/f$ noise. The second indicates that a heavy-tailed random function is in the class of $1/f$ noise. The third provides a type of stochastic differential equations that produce $1/f$ noise.

1. Introduction

The pioneering work of $1/f$ noise may refer to the paper by Schottky [1], where he introduced the concept of two classes of noise. One class is thermal noise, such as the random motion of molecules in the conductors. The other is the shot noise, which may be caused by randomness of the emission from the cathode and the randomness of the velocity of the emitted electrons [1, 2]. Johnson described the latter using the term of the Schottky effect [3], which may be the first paper in the sense of expressing such a type of processes by the term of $1/f$ noise.

Let $x(t)$ $(-\infty < t < \infty)$ be a random function. Let $S_{xx}(f)$ be its power spectrum density (PSD) function, where $f = \omega/2\pi$ is frequency and ω is radian frequency. Then, by $1/f$ noise, one means that $S_{xx}(f) \to \infty$ for $f \to 0$. Note that the PSD of a conventional random function, such as $(1/(a^2+(2\pi f)^2))$ $(a \neq 0)$, is convergent at $f = 0$. In the field, the term "$1/f$ noise" is a collective noun, which implies in fact $(1/f^\beta)(\beta > 0)$ noise. In the general case, $1/f$ noise has the meaning of $1/f^\alpha$ noise for $\alpha \in \mathbf{R}$, where $\mathbf{R}$ is the set of real numbers; see, for example, [4]. However, since $\lim_{f\to 0}(1/f^\alpha) = 0$ for $\alpha < 0$, one may usually not be interested in the case of $\alpha < 0$. In what follows, we discuss the noise of $1/f^\beta$ type for $\beta > 0$ unless otherwise stated.

Since the notion of $1/f$ noise appeared [3], it has gained increasing interests of scientists in various fields, ranging from bioengineering to computer networks; see, for example, [4–54], simply to mention a few. That fact gives rise to a question of what $1/f$ noise is. The question may be roughly answered in a way that a $1/f$ noise $x(t)$ is such that its PSD $S_{xx}(f)$ is divergent at $f = 0$ as previously mentioned. Nonetheless, that answer may be never enough to describe the full picture of $1/f$ noise. By full picture, we mean that we should describe a set of main properties of $1/f$ noise, in addition to the property $\lim_{f\to 0}(1/f^{\beta}) = \infty$ in frequency domain. When regarding $\lim_{f\to 0}(1/f^{\beta}) = \infty$ as its first property, denoted by P1, we would like to list other three as follows.

P2: What is the qualitative structure of an autocorrelation function (ACF) of $1/f$ noise? For this property, we will discuss the statistical dependence based on the hyperbolically decayed ACF structure.

P3: What is the main property of its probability density function (PDF)? With P3, we will explain the heavy-tailed property of $1/f$ noise, which may produce random functions without mean or variance.

P4: What is the possible structure of the differential equation to synthesize $1/f$ noise?

For facilitating the description of the full picture of $1/f$ noise, we will brief the preliminaries in Section 2. Then, P2–P4 will be discussed in Sections 3–5, respectively. After that, we will conclude the paper in Section 6.

2. Preliminaries

2.1. Dependence of Random Variables

A time series $x(t)$ may also be called a random function [55]. The term random function apparently exhibits that $x(t_i)$ $(i = 0, 1, \ldots)$ is a random variable, implying

$$x(t_i) \neq x(t_j) \quad \text{for } i \neq j \ (i, j = 0, 1, \ldots). \tag{2.1}$$

We would like to discuss the dependence of $x(t_i)$ and $x(t_j)$ for $i \neq j$, as well as $x(t)$.

2.1.1. Dependence Description of Random Variables with Probability

Let $P\{x(t_i) < x_1\}$ be the probability of the event $\{x(t_i) < x_1\}$. Denote by $P\{x(t_j) < x_2\}$ the probability of the event $\{x(t_j) < x_2\}$. Then [56], $\{x(t_i) < x_1\}$ and $\{x(t_j) < x_2\}$ are said to be independent events if

$$P\left[\{x(t_i) < x_1\}\{x(t_j) < x_2\}\right] = P\{x(t_i) < x_1\}P\{x(t_j) < x_2\}. \tag{2.2}$$

If $\{x(t_i) < x_1\}$ and $\{x(t_j) < x_2\}$ are dependent, on the other side,

$$P\left[\{x(t_i) < x_1\}\{x(t_j) < x_2\}\right] = P\{x(t_j) < x_2\}P\left[\{x(t_i) < x_1\} \mid \{x(t_j) < x_2\}\right], \tag{2.3}$$

where $P[\{x(t_i) < x_1\} \mid \{x(t_j) < x_2\}]$ is the conditional probability, implying that the probability of the event $\{x(t_i) < x_1\}$ provided that $\{x(t_j) < x_2\}$ has occurred.

Note 1. The dependence of $\{x(t_i) < x_1\}$ and $\{x(t_j) < x_2\}$ is reflected in the conditional probability $P[\{x(t_i) < x_1\} \mid \{x(t_j) < x_2\}]$. If $\{x(t_i) < x_1\}$ and $\{x(t_j) < x_2\}$ are independent, $P[\{x(t_i) < x_1\} \mid \{x(t_j) < x_2\}] = P\{x(t_i) < x_1\}$.

2.1.2. Dependence Description of Gaussian Random Variables with Correlation

Let $x(t)$ be a Gaussian random function. Denote by $\operatorname{corr}[\{x(t_i) < x_1\}, \{x(t_j) < x_2\}]$ the correlation coefficient between $\{x(t_i) < x_1\}$ and $\{x(t_j) < x_2\}$. Then [57, 58], $\{x(t_i) < x_1\}$ and $\{x(t_j) < x_2\}$ are independent if

$$\operatorname{corr}\left[\{x(t_i) < x_1\}, \{x(t_j) < x_2\}\right] = a = 0. \tag{2.4}$$

On the other hand, $\{x(t_i) < x_1\}$ and $\{x(t_j) < x_2\}$ are dependent provided that

$$\operatorname{corr}\left[\{x(t_i) < x_1\}, \{x(t_j) < x_2\ \}\right] = b \neq 0. \tag{2.5}$$

The condition (2.4) or (2.5) expressed by the correlation coefficient regarding the independence or dependence may not be enough to identify the independence or dependence of a Gaussian random function completely. For example, when $t_j = t_i + \tau$, that (2.4) holds for large τ may not imply that it is valid for small τ. In other words, one may encounter the situations expressed by

$$\operatorname{corr}[\{x(t_i) < x_1\}, \{x(t_i + \tau) <\ x_2\}] = 0 \quad \text{for large } \tau, \tag{2.6}$$

$$\operatorname{corr}[\{x(t_i) < x_1\}, \{x(t_i\ +\ \tau) < x_2\}] \neq 0 \quad \text{for small } \tau. \tag{2.7}$$

By using the concept of probability, (2.6) corresponds to

$$P[\{x(t_i) < x_1\}\{x(t_i\ + \tau) < x_2\}] = P\{x(t_i) < x_1\}P\{x(t_i + \tau) < x_2\} \quad \text{for large } \tau. \tag{2.8}$$

Similarly, (2.7) corresponds to (2.9)

$$\begin{aligned} &P[\{x(t_i) < x_1\}\{x(t_i +\ \tau) < x_2\}] \\ &\quad = P\{x(t_i + \tau) < x_2\}P[\{x(t_i) < x_1\} \mid \{x(t_i + \tau) < x_2\}] \quad \text{for small } \tau. \end{aligned} \tag{2.9}$$

Note 2. The notion of the dependence or independence of a set of random variables plays a role in the axiomatic approach of probability theory and stochastic processes; see Kolmogorov [59].

The above example exhibits an interesting fact that the dependence or independence relies on the observation scale or observation range. In conventional time series, we do not usually consider the observation scale. That is, (2.2) and (2.3) or (2.4) and (2.5) hold for all observation ranges no matter whether τ is small or large; see, for example, [60, 61]. Statistical properties that depend on observation ranges are briefed by Papoulis and Pillai [56] and Fuller [62] but detailed in Beran [63, 64].

Note 3. The Kolmogorov's work on axiomatic approach of probability theory and stochastic processes needs the assumption that $\{x(t_i) < x_1\}$ and $\{x(t_j) < x_2\}$ are independent in most cases likely for the completeness of his theory. Nevertheless, he contributed a lot in random functions that are range dependent; see, for example, [65, 66].

2.1.3. Dependence Description of Gaussian Random Variables with ACF

Denote by $r_{xx}(\tau, t)$ the ACF of $x(t)$. Denote by E the operator of mean. Then, it is given by

$$r_{xx}(\tau, t) = \mathrm{E}[x(t)x(t+\tau)]. \tag{2.10}$$

It represents the correlation between the one point $x(t)$ and the other τ apart, that is, $x(t+\tau)$. For facilitating the discussions, we assume that $x(t)$ is stationary in the wide sense (stationary for short). With this assumption, $r_{xx}(\tau, t)$ only replies on the lag τ. Therefore, one has

$$r_{xx}(\tau, t) = r_{xx}(\tau) = \mathrm{E}[x(t)x(t+\tau)]. \tag{2.11}$$

In the normalized case,

$$0 \le |r_{xx}(\tau)| \le 1. \tag{2.12}$$

ACF is a convenient tool of describing the dependence of a Gaussian random function $x(t)$. For instance, on the one hand, we say that any two different points of $x(t)$ are uncorrelated, accordingly independent as $x(t)$ is Gaussian, if $r_{xx}(\tau) = 0$. That is the case of Gaussian white noise. On the other hand, any two different points of $x(t)$ are strongest dependent if $r_{xx}(\tau) = 1$. This is the case of the strongest long-range dependence. In the case of $0 < |r_{xx}(\tau)| < 1$, the value of $|r_{xx}(\tau)|$ varies with the lag $\tau = (t+\tau) - t$.

A useful measure called correlation time, which is denoted by τ_c [67, page 74], is defined in the form

$$\tau_c = \frac{1}{\lim_{T\to\infty} \int_0^T r_{xx}(\tau)d\tau} \lim_{T\to\infty} \int_0^T \tau r_{xx}(\tau)d\tau. \tag{2.13}$$

By correlation time, we say that the correlation can be neglected if $\tau_c \le \tau$, where τ is the time scale of interest [67].

Note 4. For a conventional Gaussian random function $x(t)$, its correlation can be neglected if $\tau_c \le \tau$. This implies that the statistical dependence or independence of $x(t)$ relies on its correlation time τ_c. However, we will show in Note 8 that correlation time fails if $x(t)$ is a $1/f$ noise.

2.2. ACF and PSD

The PSD of $x(t)$ is the Fourier transform of its ACF. That is,

$$S_{xx}(f) = \int_{-\infty}^{\infty} r_{xx}(t)e^{-j2\pi ft}dt, \quad j = \sqrt{-1}. \tag{2.14}$$

Equivalently,

$$r_{xx}(\tau) = \int_{-\infty}^{\infty} S_{xx}(f)e^{j2\pi f\tau}df. \tag{2.15}$$

Letting $f = 0$ in (2.14) yields

$$S_{xx}(0) = \int_{-\infty}^{\infty} r_{xx}(\tau)d\tau. \tag{2.16}$$

Similarly, letting $\tau = 0$ in (2.15) produces

$$r_{xx}(0) = \int_{-\infty}^{\infty} S_{xx}(f)df. \tag{2.17}$$

Note 5. Equation (2.16) implies that an ACF $r_{xx}(\tau)$ is integrable if $S_{xx}(0) < \infty$. On the other side, a PSD $S_{xx}(f)$ is integrable when $r_{xx}(0) < \infty$ as indicated in (2.17). Both are usual cases in conventional random functions.

Note 6. The noise of $1/f$ type has the property $S_{xx}(0) = \infty$, which makes $1/f$ noise substantially different from conventional random functions.

2.3. Mean and Variance

Denote by $p(x)$ the PDF of $x(t)$. Then, the mean denoted by $\mu(t)$ is given by

$$\mu(t) = \mathrm{E}[x(t)] = \int_{-\infty}^{\infty} xp(x)dx. \tag{2.18}$$

The variance denoted by $\sigma^2(t)$ is in the form

$$\sigma^2(t) = \mathrm{E}\left\{\left[x(t) - \mu(t)\right]^2\right\} = \int_{-\infty}^{\infty} \left[x(t) - \mu(t)\right]^2 p(x)dx. \tag{2.19}$$

Assume that $x(t)$ is stationary. Then, $\mu(t)$ and $\sigma^2(t)$ do not rely on time t. In this case, they are expressed by

$$\mu(t) = \mu, \qquad \sigma^2(t) = \sigma^2. \tag{2.20}$$

Without generality losing, we always assume that $x(t)$ is stationary in what follows unless otherwise stated.

Denote by $C_{xx}(\tau)$ the autocovariance of $x(t)$. Then,

$$C_{xx}(\tau) = \mathrm{E}\{[x(t)-\mu][x(t+\tau)-\mu]\} = \int_{-\infty}^{\infty}[x(t)-\mu][x(t+\tau)-\mu]p(x)dx. \tag{2.21}$$

Taking into account (2.19) and (2.21), one has

$$C_{xx}(0) = \sigma^2. \tag{2.22}$$

Denote by Ψ the mean square value of $x(t)$. Then,

$$\Psi = \mathrm{E}\left[x^2(t)\right] = \int_{-\infty}^{\infty} x^2 p(x)dx. \tag{2.23}$$

Considering $r_{xx}(\tau)$ given by

$$r_{xx}(\tau) = \mathrm{E}[x(t)x(t+\tau)] = \int_{-\infty}^{\infty} x(t)x(t+\tau)p(x)dx, \tag{2.24}$$

one has

$$\Psi = r_{xx}(0). \tag{2.25}$$

Therefore,

$$\sigma^2 = \Psi - \mu^2. \tag{2.26}$$

It is worth noting that the above number characteristics are crucial to the analysis of $x(t)$ in practice; see, for example, [60–62, 65–87], just to cite a few. However, things turn to be complicated if $x(t)$ is in the class of $1/f$ noise. In Section 4 below, we will show that μ and/or σ^2 may not exist for specific types of $1/f$ noise.

3. Hyperbolically Decayed ACFs and $1/f$ Noise

The qualitative structure of $1/f$ noise is in the form K/f^{β}, where K is a constant. Since $\lim_{f\to 0}(K/f^{\beta}) = \infty$, as we previously mentioned several times, its ACF is nonintegrable over $(-\infty, \infty)$. That is,

$$S_{xx}(0) = \int_{-\infty}^{\infty} r_{xx}(\tau)d\tau = \infty. \tag{3.1}$$

Note that the above may be taken as a definition of LRD property of a random function; see, for example, [4, 63, 64, 88–90]. Thus, the above exhibits that $1/f$ noise is LRD. Consequently, $r_{xx}(\tau)$ of a $1/f$ noise $x(t)$ may have the asymptotic property given by

$$r_{xx}(\tau) \sim c|\tau|^{-\gamma}, \quad c > 0,\ 0 < \gamma < 1,\ \tau \longrightarrow \infty. \tag{3.2}$$

Following [91–93], we have the Fourier transform of $|\tau|^{-\gamma}$ in the form

$$\mathrm{F}\left(|\tau|^{-\gamma}\right) = \int_{-\infty}^{\infty} |\tau|^{-\gamma} e^{-j2\pi f\tau} d\tau = 2\sin\left(\frac{\pi\gamma}{2}\right)\Gamma(1-\gamma)\left|2\pi f\right|^{\gamma-1}. \tag{3.3}$$

From the above, we have

$$\gamma = 1 - \beta. \tag{3.4}$$

Thus, we have the following note.

Note 7. Qualitatively, the ACF of $1/f$ noise is in the structure of power function. It follows power law. This is the answer to P2 explained in the Introduction.

Example 3.1. A well-known example of $1/f$ noise is the fractional Gaussian noise (fGn) [93, 94]. Its ACF is in the form

$$r_{\mathrm{fGn}}(\tau) = \frac{V_H}{2}\left[(|\tau|+1)^{2H} + ||\tau|-1|^{2H} - 2|\tau|^{2H}\right], \tag{3.5}$$

where $0 < H < 1$ is the Hurst parameter; $V_H = (H\pi)^{-1}\Gamma(1-2H)\cos(H\pi)$ is the strength of fGn. Its PSD is in the form [93]

$$S_{\mathrm{fGn}}(f) = V_H \sin(H\pi)\Gamma(2H+1)\left|2\pi f\right|^{1-2H}. \tag{3.6}$$

Example 3.2. Another example of $1/f$ noise is fractional Brownian motion (fBm). The PSD of the fBm of the Weyl type is in the form [95, 96]

$$S_{\mathrm{fBm,Weyl}}(t,f) = \frac{1}{\left|2\pi f\right|^{2H+1}}\left(1 - 2^{1-2H}\cos 4\pi f t\right), \quad 0 < H < 1. \tag{3.7}$$

The PSD of the fBm of the Riemann-Liouville type is given by [97–99]

$$S_{\mathrm{fBm,Riemann\text{-}Liouville}}(t,f) = \frac{\pi\omega t}{(2\pi f)^{2H+1}}\left[J_H(4\pi f t)\mathbb{H}_{H-1}(4\pi f t) - J_{H-1}(4\pi f t)\mathbb{H}_H(4\pi f t)\right], \tag{3.8}$$

where J_H is the Bessel function of order H and $\mathbb{H}_H$ is the Struve function of order H, respectively.

Example 3.3. The Cauchy-class process with LRD discussed in [100, 101] is a case of $1/f$ noise. Its ACF is in the form

$$r_{\text{Cauchy-class}}(\tau) = \left(1 + |\tau|^2\right)^{-b/2}, \quad 0 < b < 1. \tag{3.9}$$

Its PSD is given by

$$S_{\text{Cauchy-class}}(f) = \int_{-\infty}^{\infty} \left(1 + |\tau|^2\right)^{-b/2} e^{-j2\pi f\tau} d\tau = \frac{2^{(1-b)/2}}{\sqrt{\pi}\Gamma(b/2)} |2\pi f|^{1/2(b-1)} K_{1/2(b-1)}(|2\pi f|), \tag{3.10}$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind. It has the asymptotic expression for $f \to 0$ in the form, being $1/f$ noise,

$$S_{\text{Cauchy-class}}(f) \sim \frac{\Gamma[(1/2)(1-b)]}{2^b\sqrt{\pi}\Gamma(b/2)} |2\pi f|^{b-1}. \tag{3.11}$$

Example 3.4. The generalized Cauchy process with LRD reported in [102–105] is an instance of $1/f$ noise. Its ACF is given by

$$r_{\text{GC}}(\tau) = \left(1 + |\tau|^\alpha\right)^{-\beta/\alpha}, \tag{3.12}$$

where $0 < \alpha \leq 2$ and $0 < \beta < 1$. Its PSD in the complete form refers to [92]. The following may suffice to exhibit its $1/f$ noise behavior [102]:

$$S_{\text{GC}}(f) \sim \frac{1}{\Gamma(\beta)\cos(\beta\pi/2)} |2\pi f|^{\beta-1}, \quad f \longrightarrow 0. \tag{3.13}$$

It may be worthwhile for us to write a theorem and a note when this section will soon finish.

Theorem 3.5. *If a random function $x(t)$ is LRD, it belongs to the class of $1/f$ noise and vice versa.*

Proof. LRD implies that the right side of (2.16) is divergent, which implies that $x(t)$ is a $1/f$ noise. On the other side, when $x(t)$ is a $1/f$ noise, $S_{xx}(0) = \infty$, which means that its ACF is nonintegrable, hence, LRD. □

Since any random function with LRD is in the class of $1/f$ noise, one may observe other types of random functions that belong to $1/f$ noise from a view of LRD processes, for example, those described in [106–112].

Note 8. As $1/f$ noise is of LRD, its ACF is nonintegrable over $(-\infty, \infty)$. Thus, in (2.13), its correlation time $\tau_c \to \infty$. That implies that correlations at any time scale cannot be neglected. Consequently, the measure of correlation time fails to characterize the statistical dependence of $1/f$ noise.

4. Heavy-Tailed PDFs and $1/f$ Noise

Heavy-tailed PDFs are widely observed in various fields of sciences and technologies, including life science and bioengineering; see, for example, [113–146]. Typical heavy-tailed PDFs are the Pareto distribution, the log-Weibull distribution, the stretched exponential distribution, the Zipfian distribution, Lévy distribution, and the Cauchy distribution; see, for example [51, 96, 125, 143–169], merely to cite a few.

By heavy tails, we mean that the tail of a PDF $p(x)$ decays slower than the tails of PDFs in the form of exponential functions. More precisely, the term of heavy tail implies that $p(x)$ of a random function $x(t)$ decays slowly such that $C_{xx}(\tau)$ in (2.21), or $r_{xx}(\tau)$ in (2.24), decays hyperbolically such that it is of LRD. Thus, we may have the theorem below.

Theorem 4.1. *Let $x(t)$ be a heavy-tailed random function. Then, it is in the class of $1/f$ noise and vice versa.*

Proof. Considering that $x(t)$ is heavy tailed, we may assume its ACF hyperbolically decays, that is, $r_{xx}(\tau) \sim c|\tau|^{-\gamma}$ for $\tau \to \infty$ and $0 < \gamma < 1$. According to (3.3), therefore, we see that $x(t)$ is $1/f$ noise. On the other side, if $x(t)$ is a $1/f$ noise, it is LRD and accordingly heavy tailed [54, 146]. This completes the proof. □

Note 9. Theorem 4.1 may be taken as an answer to P3 stated in the Introduction. Since γ in Theorem 4.1 is restricted to $(0, 1)$, Theorem 4.1 is consistent with the result of the Taqqiu's law, referring [53, 54, 89] for the details of the Taqqiu's law.

Note 10. The tail of $p(x)$ may be so heavy that the mean or variance of $x(t)$ does not exist.

A commonly used instance to clarify Note 10 is the Pareto distribution; see, for example, [96, 134, 147, 148, 161]. To clarify it further, we would like to write more. Denote by $p_{\text{Cauchy}}(x)$ the Cauchy distribution. Then, one has

$$p_{\text{Cauchy}}(x) = \frac{b}{\pi\left[(x-m)^2 + b^2\right]}, \tag{4.1}$$

where b is the half width at half maximum and m is the statistical median [170]. Its nth moment denoted by m_n is computed by

$$m_n = \mathrm{E}[x^n(t)] = \int_{-\infty}^{\infty} x^n p_{\text{Cauchy}}(x)dx = \frac{\Gamma}{2\pi}\int_{-\infty}^{\infty} \frac{x^n dx}{(x-m)^2 + (0.5\Gamma)^2}. \tag{4.2}$$

Since the above integral is divergent for $n \geq 1$, the mean and variance of $x(t)$ obeying $p_{\text{Cauchy}}(x)$ do not exist.

Note 11. Application of the Cauchy distribution to network traffic modeling refers to [169]. The generalized Cauchy distribution is reported in [171–173].

Another type of heavy-tailed random functions without mean and variance is the Lévy distribution; see, for example, [162–166, 174]. Suppose that $x(t)$ follows the Lévy distribution that is denoted by $p_{\text{Levy}}(x)$. Then, $p_{\text{Levy}}(x)$ for $x \geq \mu$ is given by

$$p_{\text{Levy}}(x) = p_{\text{Levy}}(x; \mu, c) = \sqrt{\frac{c}{2\pi}} \frac{\exp\left[-c/(2(x-\mu))\right]}{(x-\mu)^{3/2}}, \tag{4.3}$$

where μ is the location parameter and c the scale parameter. The nth moment of a Lévy distributed random function, for the simplicity by letting $\mu = 0$, is given by

$$m_n = \sqrt{\frac{c}{2\pi}} \int_0^{\infty} x^n x^{-3/2} e^{-c/(2x)} dx. \tag{4.4}$$

The integral in (4.4) is divergent for $n > 0$. Thus, its mean and variance do not exist because they approach ∞.

Note 12. Application of the Lévy distribution to network traffic modeling is discussed in [175]. Tools in Matlab for simulating Lévy distributed random data are described in [174].

The previously discussed heavy-tailed distributions, such as the Cauchy distribution and the Lévy distribution, are special cases of stable distributions, which are detailed in [89, 120, 176–179]. Denote by $p_{\text{Stable}}(x)$ the PDF of a stable-distributed random function x. Then, it is indirectly defined by its characteristic function denoted by $\varphi_{\text{Stable}}(t; \mu, c, \alpha, \beta)$. It is in the form

$$\varphi_{\text{Stable}}(t; \mu, c, \alpha, \beta) = \int_{-\infty}^{\infty} p_{\text{Stable}}(x) e^{j2\pi xt} dx = \exp\left[jt\mu - |ct|^{\alpha}(1 - j\beta \operatorname{sgn}(t)\Phi)\right], \tag{4.5}$$

where $\alpha \in (0, 2]$ is the stability parameter, $\beta \in [-1, 1]$ is the skewness parameter, $c \in (0, \infty)$ is the scale parameter, $\mu \in (-\infty, \infty)$ the location parameter, and

$$\Phi = \begin{cases} \tan\dfrac{\pi\alpha}{2}, & \text{if } \alpha \neq 1 \\ -\dfrac{2}{\pi}\log|t|, & \text{if } \alpha = 1. \end{cases} \tag{4.6}$$

It may be easy to see that x has mean μ when $\alpha > 1$. However, its mean is undefined otherwise. In addition, its variance equals to $2c^2$ if $\alpha = 2$. Otherwise, its variance is infinite.

Note 13. A stable distribution is characterized by 4 parameters. In general, the analytical expression of $p_{\text{Stable}}(x)$ is unavailable except for some specific values of parameters, due to the difficulties in performing the inverse Fourier transform of the right side of (4.5).

Note 14. A stable distribution is generally non-Gaussian except for some specific values of parameters. When $\alpha = 2$, it is Gaussian with the mean μ and the variance of $2c^2$. Generally speaking, $p_{\text{Stable}}(x)$ is heavy tailed.

Note 15. $p_{\text{Stable}}(x)$ reduces to the Landau distribution when $\alpha = \beta = 1$. Denote by Landau(μ, c) the Landau distribution [180]. Then, its characteristic function is given by

$$\varphi_{\text{Landau}}(t;\mu,c) = \int_{-\infty}^{\infty} p_{\text{Landau}}(x)e^{j2\pi xt}dx = \exp\left[jt\mu - |ct|\left(1 - \frac{2j}{\pi}\log(|t|)\right)\right], \tag{4.7}$$

where μ is the location parameter and c the scaled parameter. Its PDF is given by

$$p_{\text{Landau}}(x) = \frac{1}{\pi}\int_0^{\infty} e^{-t\log t - xt}\,\sin(\pi t)dt. \tag{4.8}$$

Applications of the Landau distribution can be found in nuclear physics [181, 182].

Note 16. $p_{\text{Stable}}(x)$ reduces to the Holtsmark distribution if $\alpha = 3/2$ and $\beta = 0$. Accordingly,

$$\varphi_{\text{Holtsmark}}(t;\mu,c) = \int_{-\infty}^{\infty} p_{\text{Holtsmark}}(x)e^{j2\pi xt}dx = \exp\left(jt\mu - |ct|^{3/2}\right). \tag{4.9}$$

Its PDF is given by

$$p_{\text{Holtsmark}}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(jt\mu - |ct|^{3/2}\right)e^{-j2\pi xt}dt. \tag{4.10}$$

This type of random functions has mean μ but its variance is infinite [183–185].

The literature regarding applications of stable distributions is rich. Their applications to network traffic modeling can also be found in [186–188].

It is worth noting that the observation of random functions without mean and variance may be traced back to the work of the famous statistician Daniel Bernoulli's cousin, Nicolas Bernoulli in 1713 [189, 190]. Nicolas Bernoulli studied a casino game as follows. A player bets on how many tosses of a coin will be needed before it first turns up heads. If it falls heads on the first toss the player wins \$2; if it falls tails, heads, the player wins \$4; if it falls tails, tails, heads, the player wins \$8, and so on. According to this game rule, if the probability of an outcome is 2^{-k}, the player wins $\$2^k$. Thus, the mean for $k \to \infty$ is given by

$$\sum_{k=1}^{\infty} 2^{-k}\left(\$2^k\right) = \infty. \tag{4.11}$$

The above may be used to express the game that is now termed "Petersburg Paradox." That paradox is now named after Daniel Bernoulli due to his presentation of the problem and his solution in the Commentaries of the Imperial Academy of Science of Saint Petersburg [191].

5. Fractionally Generalized Langevin Equation and $1/f$ Noise

The standard Langevin equation is in the form

$$\left(\frac{d}{dt}+\lambda\right)X(t)=w(t),\qquad X(0)=X_0, \tag{5.1}$$

where $\lambda>0$ [56, 192] and $w(t)$ is a standard white noise. By standard white noise, we mean that its PSD $S_{ww}(f)$ is in the form

$$S_{ww}(f)=1\quad \text{for } f\in(-\infty,\infty). \tag{5.2}$$

The solution to (5.1) in frequency domain is given by

$$S_{XX}(f)=\frac{1}{\lambda^2+(2\pi f)^2}. \tag{5.3}$$

The standard Langevin equation may not attract people much in the field of $1/f$ noise. People are usually interested in fractionally generalized Langevin equations. There are two types of fractionally generalized Langevin equations. One is in the form [193–208]

$$\left(\frac{d^{1+\varepsilon}}{dt^{1+\varepsilon}}+\lambda\right)X_1(t)=w(t),\quad \varepsilon>-1. \tag{5.4}$$

The other is expressed by [209–215]

$$\left(\frac{d}{dt}+\lambda\right)^{\alpha}X_1(t)=w(t),\quad \alpha>0. \tag{5.5}$$

Two are consistent when $\lambda=0$. We now adopt the one expressed by (5.5).

Theorem 5.1. *A solution to the stochastically fractional differential equation below belongs to* $1/f$ *noise:*

$$\frac{d^{\alpha}}{dt^{\alpha}}X_1(t)=w(t),\quad \alpha>0. \tag{5.6}$$

Proof. Denote by $g_{X_1}(t)$ the impulse response function of (5.5). That is,

$$\left(\frac{d}{dt}+\lambda\right)^{\alpha}g_{X_1}(t)=\delta(t), \tag{5.7}$$

where $\delta(t)$ is the Dirac-δ function defined by, for $f(t)$ being continuous at $t = 0$,

$$\int_{-\infty}^{\infty} \delta(t) f(t) dt = f(0), \quad \delta(t) = 0, \ t \neq 0. \tag{5.8}$$

The function $\delta(t)$ is called the impulse function in linear systems [61, 216]. According to the theory of linear systems, $g_{X_1}(t)$ is the solution to (5.7) under the zero initial condition, which is usually called the impulse response function in linear systems [61, 68, 69, 76, 77, 216–222]. Denote by $G_{X_1X_1}(f)$ the Fourier transform of $g_{X_1}(t)$. Then, with the techniques in fractional calculus [213, 215, 223], doing the Fourier transforms on (5.7) yields

$$G_{X_1X_1}(f) = \frac{1}{(\lambda + j2\pi f)^{\alpha}}. \tag{5.9}$$

Therefore, we have, by taking into account (5.2),

$$S_{X_1X_1}(f) = \left|G_{X_1X_1}(f)\right|^2 S_{ww}(f) = \left|G_{X_1X_1}(f)\right|^2 = \frac{1}{\left[\lambda^2 + (2\pi f^2)\right]^{\alpha}}. \tag{5.10}$$

If $\lambda = 0$ in (5.10), we have $1/f$ noise expressed by

$$S_{X_1X_1}(\omega) = \frac{1}{(2\pi f)^{2\alpha}}. \tag{5.11}$$

This finishes the proof. □

From the above, we see that $X_1(t)$ belongs to $1/f$ noise. As a matter of fact, by using fractional integral in (5.6), we have

$$X_1(t) = \frac{d^{-\alpha}}{dt^{-\alpha}} w(t), \quad \alpha > 0. \tag{5.12}$$

The above expression implies that a $1/f$ noise may be taken as a solution to a stochastically fractional differential equation, being an answer to P4 described in the Introduction. The following example will soon refine this point of view.

Example 5.2. Let $B(t)$ be the Wiener Brownian motion for $t \in (0, \infty)$; see [224] for the details of the Brownian motion. Then, it is nondifferentiable in the domain of ordinary functions. It is differentiable, however, in the domain of generalized functions over the Schwartz space of test functions [91, 225]. Therefore, in the domain of generalized functions, we write the stationary Gaussian white noise by

$$w(t) = \frac{dB(t)}{dt}. \tag{5.13}$$

Based on the definitions of the fractional integrals of the Riemann-Liouville's and the Weyl's [223], on the one hand, when using the Riemann-Liouville integral operator, we express the fBm of the Riemann-Liouville type, which is denoted by $B_{H,\mathrm{RL}}(t)$, in the form

$$B_{H,\mathrm{RL}}(t) = {}_0D_t^{-(H+1/2)}\,\frac{dB(t)}{dt} = \frac{1}{\Gamma(H+1/2)}\int_0^t (t-u)^{H-1/2}dB(u), \tag{5.14}$$

where ${}_0D_t^{-(H+1/2)}$ is the Riemann-Liouville integral operator of order $(0.5+H)$ for $0<H<1$. On the other hand, the fBm of the Weyl type by using the Weyl fractional integral, which we denote by $B_{H,\mathrm{Weyl}}(t)$, is given by

$$\begin{aligned} &W^{-(H+1/2)}\frac{dB(t)}{dt} \\ &\quad = B_{H,\mathrm{Weyl}}(t) - B_{H,\mathrm{Weyl}}(0) \\ &\quad = \frac{1}{\Gamma(H+1/2)}\left\{\int_{-\infty}^0 \left[(t-u)^{H-0.5} - (-u)^{H-0.5}\right]dB(u) + \int_0^t (t-u)^{H-0.5}dB(u)\right\}, \end{aligned} \tag{5.15}$$

where $W^{-(H+1/2)}$ is the Weyl integral operator of order $(0.5+H)$. Expressions (5.14) and (5.15) are the fBms introduced by Mandelbrot and van Ness in [94] but we provide a new outlook of describing them from the point of view of the fractional generalized Langevin equation with the topic of $1/f$ noise.

6. Conclusions

We have explained the main properties of $1/f$ noise as follows. First, it is LRD and its ACF is hyperbolically decayed. Second, its PDF obeys power laws and it is heavy tailed. Finally, it may be taken as a solution to a stochastic differential equation. Fractal time series, such as fGn, fBm, the generalized Cauchy process, and the Lévy flights, α-stable processes, are generally in the class of $1/f$ noise.

Acknowledgments

This work was supported in part by the 973 Plan under the Project Grant no. 2011CB302800 and by the National Natural Science Foundation of China under the Project Grant nos. 61272402, 61070214, and 60873264.

References

[1] W. Schottky, "Uber spontane Stromschwankungen in verschiedenen Elektri-zittsleitern," *Annalen der Physik*, vol. 362, no. 23, pp. 541–567, 1918.
[2] W. Schottky, "Zur Berechnung und Beurteilung des Schroteffektes," *Annalen der Physik*, vol. 373, no. 10, pp. 157–176, 1922.
[3] J. B. Johnson, "The Schottky effect in low frequency circuits," *Physical Review*, vol. 26, no. 1, pp. 71–85, 1925.
[4] B. B. Mandelbrot, *Multifractals and 1/f Noise*, Springer, New York, NY, USA, 1998.

[5] K. Fraedrich, U. Luksch, and R. Blender, "$1/f$ model for long-time memory of the ocean surface temperature," *Physical Review E*, vol. 70, no. 3, Article ID 037301, 4 pages, 2004.
[6] E. J. Wagenmakers, S. Farrell, and R. Ratcliff, "Estimation and interpretation of $1/f^{\alpha}$ noise in human cognition," *Psychonomic Bulletin and Review*, vol. 11, no. 4, pp. 579–615, 2004.
[7] F. Principato and G. Ferrante, "$1/f$ noise decomposition in random telegraph signals using the wavelet transform," *Physica A*, vol. 380, no. 1-2, pp. 75–97, 2007.
[8] V. P. Koverda and V. N. Skokov, "Maximum entropy in a nonlinear system with a $1/f$ power spectrum," *Physica A*, vol. 391, no. 1-2, pp. 21–28, 2012.
[9] Y. Nemirovsky, D. Corcos, I. Brouk, A. Nemirovsky, and S. Chaudhry, "$1/f$ noise in advanced CMOS transistors," *IEEE Instrumentation and Measurement Magazine*, vol. 14, no. 1, pp. 14–22, 2011.
[10] O. Miramontes and P. Rohani, "Estimating $1/f^{\alpha}$ scaling exponents from short time-series," *Physica D*, vol. 166, no. 3-4, pp. 147–154, 2002.
[11] C. M. van Vliet, "Random walk and $1/f$ noise," *Physica A*, vol. 303, no. 3-4, pp. 421–426, 2002.
[12] J. S. Kim, Y. S. Kim, H. S. Min, and Y. J. Park, "Theory of $1/f$ noise currents in semiconductor devices with one-dimensional geometry and its application to Si Schottky barrier diodes," *IEEE Transactions on Electron Devices*, vol. 48, no. 12, pp. 2875–2883, 2001.
[13] T. Antal, M. Droz, G. Györgyi, and Z. Rácz, "$1/f$ noise and extreme value statistics," *Physical Review Letters*, vol. 87, no. 24, Article ID 240601, 4 pages, 2001.
[14] B. Pilgram and D. T. Kaplan, "A comparison of estimators for $1/f$ noise," *Physica D*, vol. 114, no. 1-2, pp. 108–122, 1998.
[15] H. J. Jensen, "Lattice gas as a model of $1/f$ noise," *Physical Review Letters*, vol. 64, no. 26, pp. 3103–3106, 1990.
[16] E. Marinari, G. Parisi, D. Ruelle, and P. Windey, "Random walk in a random environment and $1/f$ noise," *Physical Review Letters*, vol. 50, no. 17, pp. 1223–1225, 1983.
[17] F. N. Hooge, "$1/f$ noise," *Physica B*, vol. 83, no. 1, pp. 14–23, 1976.
[18] M. B. Weissman, "Simple model for $1/f$ noise," *Physical Review Letters*, vol. 35, no. 11, pp. 689–692, 1975.
[19] F. N. Hooge, "Discussion of recent experiments on $1/f$ noise," *Physica*, vol. 60, no. 1, pp. 130–144, 1972.
[20] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.
[21] C. Cattani, "On the existence of wavelet symmetries in archaea DNA," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 673934, 21 pages, 2012.
[22] C. Cattani, E. Laserra, and I. Bochicchio, "Simplicial approach to fractal structures," *Mathematical Problems in Engineering*, vol. 2012, Article ID 958101, 21 pages, 2012.
[23] C. Cattani, "Fractional calculus and Shannon wavelet," *Mathematical Problems in Engineering*, vol. 2012, Article ID 502812, 26 pages, 2012.
[24] C. Cattani, G. Pierro, and G. Altieri, "Entropy and multifractality for the mye-loma multiple TET 2 gene," *Mathematical Problems in Engineering*, vol. 2012, Article ID 193761, 14 pages, 2012.
[25] M. S. Keshner, "$1/f$ noise," *Proceedings of the IEEE*, vol. 70, no. 3, pp. 212–218, 1982.
[26] B. Ninness, "Estimation of $1/f$ Noise," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 32–46, 1998.
[27] B. Yazici and R. L. Kashyap, "A class of second-order stationary self-similar processes for $1/f$ phenomena," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 396–410, 1997.
[28] G. W. Wornell, "Wavelet-based representations for the $1/f$ family of fractal processes," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1428–1450, 1993.
[29] B. B. Mandelbrot, "Some noises with $1/f$ spectrum, a bridge between direct current and white noise," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 289–298, 1967.
[30] N. J. Kasdin, "Discrete simulation of colored noise and stochastic processes and $1/f^{\alpha}$ power law noise generation," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 802–827, 1995.
[31] G. Corsini and R. Saletti, "$1/f^{\gamma}$ power spectrum noise sequence generator," *IEEE Transactions on Instrumentation and Measurement*, vol. 37, no. 4, pp. 615–619, 1988.
[32] W. T. Li and D. Holste, "Universal $1/f$ noise, crossovers of scaling exponents, and chromosome-specific patterns of guanine-cytosine content in DNA sequences of the human genome," *Physical Review E*, vol. 71, no. 4, Article ID 041910, 9 pages, 2005.
[33] W. T. Li, G. Stolovitzky, P. Bernaola-Galván, and J. L. Oliver, "Compositional heterogeneity within, and uniformity between, DNA sequences of yeast chromosomes," *Genome Research*, vol. 8, no. 9, pp. 916–928, 1998.

[34] W. T. Li and K. Kaneko, "Long-range correlation and partial spectrum in a noncoding DNA sequence," *Europhysics Letters*, vol. 17, no. 7, pp. 655–660, 1992.
[35] P. C. Ivanov, L. A. Nunes Amaral, A. L. Goldberger et al., "From $1/f$ noise to multifractal cascades in heartbeat dynamics," *Chaos*, vol. 11, no. 3, pp. 641–652, 2001.
[36] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge, UK, 2000.
[37] W. Q. Duan and H. E. Stanley, "Cross-correlation and the predictability of financial return series," *Physica A*, vol. 390, no. 2, pp. 290–296, 2010.
[38] B. Podobnik, D. Horvatic, A. Lam Ng, H. E. Stanley, and P. C. Ivanov, "Modeling long-range cross-correlations in two-component ARFIMA and FIARCH processes," *Physica A*, vol. 387, no. 15, pp. 3954–3959, 2008.
[39] G. Aquino, M. Bologna, P. Grigolini, and B. J. West, "Beyond the death of linear response: $1/f$ optimal information transport," *Physical Review Letters*, vol. 105, no. 4, Article ID 040601, 4 pages, 2010.
[40] B. J. West and P. Grigolini, "Chipping away at memory," *Biological Cybernetics*, vol. 103, no. 2, pp. 167–174, 2010.
[41] B. J. West and M. F. Shlesinger, "On the ubiquity of $1/f$ noise," *International Journal of Modern Physics B*, vol. 3, no. 6, pp. 795–819, 1989.
[42] A. L. Goldberger, V. Bhargava, B. J. West, and A. J. Mandell, "On the mechanism of cardiac electrical stability. The fractal hypothesis," *Biophysical Journal*, vol. 48, no. 3, pp. 525–528, 1985.
[43] T. Musha, H. Takeuchi, and T. Inoue, "$1/f$ fluctuations in the spontaneous spike discharge intervals of a giant snail neuron," *IEEE Transactions on Biomedical Engineering*, vol. 30, no. 3, pp. 194–197, 1983.
[44] M. Kobayashi and T. Musha, "$1/f$ fluctuation of heartbeat period," *IEEE Transactions on Biomedical Engineering*, vol. 29, no. 6, pp. 456–457, 1982.
[45] B. Neumcke, "$1/f$ noise in membranes," *Biophysics of Structure and Mechanism*, vol. 4, no. 3, pp. 179–199, 1978.
[46] J. R. Clay and M. F. Shlesinger, "Unified theory of $1/f$ and conductance noise in nerve membrane," *Journal of Theoretical Biology*, vol. 66, no. 4, pp. 763–773, 1977.
[47] E. Frehland, "Diffusion as a source of $1/f$ noise," *The Journal of Membrane Biology*, vol. 32, no. 1, pp. 195–196, 1977.
[48] M. E. Green, "Diffusion and $1/f$ noise," *The Journal of Membrane Biology*, vol. 28, no. 1, pp. 181–186, 1976.
[49] I. Csabai, "$1/f$ noise in computer network traffic," *Journal of Physics A*, vol. 27, no. 12, pp. L417–L421, 1994.
[50] M. Takayasu, H. Takayasu, and T. Sato, "Critical behaviors and $1/f$ noise in information traffic," *Physica A*, vol. 233, no. 3-4, pp. 824–834, 1996.
[51] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, 1995.
[52] W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker, "Scaling phenomena in the internet: critically examining criticality," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, supplement 1, pp. 2573–2580, 2002.
[53] P. Loiseau, P. Gonçalves, G. Dewaele, P. Borgnat, P. Abry, and P. V. B. Primet, "Investigating self-similarity and heavy-tailed distributions on a large-scale experimental facility," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1261–1274, 2010.
[54] P. Abry, P. Borgnat, F. Ricciato, A. Scherrer, and D. Veitch, "Revisiting an old friend: on the observability of the relation between long range dependence and heavy tail," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 147–165, 2010.
[55] A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions*, vol. 1, Springer, New York, NY, USA, 1987.
[56] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, USA, 1997.
[57] B. W. Lindgren and G. W. McElrath, *Introduction to Probability and Statistics*, The Macmillan, New York, NY, USA, 1959.
[58] J. L. Doob, "The elementary Gaussian processes," *Annals of Mathematical Statistics*, vol. 15, pp. 229–282, 1944.
[59] A. N. Kolmogorov, *Fundamental of Probability*, Business Press, Shanghai, China, 1954, Translated from Russian by S.-T. Ding.
[60] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs, NJ, USA, 1994.

[61] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedure*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, USA, 3rd edition, 2000.
[62] W. A. Fuller, *Introduction to Statistical Time Series*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2nd edition, 1996.
[63] J. Beran, *Statistics for Long-Memory Processes*, vol. 61 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, New York, NY, USA, 1994.
[64] J. Beran, "Statistical methods for data with long-range dependence," *Statistical Science*, vol. 7, no. 4, pp. 404–416, 1992.
[65] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics: Mechanics of Turbulence*, vol. 2, MIT Press, Cambridge, Mass, USA, 1971.
[66] A. N. Kolmogorov, "Local structure of turbulence in an incompressible viscous fluid at very high Reynolds numbers," *Soviet Physics Uspekhi*, vol. 10, no. 6, pp. 734–736, 1968.
[67] N. C. Nigam, *Introduction to Random Vibrations*, MIT Press, Cambridge, Mass, USA, 1983.
[68] T. T. Song and M. Grigoriu, *Random Vibration of Mechanical and Structural Systems*, Prentice Hall, New York, NY, USA, 1993.
[69] C. M. Harris, *Shock and Vibration Handbook*, McGraw-Hill, New York, NY, USA, 4th edition, 1995.
[70] H. Czichos, T. Saito, and L. Smith, *Springer Handbook of Metrology and Testing*, Springer, New York, NY, USA, 2011.
[71] W. N. Sharpe Jr., *Springer Handbook of Experimental Solid Mechanics*, Springer, New York, NY, USA, 2008.
[72] C. Tropea, A. L. Yarin, and J. F. Foss, Eds., *Springer Handbook of Experimental Fluid Mechanics*, Springer, New York, NY, USA, 2007.
[73] K. H. Grote and E. K. Antonsson, Eds., *Springer Handbook of Mechanical Engineering*, Springer, New York, NY, USA, 2009.
[74] R. Kramme, K. P. Hoffmann, and R. S. Pozos, *Springer Handbook of Medical Technology*, Springer, New York, NY, USA, 2012.
[75] W. Kresse and D. M. Danko, *Springer Handbook of Geographic Information*, Springer, New York, NY, USA, 2012.
[76] S. S. Bhattacharyya, F. Deprettere, R. Leupers, and J. Takala, Eds., *Handbook of Signal Processing Systems*, Springer, New York, NY, USA, 2010.
[77] S. K. Mitra and J. F. Kaiser, *Handbook for Digital Signal Processing*, John Wiley & Sons, New York, NY, USA, 1993.
[78] W. A. Woyczyński, *A First Course in Statistics for Signal Analysis*, Birkhäuser, Boston, Mass, USA, 2006.
[79] R. A. Bailey, *Design of Comparative Experiments*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK, 2008.
[80] ASME, *Measurement Uncertainty Part 1, Instruments and Apparatus*, Sup-plement to ASME, Performance Test Codes, ASME, New York, NY, USA, 1986.
[81] D. Sheskin, *Statistical Tests and Experimental Design: A Guidebook*, Gardner Press, New York, NY, USA, 1984.
[82] T. W. MacFarland, *Two-Way Analysis of Variance*, Springer, New York, NY, USA, 2012.
[83] A. K. Gupta, W. B. Zeng, and Y. Wu, *Probability and Statistical Models: Foundations for Problems in Reliability and Financial Mathematics*, Birkhäuser, Boston, Mass, USA, 2010.
[84] P. Fieguth, *Statistical Image Processing and Multidimensional Modeling*, Information Science and Statistics, Springer, New York, NY, USA, 2011.
[85] J. Nauta, *Statistics in Clinical Vaccine Trials*, Springer, New York, NY, USA, 2011.
[86] A. Gelman, "Analysis of variance—why it is more important than ever," *The Annals of Statistics*, vol. 33, no. 1, pp. 1–53, 2005.
[87] C. G. Pendse, "A note on mathematical expectation," *The Mathematical Gazette*, vol. 22, no. 251, pp. 399–402, 1938.
[88] B. B. Mandelbrot, *Gaussian Self-Affinity and Fractals*, Springer, New York, NY, USA, 2001.
[89] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*, Stochastic Modeling, Chapman & Hall, New York, NY, USA, 1994.
[90] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 1566–1579, 1995.
[91] I. M. Gelfand and K. Vilenkin, *Generalized Functions*, vol. 1, Academic Press, New York, NY, USA, 1964.
[92] M. Li and S. C. Lim, "Power spectrum of generalized Cauchy process," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 219–222, 2010.

[93] M. Li and S. C. Lim, "A rigorous derivation of power spectrum of fractional Gaussian noise," *Fluctuation and Noise Letters*, vol. 6, no. 4, pp. C33–C36, 2006.
[94] B. B. Mandelbrot and J. W. van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, pp. 422–437, 1968.
[95] P. Flandrin, "On the spectrum of fractional Brownian motions," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 197–199, 1989.
[96] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
[97] S. V. Muniandy and S. C. Lim, "Modeling of locally self-similar processes using multifractional Brownian motion of Riemann-Liouville type," *Physical Review E*, vol. 63, no. 4, Article ID 046104, 7 pages, 2001.
[98] V. M. Sithi and S. C. Lim, "On the spectra of Riemann-Liouville fractional Brownian motion," *Journal of Physics A*, vol. 28, no. 11, pp. 2995–3003, 1995.
[99] S. C. Lim and S. V. Muniandy, "On some possible generalizations of fractional Brownian motion," *Physics Letters A*, vol. 266, no. 2-3, pp. 140–145, 2000.
[100] J. P. Chilès and P. Delfiner, *Geostatistics, Modeling Spatial Uncertainty*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1999.
[101] M. Li, C. Cattani, and S. Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 13 pages, 2011.
[102] S. C. Lim and M. Li, "A generalized Cauchy process and its application to relaxation phenomena," *Journal of Physics A*, vol. 39, no. 12, pp. 2935–2951, 2006.
[103] S. C. Lim and L. P. Teo, "Gaussian fields and Gaussian sheets with generalized Cauchy covariance structure," *Stochastic Processes and Their Applications*, vol. 119, no. 4, pp. 1325–1356, 2009.
[104] M. Li and S. C. Lim, "Modeling network traffic using generalized Cauchy process," *Physica A*, vol. 387, no. 11, pp. 2584–2594, 2008.
[105] P. Vengadesh, S. V. Muniandy, and W. H. A. Majid, "Fractal morphological analysis of bacteriorhodopsin (bR) layers deposited onto indium tin oxide (ITO) electrodes," *Materials Science and Engineering C*, vol. 29, no. 5, pp. 1621–1626, 2009.
[106] R. J. Martin and A. M. Walker, "A power-law model and other models for long-range dependence," *Journal of Applied Probability*, vol. 34, no. 3, pp. 657–670, 1997.
[107] R. J. Martin and J. A. Eccleston, "A new model for slowly-decaying correlations," *Statistics and Probability Letters*, vol. 13, no. 2, pp. 139–145, 1992.
[108] W. A. Woodward, Q. C. Cheng, and H. L. Gray, "A k-factor GARMA long-memory model," *Journal of Time Series Analysis*, vol. 19, no. 4, pp. 485–504, 1998.
[109] C. Ma, "Power-law correlations and other models with long-range dependence on a lattice," *Journal of Applied Probability*, vol. 40, no. 3, pp. 690–703, 2003.
[110] C. Ma, "A class of stationary random fields with a simple correlation structure," *Journal of Multivariate Analysis*, vol. 94, no. 2, pp. 313–327, 2005.
[111] M. Li, W. Jia, and W. Zhao, "Correlation form of timestamp increment sequences of self-similar traffic on Ethernet," *Electronics Letters*, vol. 36, no. 19, pp. 1668–1669, 2000.
[112] M. Li and W. Zhao, "Quantitatively investigating locally weak stationarity of modified multifractional Gaussian noise," *Physica A*, vol. 391, no. 24, pp. 6268–6278, 2012.
[113] E. G. Tsionas, "Estimating multivariate heavy tails and principal directions easily, with an application to international exchange rates," *Statistics and Probability Letters*, vol. 82, no. 11, pp. 1986–1989, 2012.
[114] J. Lin, "Second order asymptotics for ruin probabilities in a renewal risk model with heavy-tailed claims," *Insurance: Mathematics and Economics*, vol. 51, no. 2, pp. 422–429, 2012.
[115] K. Yu, M. L. Huang, and P. H. Brill, "An algorithm for fitting heavy-tailed distributions via generalized hyperexponentials," *INFORMS Journal on Computing*, vol. 24, no. 1, pp. 42–52, 2012.
[116] R. Luger, "Finite-sample bootstrap inference in GARCH models with heavy-tailed innovations," *Computational Statistics and Data Analysis*, vol. 56, no. 11, pp. 3198–3211, 2012.
[117] T. Ishihara and Y. Omori, "Efficient Bayesian estimation of a multivariate stochastic volatility model with cross leverage and heavy-tailed errors," *Computational Statistics and Data Analysis*, 2010.
[118] J. Diebolt, L. Gardes, S. Girard, and A. Guillou, "Bias-reduced extreme quantile estimators of Weibull tail-distributions," *Journal of Statistical Planning and Inference*, vol. 138, no. 5, pp. 1389–1401, 2008.
[119] J. Beran, B. Das, and D. Schell, "On robust tail index estimation for linear long-memory processes," *Journal of Time Series Analysis*, vol. 33, no. 3, pp. 406–423, 2012.

[120] P. Barbe and W. P. McCormick, "Heavy-traffic approximations for fractionally integrated random walks in the domain of attraction of a non-Gaussian stable distribution," *Stochastic Processes and Their Applications*, vol. 122, no. 4, pp. 1276–1303, 2012.
[121] C. Weng and Y. Zhang, "Characterization of multivariate heavy-tailed distribution families via copula," *Journal of Multivariate Analysis*, vol. 106, pp. 178–186, 2012.
[122] C. B. García, J. García Pérez, and J. R. van Dorp, "Modeling heavy-tailed, skewed and peaked uncertainty phenomena with bounded support," *Statistical Methods and Applications*, vol. 20, no. 4, pp. 146–166, 2011.
[123] V. H. Lachos, T. Angolini, and C. A. Abanto-Valle, "On estimation and local influence analysis for measurement errors models under heavy-tailed distributions," *Statistical Papers*, vol. 52, no. 3, pp. 567–590, 2011.
[124] V. Ganti, K. M. Straub, E. Foufoula-Georgiou, and C. Paola, "Space-time dynamics of depositional systems: experimental evidence and theoretical modeling of heavy-tailed statistics," *Journal of Geophysical Research F: Earth Surface*, vol. 116, no. 2, Article ID F02011, 17 pages, 2011.
[125] U. J. Dixit and M. J. Nooghabi, "Efficient estimation in the Pareto distribution with the presence of outliers," *Statistical Methodology*, vol. 8, no. 4, pp. 340–355, 2011.
[126] P. Nándori, "Recurrence properties of a special type of heavy-tailed random walk," *Journal of Statistical Physics*, vol. 142, no. 2, pp. 342–355, 2011.
[127] D. Ceresetti, G. Molinié, and J. D. Creutin, "Scaling properties of heavy rainfall at short duration: a regional analysis," *Water Resources Research*, vol. 46, no. 9, Article ID W09531, 12 pages, 2010.
[128] A. Charpentier and A. Oulidi, "Beta kernel quantile estimators of heavy-tailed loss distributions," *Statistics and Computing*, vol. 20, no. 1, pp. 35–55, 2010.
[129] P. Embrechts, J. Nešlehová, and M. V. Wüthrich, "Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness," *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 164–169, 2009.
[130] I. F. Alves, L. de Haan, and C. Neves, "A test procedure for detecting super-heavy tails," *Journal of Statistical Planning and Inference*, vol. 139, no. 2, pp. 213–227, 2009.
[131] J. Beirlant, E. Joossens, and J. Segers, "Second-order refined peaks-over-threshold modelling for heavy-tailed distributions," *Journal of Statistical Planning and Inference*, vol. 139, no. 8, pp. 2800–2815, 2009.
[132] R. Ibragimov, "Heavy-tailedness and threshold sex determination," *Statistics and Probability Letters*, vol. 78, no. 16, pp. 2804–2810, 2008.
[133] R. Delgado, "A reflected fBm limit for fluid models with ON/OFF sources under heavy traffic," *Stochastic Processes and Their Applications*, vol. 117, no. 2, pp. 188–201, 2007.
[134] M. S. Taqqu, "The modelling of ethernet data and of signals that are heavy-tailed with infinite variance," *Scandinavian Journal of Statistics*, vol. 29, no. 2, pp. 273–295, 2002.
[135] B. G. Lindsay, J. Kettenring, and D. O. Siegmund, "A report on the future of statistics," *Statistical Science*, vol. 19, no. 3, pp. 387–413, 2004.
[136] S. Resnick, "On the foundations of multivariate heavy-tail analysis," *Journal of Applied Probability*, vol. 41, pp. 191–212, 2004.
[137] S. Resnick and H. Rootzén, "Self-similar communication models and very heavy tails," *The Annals of Applied Probability*, vol. 10, no. 3, pp. 753–778, 2000.
[138] J. Cai and Q. Tang, "On max-sum equivalence and convolution closure of heavy-tailed distributions and their applications," *Journal of Applied Probability*, vol. 41, no. 1, pp. 117–130, 2004.
[139] V. Limic, "A LIFO queue in heavy traffic," *The Annals of Applied Probability*, vol. 11, no. 2, pp. 301–331, 2001.
[140] H. Le and A. O. 'Hagan, "A class of bivariate heavy-tailed distributions," *San-Khyā: The Indian Journal of Statistics, Series B*, vol. 60, no. 1, pp. 82–100, 1998.
[141] M. C. Bryson, "Heavy-tailed distributions: properties and tests," *Technometrics*, vol. 16, no. 1, pp. 61–68, 1974.
[142] J. Beran, "Discussion: heavy tail modeling and teletraffic data," *The Annals of Statistics*, vol. 25, no. 5, pp. 1852–1856, 1997.
[143] S. Ahn, J. H. T. Kim, and V. Ramaswami, "A new class of models for heavy tailed distributions in finance and insurance risk," *Insurance: Mathematics and Economics*, vol. 51, no. 1, pp. 43–52, 2012.
[144] V. Pisarenko and M. Rodkin, *Heavy-Tailed Distributions in Disaster Analysis*, Springer, New York, NY, USA, 2010.
[145] S. I. Resnick, *Heavy-Tail Phenomena Probabilistic and Statistical Modeling*, Springer, New York, NY, USA, 2007, Probabilistic and statistical modeling.

[146] R. J. Adler, R. E. Feldman, and M. S. Taqqu, Eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*,, Birkhäuser, Boston, Mass, USA, 1998.

[147] M. Li and W. Zhao, "Visiting power laws in cyber-physical networking systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 302786, 13 pages, 2012.

[148] L. Xu, P. C. Ivanov, K. Hu, Z. Chen, A. Carbone, and H. E. Stanley, "Quantifying signals with power-law correlations: a comparative study of detrended fluctuation analysis and detrended moving average techniques," *Physical Review E*, vol. 71, no. 5, Article ID 051101, 14 pages, 2005.

[149] M. Li and J. Y. Li, "On the predictability of long-range dependent series," *Mathematical Problems in Engineering*, vol. 2010, Article ID 397454, 9 pages, 2010.

[150] W. Hürlimann, "From the general affine transform family to a Pareto type IV model," *Journal of Probability and Statistics*, vol. 2009, Article ID 364901, 10 pages, 2009.

[151] A. André, "Limit theorems for randomly selected adjacent order statistics from a Pareto distribution," *International Journal of Mathematics and Mathematical Sciences*, vol. 2005, no. 21, pp. 3427–3441, 2005.

[152] H. E. Stanley, "Power laws and universality," *Nature*, vol. 378, no. 6557, p. 554, 1995.

[153] I. Eliazar and J. Klafter, "A probabilistic walk up power laws," *Physics Reports*, vol. 511, no. 3, pp. 143–175, 2012.

[154] A. R. Bansal, G. Gabriel, and V. P. Dimri, "Power law distribution of susceptibility and density and its relation to seismic properties: an example from the German Continental Deep Drilling Program (KTB)," *Journal of Applied Geophysics*, vol. 72, no. 2, pp. 123–128, 2010.

[155] S. Milojević, "Power law distributions in information science: making the case for logarithmic binning," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2417–2425, 2010.

[156] Y. Wu, Q. Ye, J. Xiao, and L. X. Li, "Modeling and statistical properties of human view and reply behavior in on-line society," *Mathematical Problems in Engineering*, vol. 2012, Article ID 969087, 7 pages, 2012.

[157] A. Fujihara, M. Uchida, and H. Miwa, "Universal power laws in the threshold network model: a theoretical analysis based on extreme value theory," *Physica A*, vol. 389, no. 5, pp. 1124–1130, 2010.

[158] A. Saiz, "Boltzmann power laws," *Physica A*, vol. 389, no. 2, pp. 225–236, 2010.

[159] A. Jaishankar and G. H. McKinley, "Power-law rheology in the bulk and at the interface: quasi-properties and fractional constitutive equations," *Proceedings of the Royal Society of London Series A*, vol. 469, no. 2149, Article ID 20120284, 2013.

[160] X. Zhao, P. J. Shang, and Y. L. Pang, "Power law and stretched exponential effects of extreme events in Chinese stock markets," *Fluctuation and Noise Letters*, vol. 9, no. 2, pp. 203–217, 2010.

[161] P. Kokoszka and T. Mikosch, "The integrated periodogram for long-memory processes with finite or infinite variance," *Stochastic Processes and Their Applications*, vol. 66, no. 1, pp. 55–78, 1997.

[162] D. Belomestny, "Spectral estimation of the Lévy density in partially observed affine models," *Stochastic Processes and Their Applications*, vol. 121, no. 6, pp. 1217–1244, 2011.

[163] T. Simon, "Fonctions de Mittag-Leffler et processus de Lévy stables sans sauts négatifs," *Expositiones Mathematicae*, vol. 28, no. 3, pp. 290–298, 2010.

[164] R. Lambiotte and L. Brenig, "Truncated Lévy distributions in an inelastic gas," *Physics Letters A*, vol. 345, no. 4–6, pp. 309–313, 2005.

[165] G. Terdik, W. A. Woyczynski, and A. Piryatinska, "Fractional- and integer-order moments, and multiscaling for smoothly truncated Lévy flights," *Physics Letters A*, vol. 348, no. 3–6, pp. 94–109, 2006.

[166] I. Koponen, "Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process," *Physical Review E*, vol. 52, no. 1, pp. 1197–1199, 1995.

[167] J. Behboodian, A. Jamalizadeh, and N. Balakrishnan, "A new class of skew-Cauchydistributions," *Statistics and Probability Letters*, vol. 76, no. 14, pp. 1488–1493, 2006.

[168] P. Garbaczewski, "Cauchy flights in confining potentials," *Physica A*, vol. 389, no. 5, pp. 936–944, 2010.

[169] A. J. Field, U. Harder, and P. G. Harrison, "Measurement and modelling of self-similar traffic in computer networks," *IEE Proceedings-Communications*, vol. 151, no. 4, pp. 355–363, 2004.

[170] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, McGraw-Hill, New York, NY, USA, 1961.

[171] H. Konno and Y. Tamura, "A generalized Cauchy process having cubic non-linearity," *Reports on Mathematical Physics*, vol. 67, no. 2, pp. 179–195, 2011.

[172] H. Konno and F. Watanabe, "Maximum likelihood estimators for generalized Cauchy processes," *Journal of Mathematical Physics*, vol. 48, no. 10, Article ID 103303, 19 pages, 2007.

[173] I. A. Lubashevsky, "Truncated Lévy flights and generalized Cauchy processes," *European Physical Journal B*, vol. 82, no. 2, pp. 189–195, 2011.

[174] Y. Liang and W. Chen, "A survey on computing Lévy stable distributions and a new MATLAB toolbox," *Signal Processing*, vol. 93, no. 1, pp. 244–251, 2013.

[175] G. Terdik and T. Gyires, "Lévy flights and fractal modeling of internet traffic," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 120–129, 2009.

[176] E. E. Kuruoğlu, "Density parameter estimation of skewed α-stable distributions," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2192–2201, 2001.

[177] A. P. Petropulu, J. C. Pesquet, X. Yang, and J. J. Yin, "Power-law shot noise and its relationship to long-memory α-stable processes," *IEEE Transactions on Signal Processing*, vol. 48, no. 7, pp. 1883–1892, 2000.

[178] S. Cohen and G. Samorodnitsky, "Random rewards, fractional brownian local times and stable self-similar processes," *The Annals of Applied Probability*, vol. 16, no. 3, pp. 1442–1461, 2006.

[179] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986–1010, 1993.

[180] L. Landau, "On the energy loss of fast particles by ionization," *Journal of Physics*, vol. 8, pp. 201–205, 1944.

[181] D. H. Wilkinson, "Ionization energy loss by charged particles part I. The Landau distribution," *Nuclear Instruments and Methods in Physics Research A*, vol. 383, no. 2-3, pp. 513–515, 1996.

[182] T. Tabata and R. Ito, "Approximations to Landau's distribution functions for the ionization energy loss of fast electrons," *Nuclear Instruments and Methods*, vol. 158, pp. 521–523, 1979.

[183] J. Holtsmark, "Uber die Verbreiterung von Spektrallinien," *Annalen der Physik*, vol. 363, no. 7, pp. 577–630, 1919.

[184] B. Pittel, W. A. Woyczynski, and J. A. Mann, "Random tree-type partitions as a model for acyclic polymerization: holtsmark (3/2-stable) distribution of the supercritical gel," *The Annals of Probability*, vol. 18, no. 1, pp. 319–341, 1990.

[185] D. G. Hummer, "Rational approximations for the holtsmark distribution, its cumulative and derivative," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 36, no. 1, pp. 1–5, 1986.

[186] R. G. Garroppo, S. Giordano, M. Pagano, and G. Procissi, "Testing α-stable processes in capturing the queuing behavior of broadband teletraffic," *Signal Processing*, vol. 82, no. 12, pp. 1861–1872, 2002.

[187] J. R. Gallardo, D. Makrakis, and L. Orozco-Barbosa, "Use of α-stable self-similar stochastic processes for modeling traffic in broadband networks," *Performance Evaluation*, vol. 40, no. 1, pp. 71–98, 2000.

[188] A. Karasaridis and D. Hatzinakos, "Network heavy traffic modeling using α-stable self-similar processes," *IEEE Transactions on Communications*, vol. 49, no. 7, pp. 1203–1214, 2001.

[189] P. R. de Montmort, "Essay d'analyse sur les jeux de hazard," 1713.

[190] P. R. de Montmort, "Essay d'analyse sur les jeux de hazard," American Mathematical Society, 1980.

[191] D. Bernoulli, "Exposi-tion of a new theory on the measurement of risk," *Econometrica*, vol. 22, no. 1, pp. 22–36, 1954.

[192] W. T. Coffey, Y. P. Kalmykov, and J. T. Waldron, *The Langevin Equation*, World Scientific, Singapore, 2nd edition, 2004.

[193] S. F. Kwok, "Langevin equation with multiplicative white noise: transfor-mation of diffusion processes into the Wiener process in different prescrip-tions," *Annals of Physics*, vol. 327, no. 8, pp. 1989–1997, 2012.

[194] A. V. Medino, S. R. C. Lopes, R. Morgado, and C. C. Y. Dorea, "Generalized Langevin equation driven by Lévy processes: a probabilistic, numerical and time series based approach," *Physica A*, vol. 391, no. 3, pp. 572–581, 2012.

[195] D. Panja, "Generalized langevin equation formulation for anomalous polymer dynamics," *Journal of Statistical Mechanics*, vol. 2010, no. 2, Article ID L02001, 2010.

[196] A. Bazzani, G. Bassi, and G. Turchetti, "Diffusion and memory effects for stochastic processes and fractional Langevin equations," *Physica A*, vol. 324, no. 3-4, pp. 530–550, 2003.

[197] E. Lutz, "Fractional Langevin equation," *Physical Review E*, vol. 64, no. 5, Article ID 051106, 4 pages, 2001.

[198] M. G. McPhie, P. J. Daivis, I. K. Snook, J. Ennis, and D. J. Evans, "Generalized Langevin equation for nonequilibrium systems," *Physica A*, vol. 299, no. 3-4, pp. 412–426, 2001.

[199] K. S. Fa, "Fractional Langevin equation and Riemann-Liouville fractional derivative," *European Physical Journal E*, vol. 24, no. 2, pp. 139–143, 2007.

[200] B. Ahmad, J. J. Nieto, A. Alsaedi, and M. El-Shahed, "A study of nonlinear Langevin equation involving two fractional orders in different intervals," *Non-Linear Analysis: Real World Applications*, vol. 13, no. 2, pp. 599–606, 2012.

[201] B. Ahmad and J. J. Nieto, "Solvability of nonlinear Langevin equation involving two fractional orders with Dirichlet boundary conditions," *International Journal of Differential Equations*, vol. 2012, Article ID 649486, 10 pages, 2010.

[202] S. C. Kou and X. S. Xie, "Generalized langevin equation with fractional gaussian noise: subdiffusion within a single protein molecule," *Physical Review Letters*, vol. 93, no. 18, Article ID 180603, 4 pages, 2004.

[203] H. C. Fogedby, "Langevin equations for continuous time Lévy flights," *Physical Review E*, vol. 50, no. 2, pp. 1657–1660, 1994.

[204] Y. Fukui and T. Morita, "Derivation of the stationary generalized Langevin equation," *Journal of Physics A*, vol. 4, no. 4, pp. 477–490, 1971.

[205] S. C. Kou, "Stochastic modeling in nanoscale biophysics: subdiffusion within proteins," *The Annals of Applied Statistics*, vol. 2, no. 2, pp. 501–535, 2008.

[206] V. V. Anh, C. C. Heyde, and N. N. Leonenko, "Dynamic models of long-memory processes driven by Lévy noise," *Journal of Applied Probability*, vol. 39, no. 4, pp. 730–747, 2002.

[207] B. N. N. Achar, J. W. Hanneken, and T. Clarke, "Damping characteristics of a fractional oscillator," *Physica A*, vol. 339, no. 3-4, pp. 311–319, 2004.

[208] B. N. N. Achar, J. W. Hanneken, and T. Clarke, "Response characteristics of a fractional oscillator," *Physica A*, vol. 309, no. 3-4, pp. 275–288, 2002.

[209] C. H. Eab and S. C. Lim, "Fractional generalized Langevin equation approach to single-file diffusion," *Physica A*, vol. 389, no. 13, pp. 2510–2521, 2010.

[210] C. H. Eab and S. C. Lim, "Fractional Langevin equations of distributed order," *Physical Review E*, vol. 83, no. 3, Article ID 031136, 10 pages, 2011.

[211] S. C. Lim and L. P. Teo, "Modeling single-file diffusion with step fractional Brownian motion and a generalized fractional Langevin equation," *Journal of Statistical Mechanics*, vol. 2009, no. 8, Article ID P08015, 2009.

[212] S. C. Lim, L. Ming, and L. P. Teo, "Locally self-similar fractional oscillator processes," *Fluctuation and Noise Letters*, vol. 7, no. 2, pp. L169–L179, 2007.

[213] M. Li, S. C. Lim, and S. Chen, "Exact solution of impulse response to a class of fractional oscillators and its stability," *Mathematical Problems in Engineering*, vol. 2011, Article ID 657839, 9 pages, 2011.

[214] S. C. Lim, C. H. Eab, K. H. Mak, M. Li, and S. Chen, "Solving linear coupled fractional differential equations and their applications," *Mathematical Problems in Engineering*, vol. 2012, Article ID 653939, 28 pages, 2012.

[215] S. C. Lim, M. Li, and L. P. Teo, "Langevin equation with two fractional orders," *Physics Letters A*, vol. 372, no. 42, pp. 6309–6320, 2008.

[216] R. A. Gabel and R. A. Roberts, *Signals and Linear Systems*, John Wiley & Sons, New York, NY, USA, 1973.

[217] M. Carlini, T. Honorati, and S. Castellucci, "Photovoltaic greenhouses: comparison of optical and thermal behaviour for energy savings," *Mathematical Problems in Engineering*, vol. 2012, Article ID 743764, 10 pages, 2012.

[218] M. Carlini and S. Castellucci, "Modelling the vertical heat exchanger in thermal basin," in *Proceedings of the ICCSA, 2011, Part 4*, vol. 6785 of *Springer Lecture Notes in Computer Science*, pp. 277–286, Springer, New York, NY, USA, 2011.

[219] M. Carlini, C. Cattani, and A. Tucci, "Optical modelling of square solar con-centrator," in *Proceedins of the ICCSA, 2011, Part 4*, vol. 6785 of *Springer Lecture Notes in Computer Science*, pp. 287–295, Springer, New York, NY, USA, 2011.

[220] L. Qiu, B. G. Xu, and S. B. Li, "H_2/H_∞ control of networked control system with random time delays," *Science China Information Sciences*, vol. 54, no. 12, pp. 2615–2630, 2011.

[221] J. Li, J. Z. Wang, S. K. Wang, L. L. Ma, and W. Shen, "Dynamic image stabilization precision test system based on the Hessian matrix," *Science China Information Sciences*, vol. 55, no. 9, pp. 2056–2074, 2012.

[222] W. X. Zhao and H. F. Chen, "Markov chain approach to identifying Wiener systems," *Science China Information Sciences*, vol. 55, no. 5, pp. 1201–1217, 2012.
[223] K. S. Miller and B. Ross, *An Introduction to the Fractional Calculus and Fractional Differential Equations*, John Wiley & sons, NewYork, NY, USA, 1993.
[224] T. Hida, *Brownian Motion*, Springer, New York, NY, USA, 1980.
[225] A. H. Zemanian, "An introduction to generalized functions and the generalized Laplace and Legendre transformations," *SIAM Review*, vol. 10, no. 1, pp. 1–24, 1968.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 879614, 19 pages
doi:10.1155/2012/879614

Research Article

Multiobjective Quantum Evolutionary Algorithm for the Vehicle Routing Problem with Customer Satisfaction

Jingling Zhang,[1] Wanliang Wang,[2] Yanwei Zhao,[1] and Carlo Cattani[3]

[1] *College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou 310014, China*
[2] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310014, China*
[3] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

Correspondence should be addressed to Yanwei Zhao, zyw@zjut.edu.cn

Received 20 August 2012; Accepted 10 October 2012

Academic Editor: Sheng-yong Chen

The multiobjective vehicle routing problem considering customer satisfaction (MVRPCS) involves the distribution of orders from several depots to a set of customers over a time window. This paper presents a self-adaptive grid multi-objective quantum evolutionary algorithm (MOQEA) for the MVRPCS, which takes into account customer satisfaction as well as travel costs. The degree of customer satisfaction is represented by proposing an improved fuzzy due-time window, and the optimization problem is modeled as a mixed integer linear program. In the MOQEA, nondominated solution set is constructed by the Challenge Cup rules. Moreover, an adaptive grid is designed to achieve the diversity of solution sets; that is, the number of grids in each generation is not fixed but is automatically adjusted based on the distribution of the current generation of nondominated solution set. In the study, the MOQEA is evaluated by applying it to classical benchmark problems. Results of numerical simulation and comparison show that the established model is valid and the MOQEA is effective for MVRPCS.

1. Introduction

The vehicle routing problem (VRP) is one of the most important and widely studied combinatorial optimization problems, with many real-world applications in logistic distribution and transportation [1]. Since the VRP was firstly proposed by Dantzig and Ramser in 1959 [2], it has been focused in the field of operational research and combinatorial optimization [3–5].

The aim of VRP is to find optimal routes for a fleet of vehicles serving a set of customers with known demands. Each customer is serviced exactly once and must be assigned a

satisfactory vehicle without exceeding vehicle capacities. A solution for this problem is to find out a set of minimum cost routes that are used to represent vehicles distribution and clients' permutation. However, current studies on VRP [6, 7] are mainly focused on the single objective problem and the objective is to optimize the number of vehicles dispatched and the travel distance, that is, reducing the service costs of the provider.

Actually, to achieve competitive advantage, a service provider needs to consider not only service costs but also service quality that can determine customers' satisfaction. Most of the research on multiobjective VRP (MOVRP) does not take into account this objective, only focusing on the traditional objectives of minimum costs and the length of the longest route [8, 9]. Hong and Park [10] constructed a linear goal programming (GP) model for the biobjective vehicle routing with time window constraints (BVRPTW) and proposed a heuristic algorithm to relieve the computational burden inherent to the application of the GP model. Zitzler and Thiele [11] proposed a multiobjective evolutionary algorithm based on the Pareto approach for VRP. Lim and wang [12] proposed a method to deal with MOVRP by assigning different weights of the objectives. Tan et al. [13] proposed a hybrid multiobjective evolutionary algorithm (HMOEA) that incorporates various heuristics for local exploitation in the evolutionary search and the concept of Pareto's optimality for solving multiobjective optimization in vehicle routing problem with time window constraints (VRPTW). Garcia-Najera and Bullinaria [14] studied an improved multiobjective evolutionary algorithm for VRP with time windows.

The VRPTW is developed from VRP and has been widely studied in the last decade [15–19]. In the VRPTW, each customer has a time window with values about the deadline and the earliest time constraints for the service he/she requires. Thus this problem involves a routing combination and scheduling component. Routes must be designed to minimize the total cost, but, at the same time, scheduling must be performed to ensure time feasibility.

In practice, this time window actually does not well describe customers' satisfaction. A major reason is that customers are asked to provide a fixed time window for service, but in reality they really hope to be served at a desired time. Cheng and Gen [20, 21] called such a desired time the due-time and proposed to use the concept of fuzzy due-time to replace this time window because, as they claimed, it can describe customers' satisfaction better. Thus customers' satisfaction can be also described as a convex fuzzy number [22–24].

Cheng and Gen [20, 21] introduced the fuzzy due-time, used triangle fuzzy numbers to describe customers' satisfaction, and solved the VRP by genetic algorithms (GAs). Zhang et al. [25] proposed a multiobjective fuzzy VRP and used trapezoid fuzzy numbers to describe customers' satisfaction. Jia [26] used multiobjective hybrid GA for this problem. Wu [27] studied the open VRP based on customers' satisfaction. Lin [28] proposed a GA-based multiobjective decision-making method for optimal vehicle transportation, which is focused on a fuzzy vehicle routing and scheduling problem (FVRSP) based on five attributes, namely, space utility, service satisfaction, waiting time, delay time, and transportation distance. Wang and Li [29] proposed a hybrid algorithm based on GA and incorporated some methods based on greedy algorithms to solve the MOP model for which depot desires and clients expectations are considered simultaneously.

The above studies use the weighted sums of objectives to solve the multiobjective problem; the higher an objective's importance, the larger its corresponding weight coefficient. In general, no single solution can attain the optimum of all objectives at the same time. Therefore, it is desirable to obtain a set of Pareto optimal solution, that is, the Pareto set. The points in the objective space that correspond to the results in the set are usually called the Pareto front.

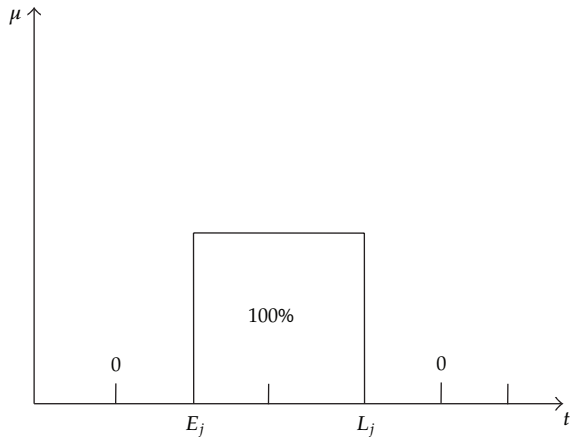


Figure 1: Traditional time windows.

In this paper, a self-adaptive grid multiobjective quantum evolutionary algorithm (MOQEA) is proposed to solve the MVRPCS problem. In particular, the quantum evolutionary algorithm (QEA) is used in the MOQEA due to its high efficiency, convergence speed, strong full-searching optimization ability [30]. With the MOQEA, an optimal or a nearly optimal set of vehicle routes solution with the minimal total travel cost and maximal customers' satisfaction can be obtained by decoding the chromosome and simultaneously obtains several solution sets. This method can support the dispatcher to more efficiently determine how to distribute the shipment to serve customers by available vehicles.

The remainder of the paper is organized as follows. Section 2 presents the mathematical model for the MVRPCS with consideration of customer satisfaction. Section 3 describes the proposed MOQEA in detail. In Section 4, the application of the proposed algorithm to a classic problem is introduced and the simulation results are discussed and compared with an early algorithm. Finally, conclusions are given in Section 5.

2. Model for the MVRPCS

2.1. Representation of Customer Satisfaction

In traditional VRP, customers' time constraints are represented by time windows as shown in Figure 1. For this method, if the customer is serviced at a time within the window, then the satisfaction degree is 100%, otherwise the satisfaction degree is 0. It is really unrealistic to measure the degree of satisfaction accordingly based on time window because customer satisfaction is not necessarily the same if they are serviced at different times within the window. In fact, service time can be divided into two categories; namely, the service time can be tolerated and the desirable service time.

Fuzzy due-time windows have been introduced to describe different degrees of satisfaction. Generally, the tolerable service time for customer i can be described as $[E_i, L_i]$, where E_i is the earliest time and L_i is the latest time. For example, in [20, 25] customers satisfaction is represented by the triangular fuzzy number, as shown in Figure 2. When the desirable service time is taken into account, customer satisfaction can be described using trapezoidal fuzzy number. In [26], this method is used and the desirable service time is

described by the due-time $[a_i, b_i]$, as shown in Figure 3. In this case, customer satisfaction is zero no matter vehicles arrive early or late if the expected service time slot is not achieved.

In this paper, an improved fuzzy due-time window is proposed, as shown in Figure 4. In this method, if a customer is served at the desired time, the degree of satisfaction is 1; otherwise, the degree of satisfaction decreases as the difference between the actual service time and the desired time increases. The degree of satisfaction for the cases in which vehicles arrive early than the earliest expected time is not zero, but equals to that of the case when the earliest expected time is met. The degree of customer satisfaction is represented by the membership function of the improved fuzzy due-time window, that is, an improved trapezoidal fuzzy number. For customer i, mark his/her satisfaction as $\mu_i(t_i)$ for a given service time t_i. Then the degree of satisfaction can be calculated using the membership function as follows:

$$\mu_i(t_i) = \begin{cases} \exp k(E_i - a_i), & t_i < E_i \\ \exp k(t_i - a_i), & E_i < t_i < a_i \\ 1, & a_i < t_i < b_i \\ \exp k(b_i - t_i), & b_i < t_i < L_i \\ 0, & t_i > L_i. \end{cases} \tag{2.1}$$

2.2. Mathematics Model

The MVRPCS can be described as follows: there are M depots each of which has $K_m (m = 1, 2, \ldots, M)$ vehicles with a capacity of b_k. These vehicles will be dispatched to L customers to meet their demands.

Mark the demand of customer i as $d_i (i = 1, 2, \ldots, L)$, and assume $d_i < b_k$. Each customer can be served by any vehicles from a depot, but the service will only take place once. In addition, each vehicle can complete the shipping task without having to return to the original depot. Thus an appropriate vehicle scheduling program is required to meet the needs of all customers. The meanings of variables used in this research are described as follows.

Customer number is $1, 2, \ldots, L$. Depot number is $L + 1, L + 2, \ldots, L + M$.

Fixed cost of sending a vehicle is $F_k (m \in \{L + 1, L + 2, \ldots, L + M\}; k \in \{1, 2, \ldots, K_m\})$.

Distribution cost from customer i to customer j is c_{ij}, and $i, j \in (1, 2, \ldots, L, L+1, \ldots, L+M)$.

Time window of customer i is $[E_i, L_i]$; the arrival time of customer i is t_i; the travel time from i to j is t_{ij}, s_i is the service time of customer i and $\mu_i(t_i)$ is the degree of satisfaction for customer i.

If vehicle k travels directly from customer i to customer j and arrives too early at j, it will wait; $w_j(t_j)$ is the waiting time of customer j for a vehicle. Thus, the MVRPCS model can be established, as discussed in detail in the following paragraphs. Two decision variables are defined as follows:

$$x_{ijk}^m = \begin{cases} 1 & \text{vehicle } k \text{ from } i \text{ to } j \text{ of depot } m, \\ 0 & \text{other}, \end{cases}$$
$$y_k^m = \begin{cases} 1 & \text{vehicle } k \text{ dispatched from depot } m, \\ 0 & \text{other}. \end{cases} \tag{2.2}$$

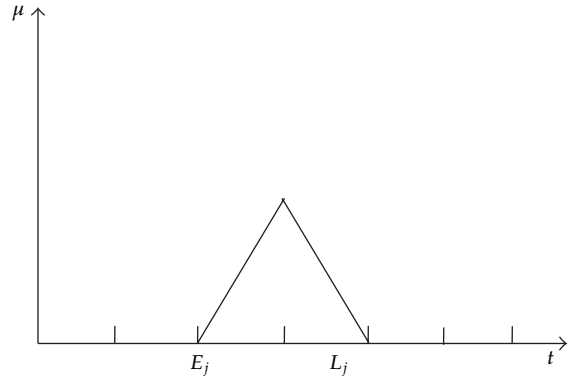


Figure 2: Triangular fuzzy number.

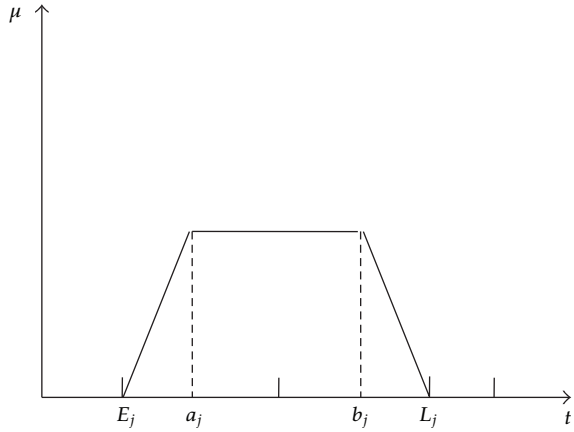


Figure 3: Trapezoidal fuzzy number.

This problem has two objectives that are determined by both the service quality and the service costs. Considering the two objectives can help the dispatcher make better decision without compromising any of the two objectives. The dispatcher can still have preferences on different objectives by selecting different parameters.

The service quality objective is to maximize average customers satisfaction:

$$\max \frac{1}{L}\sum_{i=1}^{L}\mu_i(t_i). \tag{2.3}$$

This objective is equivalent to minimizing the average customer dissatisfaction:

$$\min\left(1-\frac{1}{L}\sum_{i=1}^{L}\mu_i(t_i)\right). \tag{2.4}$$

The other objective of the service costs is to minimize travel cost, fixed cost and waiting cost. For this objective, the fixed cost of sending a vehicle is considered because

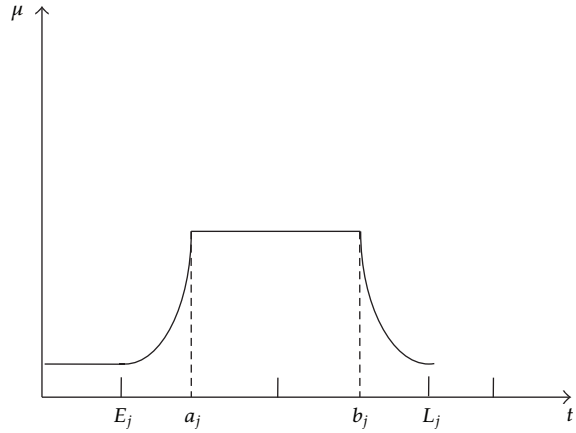


Figure 4: Improved trapezoidal fuzzy number.

vehicles in operation have depreciation and fuel consumption. Also the fixed cost is related to the number of vehicle, that is, the more vehicles, the higher fixed cost. To the best of our knowledge, no previous work has been done to take into account this fixed cost when solving multiobjective VRP with fuzzy due-time. Based on the above discussion, the mathematical model for the MVRPCS can be established as follows:

$$\min Z = \sum_{i=1}^{L+M}\sum_{j=1}^{L+M}\sum_{k=1}^{K_m}\sum_{m=1}^{M} x_{ijk}^{m} c_{ij} + \sum_{k=1}^{K_m}\sum_{m=1}^{M} y_{k}^{m} F_k + \sum_{i=1}^{L} w_i(t_i) \tag{2.5}$$

s.t.,

$$\sum_{i=1}^{L}\sum_{j=1}^{L+M} x_{ijk}^{m} d_i \le b_k, \quad m \in \{L+1, L+2, \ldots, L+M\};\ k \in \{1, 2, \ldots K_m\}, \tag{2.6}$$

$$\sum_{j=1}^{L}\sum_{k=1}^{K_m} x_{ijk}^{m} \le K_m, \quad i = m \in \{L+1, L+2, \ldots, L+M\}, \tag{2.7}$$

$$\sum_{j=1}^{L+M}\sum_{k=1}^{K_m}\sum_{m=1}^{M} x_{ijk}^{m} = 1, \quad i \in \{1, 2, \ldots, L\}, \tag{2.8}$$

$$\sum_{i=1}^{L+M}\sum_{k=1}^{K_m}\sum_{m=1}^{M} x_{ijk}^{m} = 1, \quad j \in \{1, 2, \ldots, L\}, \tag{2.9}$$

$$t_j = \begin{cases} t_i + s_i + t_{ij}, & t_i \ge E_i, \\ E_i + s_i + t_{ij}, & t_i < E_i, \end{cases} \tag{2.10}$$

$$t_j < L_j, \tag{2.11}$$

$$w_j(t_j) = \operatorname{Max}\{0, E_j - t_j\}. \tag{2.12}$$

In the model, formula (2.6) assures the carrying capacity of each vehicle; formula (2.7) assures the number of vehicles that are dispatched from each depot does not exceed the capacity of the depot; formulas (2.8) and (2.9) assure each client is served only by one vehicle; formulas (2.10), (2.11), and (2.12) assure customers can be served within time windows; and from (2.10) and (2.12), the waiting cost for the vehicle can be obtained.

3. MOQEA for the MVRPCS

In this research, the multiobjective optimization method of Pareto optimal solution [31] is used. Its main advantage is to approximate the Pareto front in order to provide a set of equivalent solutions to the decision maker [32, 33]. The algorithm to solve multiobjective optimization problem of Pareto optimality involves two main questions:

(i) how to construct a Pareto optimal solution set, namely non-dominated solutions set, and make it close to the Pareto optimal front as much as possible?

(ii) how to attain the diversity and variety of solutions?

To address these two issues, a self-adaptive grid multiobjective quantum evolutionary algorithm (MOQEA) to solve the MVRPCS problem is proposed. The method of constructing non-dominated solution set and attaining the diversity and variety of solutions is described in the following sections.

3.1. Constructing Nondominated Solution Set

In this paper, the Challenge Cup rule [34] is used to construct non-dominated solution set. Assuming P is an evolution group and Q is a constructed set, let $Q = P$ initially. Assume *Ndset* is a non-dominated set which is empty initially. The basic idea of this rule is to take any individual x from Q, followed by comparison with all other individuals y in Q : if $x \succ y$, then clear y from Q; if $y \succ x$, then use y to replace x, and then y is the new champion and continues to be compared with other individuals. After a comparison, cluster$(x) = \{y \mid x \succ y, x, y \in P\}$ is formed, where x is the smallest element. Add x to *Ndset* and continue the comparison until Q is empty.

3.2. Method of Attaining the Diversity and Variety of Set Based on Self-Adaptive Grid

To attain the variety of the set, the individual space is divided into several small areas each of which is a called a grid, as shown in Figure 5. Thus each individual is associated with a grid in the figure, and the number of individuals in each grid can be defined as extrusion coefficient.

A grid is used in many different ways to maintain the diversity and variety of solutions. Knowles and Crone [35] proposed a pareto archived evolution strategy for pareto multiobjective optimization. Corne et al. [36] proposed a pareto envelope-based selection algorithm for multiobjective optimization.

When a grid contains more than one individual, these individuals are treated as the same solution. As such, the size of grid is very important. When the grid is too large, multiple individuals will exist in the same grid, and the resultant solution distribution is not accurate.

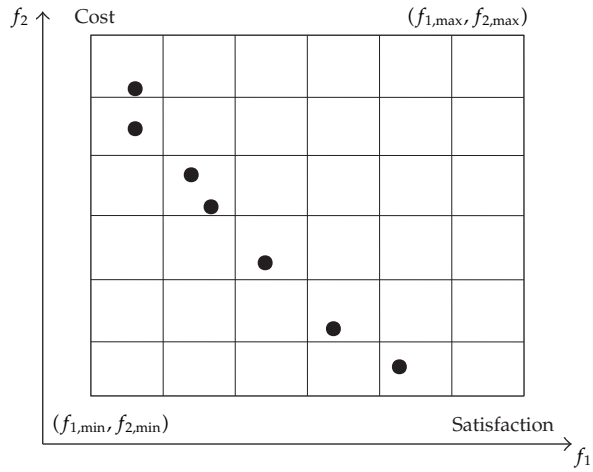


Figure 5: Individual space divided by grid.

When the grid is too small, it is likely that there are no individuals in some grids, and so it takes longer computation time though the resultant solution distribution is accurate. Therefore, computation time and accuracy must be traded off when determining the grid size.

There are two objective functions in this optimization problem. The range of the customers' satisfaction is $[0,1]$, and the range of travel costs and waiting costs is changed with the experimental data. The boundaries of the grid in the target space can be determined by the range of the above two objective functions.

In this paper, the number of grids is not fixed in each generation but automatically adjusted based on the distribution of the current generation of non-dominated solution set. The grid boundary is a fixed value. In the process of each evolutionary generation, the number of grid is adjusted by the D-value between the maximum and minimum values of each dimension in the non-dominated solution set. The method of self-adaptive grid is designed as follows.

The two objections can be described by $k = \{1,2\}$. Mark the number of each dimension grid in generation 1 as $N_{1,k}$, $Ndset_1$ as the non-dominated solutions set in generation 1, and $Ndset_t$ as the non-dominated solutions set in generation t. If $f_{t,k}$ is the objection of each dimension, then the D-value of each dimension can be described as follows:

generation 1: $\forall k = \{1,2\}$

$$\text{Diffence}_{1,k} = \max\{f_{1,k} \mid f_1 \in Ndset_1\} - \min\{f_{1,k} \mid f_1 \in Ndset_1\}, \tag{3.1}$$

generation t : $\forall k = \{1,2\}$

$$\text{Diffence}_{t,k} = \max\{f_{t,k} \mid f_t \in Ndset_t\} - \min\{f_{t,k} \mid f_t \in Ndset_t\}. \tag{3.2}$$

The number of each dimension grid in generation t is

$$N_{t,k} = \left[N_{1,k} * \frac{\text{Diffence}_{1,k}}{\text{Diffence}_{t,k}} \right], \quad \text{where } [\ ] \text{ means being rounded.} \tag{3.3}$$

To keep the diversity and variety of the non-dominated solution set, choose the individual with the maximum extrusion coefficient and delete it form the non-dominated solution set.

3.3. Representation

Quantum evolutionary algorithm (QEA) [37] is based on the concept and principles of quantum computing such as a quantum bit and superposition of states. Instead of binary and numeric representation, QEA uses a Q-bit chromosome as a probabilistic representation and a Q-bits individual is modeled by a string of Q-bits.

The smallest unit of information stored in two-state quantum computer is called a Q-bit, which may be in the "1" state, or in the "0" state, or in any superposition of the two. The state of a Q-bit can be represented as follows

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{3.4}$$

where α and β are complex numbers. $|\alpha|^2$ and $|\beta|^2$ donate the probabilities that the Q-bit will be found in the "0" state and in the "1" state, respectively. Normalization of the state to unity is used to meet $|\alpha|^2 + |\beta|^2 = 1$.

So a Q-bit individual with a string of m Q-bits can be expressed as follows

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m \\ \beta_1 & \beta_2 & & \beta_m \end{bmatrix}, \tag{3.5}$$

where $|\alpha_i|^2 + |\beta_i|^2 = 1,\ i = 1, 2, \ldots, m$.

The main advantage of the representation is that any linear superposition of solutions can be represented. For example, a three-Q-bit system can contain the information of eight states. QEA with Q-bit representation has a better characteristic of population diversity than other representations, as it can represent linear superposition of state's probabilities.

In this paper, a method of converting integer representation to Q-bit representation is designed. For the MVRPCS with L customers, the representation of customers route is described as the permutation of $1 \sim L$. Note the permutation of $1 \sim L$ is $(g_1, g_2, \ldots, g_L)$, and represent each gene $g_j (j = 1, 2, \ldots, L)$ as a string of r-Q-bit, then L groups n-Q-bit are obtained. So, a quantum individual is described as a $rL \times 2$ Q-bit matrix. Here $r = \{m\}$, where $\{m\}$ is the smallest integer not less than m and $2^m \geq L$, thus we can get $m \geq \log_2 L$.

3.4. Decoding Method

The "Customers permutation Route First, Vehicles distribution Cluster Second" rule is adopted for decoding.

Table 1: Lookup table of rotation angle.

r_i	y_i	$f(y) \succ f(r)$	$\Delta\theta_i$	$s(\alpha_i, \beta_i)$			
				$\alpha_i\beta_i > 0$	$\alpha_i\beta_i < 0$	$\alpha_i = 0$	$\beta_i = 0$
0	0	False	0	0	0	0	0
0	0	True	θ	+1	−1	0	±1
0	1	False	0	0	0	0	0
0	1	True	θ	−1	+1	±1	0
1	0	False	θ	−1	+1	±1	0
1	0	True	θ	−1	+1	0	±1
1	1	False	0	0	0	0	0
1	1	True	θ	−1	+1	0	±1

(1) Firstly, get the customers permutation route. The solution of MVRPCS is a permutation of all customers and Q-bit representation cannot be evaluated directly. So it should be converted to permutation for evaluation.

The Q-bit string is firstly converted to binary string γ. Specifically, a random number η between $[0, 1]$ is generated, if the ith bit α_i of Q-bit string satisfied $|\alpha_i|^2 > \eta$, then let the corresponding bit γ_i of the binary string γ be 1, otherwise let it be 0. Then the binary representation is converted to integer representation, which is viewed as random key representation [38], and customer permutation is constructed based on the generated random key. If two random key values are different, the smaller random key denotes the customer with smaller number; otherwise, let the one that first appears denote the customer with smaller number.

(2) Secondly, distribute the vehicles and get the subroute. A vehicle is dispatched to service customers according to the customers' permutation route, if the vehicle cannot serve the next customer when it cannot meet the time window or loading capacity constraints, a new vehicle will be dispatched. For example, the customers' permutation route is [8 5 9 3 4 1 2 6 7], the 3 depots notes [10 11 12], use this decoding method, the subroute is: Route 1: 11-8-5-9; Route 2: 10-3-4-1; Route 3: 12-2-6-7.

3.5. Strategy of Updating by Q-Gate

In the MOQEA, a Q-gate is an evolution operator which is the same as the QEA in [39]. A rotation gate is often used to update a Q-bit individual, as shown in (3.6)

$$\begin{bmatrix} \alpha_i' \\ \beta_i' \end{bmatrix} = U(\theta_i) \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}, \tag{3.6}$$

where $[\alpha_i, \beta_i]^T$ is the ith Q-bit and θ_i is the rotation angle of each Q-bit. $\theta_i = s(\alpha_i, \beta_i)\Delta\theta_i$.

The lookup table of θ_i is shown in Table 1. In this paper, a non-dominated solution is randomly selected as the current objective solution from the non-dominated solution set. In the multiobjective optimization, it is unable to find the optimal value with all objectives met. So we need to choose an objective solution y for each individual r only to find the non-dominated solution set.

Table 2: Optimization results of coefficient k ($s = 0.2$).

Problem		pr02 (96, 4)			pr07 (72, 6)	
k	VN	CS	Cost	VN	CS	Cost
0.08	16	0.402	4539	15	0.492	4116
0.06	15	0.378	4045	14	0.424	3904
0.05	14	0.355	3906	13	0.399	3331
0.02	15	0.364	3998	14	0.431	3969

Table 3: Optimization results of constant s ($k = 0.06$).

Problem		pr02 (96, 4)			pr07 (72, 6)	
s	VN	CS	Cost	VN	CS	Cost
0.1	15	0.381	4013	14	0.435	3391
0.2	14	0.357	3902	13	0.389	3367
0.3	15	0.361	3926	15	0.486	4124
0.4	16	0.411	4625	15	0.498	4102

In the above table, $\Delta\theta_i$ is the magnitude of rotation angle. $s(\alpha_i, \beta_i)$ is the sign of θ_i that determines the direction. r_i and y_i are the ith Q-bit of the binary solution in individual r and the objective solution y respectively. $f(\bullet)$ is the objective. θ is the rotation angle of size, which affects the convergence speed and search capability. In this paper, a method of dynamically adjusting the rotation angle is proposed, that is; θ will be changed with the extrusion coefficient as discussed in Section 3:

$$\theta = \frac{0.5s}{n}\pi, \tag{3.7}$$

where n is the extrusion coefficient. s is a constant in the range $[0, 1]$. From (3.7), we can see when the extrusion coefficient n is small, the rotation angle θ is big to accelerate the convergence speed; when the Extrusion Coefficient n is big, and the search step size will be reduced to enhance the solution diversity.

3.6. Procedure of MOQEA

The flow chart of the MOQEA for this problem is illustrated in Figure 6.

The detailed procedure of the MOQEA is as follows.

Step 1. Let $t = 0$ and randomly generate an initial population $Q(t) = \{q_1^t, q_2^t \cdots q_n^t\} = \left\{ \begin{matrix} \alpha_1^t \\ \beta_1^t \end{matrix} \middle| \begin{matrix} \alpha_2^t \\ \beta_2^t \end{matrix} \middle| \cdots \begin{matrix} \alpha_n^t \\ \beta_n^t \end{matrix} \right\}$, that is, randomly generate any value in $[0, 1]$ for α_i and β_i, where q_j^t denotes the jth individual in the tth generation. At the same time, construct initially empty external set $O(t)$ with size of m.

Step 2. Convert $Q(t)$ to binary population $R(t)$, then convert it to integer population $P(t)$.

Step 3. According to the decoding method to get the subroute, evaluate the objectives to get the MVRPCS solution set $M(t) = \{(f_{11}^t, f_{21}^t) \cdots (f_{1n}^t, f_{2n}^t)\}$, where (f_{1j}^t, f_{2j}^t) is the jth value of the two objectives in the tth generation. And let $M(t)$ be the construction set.

Table 4: The distance and demand of each client.

CN	Coordinate/km		ST	De/t	TW
1	−29.730	64.136	2	12	[399 525]
2	−30.664	5.463	7	8	[121 299]
3	51.642	5.469	21	16	[389 483]
4	−13.171	69.336	24	5	[204 304]
5	−67.413	68.323	1	12	[317 458]
6	48.907	6.274	17	5	[160 257]
7	5.243	22.260	6	13	[170 287]
8	−65.002	77.234	5	20	[215 321]
9	−4.175	−1.569	7	13	[80 233]
10	23.029	11.639	1	18	[90 206]
11	25.482	6.287	4	7	[397 525]
12	−42.615	−26.392	10	6	[271 420]
13	−76.672	99.341	2	9	[108 266]
14	−20.673	57.892	16	9	[340 462]
15	−52.039	6.567	23	4	[226 377]
16	−41.376	50.824	18	25	[446 604]
17	−91.943	27.588	3	5	[444 566]
18	−65.118	30.212	15	17	[434 557]
19	18.597	96.716	13	3	[319 460]
20	−40.942	83.209	10	16	[192 312]
21	−37.756	−33.325	4	25	[414 572]
22	23.767	29.083	23	21	[371 462]
23	−43.030	20.453	20	14	[378 472]
24	−35.297	−24.896	10	19	[308 477]
25	−54.755	14.368	4	14	[329 444]
26	−49.329	33.374	2	6	[269 377]
27	57.404	23.822	23	16	[398 494]
28	−22.754	55.408	6	9	[257 416]
29	−56.622	73.340	8	20	[198 294]
30	−38.562	−3.705	10	13	[375 467]
31	−16.779	19.537	7	10	[200 338]
32	−11.560	11.615	1	16	[456 632]
33	−46.545	97.974	21	19	[72 179]
34	16.229	9.320	6	22	[182 282]

Table 4: Continued.

CN	Coordinate/km		ST	De/t	TW
35	1.294	7.349	4	14	[159 306]
36	−26.404	29.529	13	10	[321 500]
37	4.352	14.685	9	11	[322 430]
38	−50.665	−23.126	22	15	[443 564]
39	−22.833	−9.814	22	13	[207 348]
40	−71.100	−18.616	18	15	[457 588]
41	−7.849	32.074	10	8	[203 382]
42	11.877	−24.933	25	22	[75 167]
43	−18.927	−23.730	23	24	[459 598]
44	−11.920	11.755	4	3	[174 332]
45	29.840	11.633	9	25	[130 225]
46	12.268	−55.811	17	19	[169 283]
47	−37.933	−21.613	10	21	[115 232]
48	42.883	−2.966	17	10	[414 531]

Step 4. Use the formulas (2.4) and (2.5) to evaluate the domination of each (f_{1j}^t, f_{2j}^t) in construction set $M(t)$ and use the method of Challenge Cup rules to construct non-dominated solution set $NDset(t)$.

Step 5. When $t = 0$, reproduce $NDset(0)$ to the external set $O(t)$. When $t > 0$, if a certain individual in $NDset(t)$ dominates one solution in $O(t)$ delete the solution and join the individual into the $O(t)$, and if a solution in $O(t)$ dominates a certain individual in $NDset(t)$, the solutions in $O(t)$ does not change; otherwise, join the individual of $NDset$(t) into the $O(t)$.

Step 6. Adjust the size of $O(t)$ to the number of m and satisfy the distribution of the non-dominated solution set. The method is discussed in detail as follows. If the size of $O(t)$ is less than m, randomly select individual of $NDset$(t) into the $O(t)$ until the size of $O(t)$ is m; otherwise, use the self-adaptive grid method to keep the diversity and variety of $O(t)$, divide the individual space of $O(t)$ into several small grids, choose the grid with the maximum extrusion coefficient, and randomly delete a solution set from it.

Step 7. If the stopping condition is satisfied, then output the Pareto set; otherwise, go on to the following steps.

Step 8. Randomly select some individuals from the Q-bit $Q(t)$, which is instated of the individuals from $Q(t)$ corresponding to $O(t)$.

Step 9. Use (3.6) to perform rotation operation for $Q(t)$ to generate $Q(t + 1)$.

Step 10. Let $t = t + 1$ and go back to Step 2.

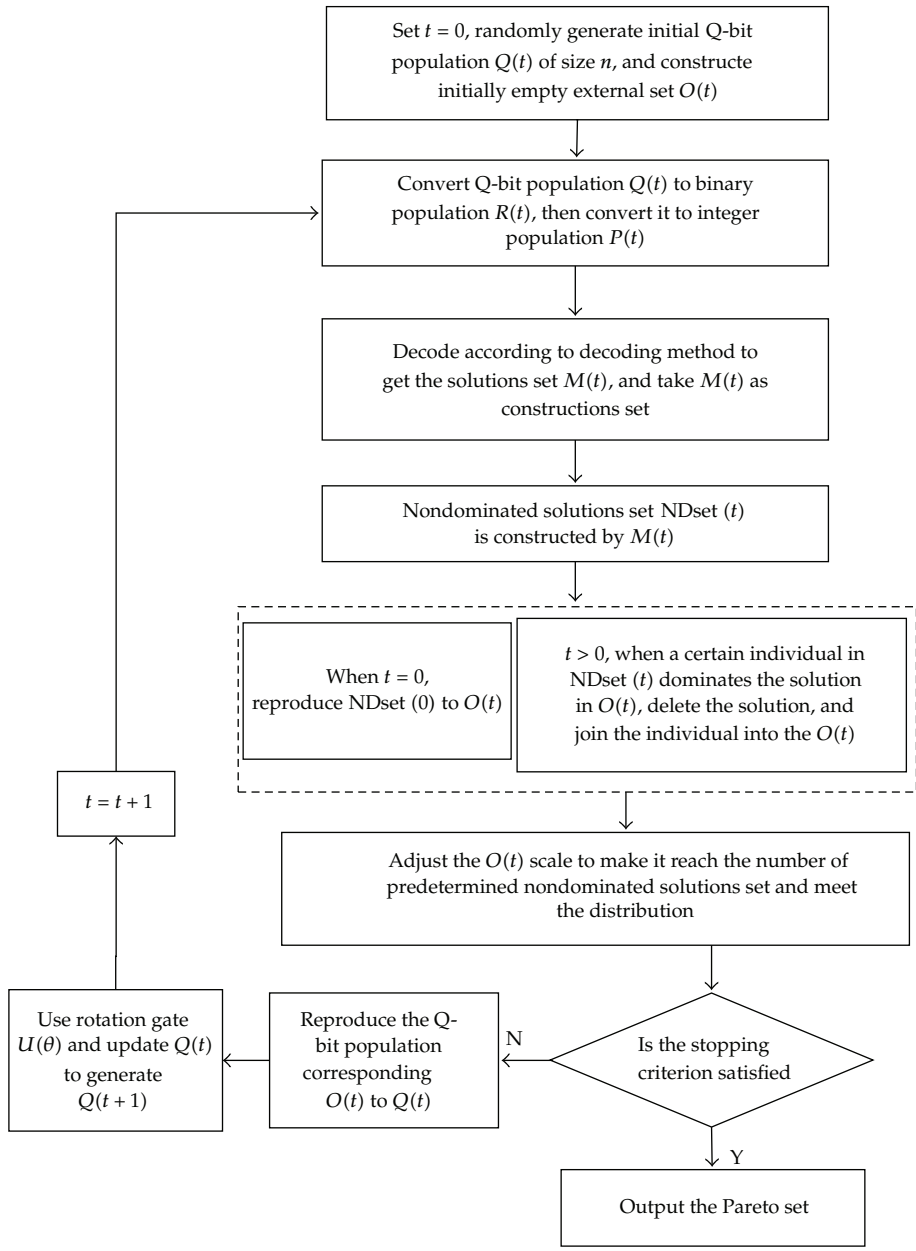


Figure 6: Flow chart of MOQEA.

4. Experiment Results and Comparisons

4.1. Experimental Data

There are few studies on multi-VRP taking into account customers' satisfaction. Among the studies that have taken into account customer satisfaction, most of them are evaluated using randomly generated test cases. Therefore, there is no standard test cases library. The

Table 5: The distance of each depot.

Depot no.	Coordinate/km		Service time	Demand/t	Time windows
49	4.163	13.559	0	0	[0 1000]
50	21.387	17.105	0	0	[0 1000]
51	−36.118	49.097	0	0	[0 1000]
52	−31.201	0.235	0	0	[0 1000]

Table 6: Optimization results of pr01.

Pareto set	The solution of maximum satisfaction (472, 4532)	The solution of minimizing travel cost, fixed cost, and waiting cost (630, 3523)
VN	9	11
Route	52 21 36 31 27 6 24	52 43 21 14 13
	51 28 13 44 25	52 30 45 44 38
	51 5 39 17	50 6 40 48 47 17
	49 32 19 20 18 42	49 7 37 11 32 26 16 36 42
	51 4 11 41 26 1 29 22	52 12 41 15 2 4
	51 14 12 43 16 33	51 33 18 3
	50 3 45 8 40 10 37	51 1 35 20 39
	52 23 34 46 38 30 15 47	51 8 29 9 22
	49 7 9 35 48 2	51 5 34 24 23
		51 19 27 28 10 31 25
		49 46

tests data used in this research is from the benchmark problems in the standard example library of MDVRPTW (multiple depot vehicle routing problem with time windows), and all examples can be downloaded from http://neo.lcc.uma.es/radi-aeb/WebVRP/. $[E_i, L_i]$ is the time windows in initial data, as the tolerable time in this paper. $[a_i, b_i]$ is the desirable service time, which can be computed using the following formula [27]:

$$\begin{aligned} a_i &= E_i + \text{rand}() * 0.5 * (L_i - E_i), \\ b_i &= L_i - \text{rand}() * 0.5 * (L_i - E_i). \end{aligned} \tag{4.1}$$

4.2. *Parameters Discussion of MOQEA*

The parameters involved in the MOQEA include coefficient k in formula (2.1) and constant s in formula (3.7). The proposed MOQEA for different parameters was discussed and analyzed results are shown in Tables 2 and 3. VN is the vehicle numbers. CS is the customer nonsatisfaction. These tables only list the solution that the total cost is the least. From the tables we can see when $s = 0.2$ and $k = 0.05$, the resultant vehicle number is the smallest.

Table 7: Comparisons of the MOQEA to the HMOEA in [34].

Algorithm	Project 1 MOQEA			Project 2 HMOEA in [34]		
Problems	VN	CS	Cost	VN	CS	Cost
pr02 (96, 4)	14	0.355	3906	15	0.391	4005
pr07 (72, 6)	13	0.399	3331	13	0.402	3397

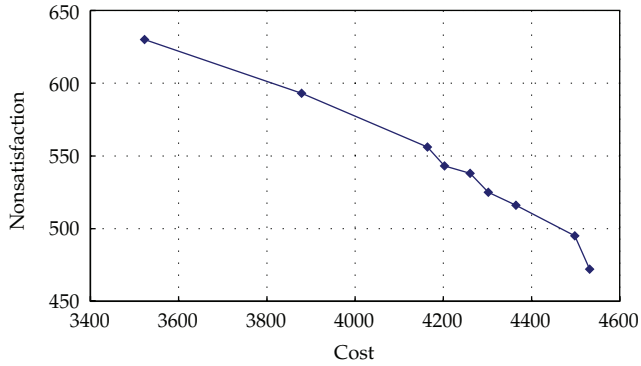


Figure 7: Pareto optimal solution set.

4.3. Simulation Experiments

All the programs in this research are developed using the JAVA language and run on a PC with Dual 2.8 GHz CPU and 1.0 GB of memory. A manufacturing company has 4 warehouses and provides goods to 48 vendors. The actual distribution process can be attributed to the open, capacity constraints, and multidepot VRP. The capacity is 20 t. The proposed MOQEA is used to solve this problem, and the distance and demand of each client and depot are shown in Tables 4 and 5. CN is the customer NO. ST is the service time. De is the demand. TW is the time windows.

The results obtained are shown in Table 6 and Figure 7. Specifically, the number of iterations is 2000, and the population size is 30. The coefficient k is set as 0.06. The constant s in formula (4.1) is 0.2. The Pareto optimal solution set is (630, 3523), (593, 3879), (556, 4164), (543, 4203), (538, 4261), (525, 4302), (516, 4365), (495, 4498), (472, 4532). The value of non-satisfaction magnified 1000 times. It can be seen from Figure 7 that the Pareto front that is close to the axis forms a more satisfactory solution set. Comparing the leftmost and the rightmost points, we can see that in this instance, it would be possible to decrease the total cost by 20% at the expense of an increase in the non-satisfaction which is about 25%.

4.4. Comparison and Discussion

In order to evaluate the performance of the algorithm, the proposed MOQEA is compared with the hybrid multiobjective evolutionary algorithm (HMOEA) developed in [34]. In the HMOEA, feasible individuals are constructed as the initial population by using the push-forward insertion heuristic (PFIH), and the GA is used to update these populations to obtain the new subpopulation and to improve the individuals of the subpopulation by the local

search method of λ-interchange with variable probability, then non-dominated solution set is constructed by using the Challenge Cup rule.

Table 7 shows the comparison of the results obtained from the two algorithms. Because the calculation result is a solution set, this table only lists the solutions in which the total cost is the least. VN is the vehicle numbers. CS is the customer non-satisfaction. From the table we can see for these two multidepot VRPs obtained from the proposed MOQEA the total cost is smaller than that from the HMOEA, and the customers' satisfaction obtained from the MOQEA is greater than that from the HMOEA.

5. Conclusions

This paper presents the modeling of vehicle scheduling problem that takes into account customer satisfaction and the development of the MVRPCS. Specifically, an improved trapezoidal fuzzy number is proposed to represent customers' satisfaction and the MOQEA for this problem is developed. The MOQEA can get multiple solutions, namely, the Pareto optimal solution set, according to his own expectations. These solutions will be used by the decision maker to choose the best one on the basis of different preferences on satisfaction maximization and travel costs minimization. In the MOQEA, the Challenge Cup rule is constructed for non-dominated solution set and a method for attaining keeping the variety of the solution set, is designed, based on self-adaptive grid. Simulation results and comparisons show that the MOQEA is an effective method. In our future work, we will focus on improving the algorithm and test it on other datasets.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (Grant no. 60970021), the Postdoctoral Science Foundation of Zhejiang Province, and the Department of Education Foundation of Zhejiang Province (No. Y201225032). The authors are also most grateful for the constructive suggestions from anonymous reviewers which led to the making of several corrections and suggestions that have greatly aided in the presentation of this paper.

References

[1] P. Toth and D. Vigo, *The Vehicle Routing Problem*, vol. 9 of *SIAM Monographs on Discrete Mathematics and Applications*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA, 2002.

[2] G. B. Dantzig and J. H. Ramser, "The truck dispatching problem," *Management Science*, vol. 6, pp. 80–91, 1959.

[3] W. Huang and S. Chen, "Epidemic metapopulation model with traffic routing in scale-free networks," *Journal of Statistical Mechanics*, vol. 2011, no. 12, Article ID P12004, 19 pages, 2011.

[4] M. Li, W. Zhao, and S. Chen, "MBm-based scalings of traffic propagated in internet," *Mathematical Problems in Engineering*, vol. 2011, Article ID 389803, 21 pages, 2011.

[5] S. Y. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.

[6] C. H. Jiang, S. G. Dai, and Y. H. Hu, "Hybrid genetic algorithm for capacitated vehicle routing problem," *Computer Integrated Manufacturing Systems*, vol. 13, no. 10, pp. 2047–2052, 2007 (Chinese).

[7] C. Prins, "A simple and effective evolutionary algorithm for the vehicle routing problem," *Computers & Operations Research*, vol. 31, no. 12, pp. 1985–2002, 2004.

[8] P. Reiter and W. J. Gutjahr, "Exact hybrid algorithms for solving a bi-objective vehicle routing problem," *Central European Journal of Operations Research*, vol. 20, no. 1, pp. 19–43, 2012.
[9] S. P. Anbuudayasankar, K. Ganesh, S. C. Lenny Koh, and Y. Ducq, "Modified savings heuristics and genetic algorithm for bi-objective vehicle routing problem with forced backhauls," *Expert Systems with Applications*, vol. 39, pp. 2296–2305, 2012.
[10] S. C. Hong and Y. B. Park, "Heuristic for bi-objective vehicle routing with time window constraints," *International Journal of Production Economics*, vol. 62, no. 3, pp. 249–258, 1999.
[11] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.
[12] A. Lim and F. Wang, "A smoothed dynamic tabu search embedded GRASP for m-VRPTW," in *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '04)*, pp. 704–708, November 2004.
[13] K. C. Tan, Y. H. Chew, and L. H. Lee, "A hybrid multiobjective evolutionary algorithm for solving vehicle routing problem with time windows," *Computational Optimization and Applications*, vol. 34, no. 1, pp. 115–151, 2006.
[14] A. Garcia-Najera and J. A. Bullinaria, "An improved multi-objective evolutionary algorithm for the vehicle routing problem with time windows," *Computers & Operations Research*, vol. 38, no. 1, pp. 287–300, 2011.
[15] P. Badeau, F. Guertin, M. Gendreau, J. Y. Potvin, and E. Taillard, "A parallel tabu search heuristic for the vehicle routing problem with time windows," *Transportation Research C*, vol. 5, no. 2, pp. 109–122, 1997.
[16] K. C. Tan, L. H. Lee, Q. L. Zhu, and K. Ou, "Heuristic methods for vehicle routing problem with time windows," *Artificial Intelligence in Engineering*, vol. 15, no. 3, pp. 281–295, 2001.
[17] J. Berger and M. Barkaoui, "A parallel hybrid genetic algorithm for the vehicle routing problem with time windows," *Computers & Operations Research*, vol. 31, no. 12, pp. 2037–2053, 2004.
[18] A. Le Bouthillier and T. G. Crainic, "A cooperative parallel meta-heuristic for the vehicle routing problem with time windows," *Computers and Operations Research*, vol. 32, no. 7, pp. 1685–1708, 2005.
[19] G. B. Alvarenga, G. R. Mateus, and G. de Tomi, "A genetic and set partitioning two-phase approach for the vehicle routing problem with time windows," *Computers and Operations Research*, vol. 34, no. 6, pp. 1561–1584, 2007.
[20] R. Cheng and M. Gen, "Vehicle routing problem with fuzzy due-time using genetic algorithms," *Japanese Journal of Fuzzy Theory and Systems*, vol. 7, no. 5, pp. 1050–1061, 1995.
[21] R. Cheng and M. Gen, "Fuzzy vehicle routing and scheduling problem using genetic algorithm," in *Genetic Algorithms and Soft Computing*, F. Herrera and J. Verdegay, Eds., pp. 683–709, Springer, 1996.
[22] S. Wen, W. Zheng, J. Zhu, X. Li, and S. Chen, "Elman fuzzy adaptive control for obstacle avoidance of mobile robots using hybrid force/position incorporation," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 42, no. 4, pp. 603–608, 2012.
[23] S. Y. Chen, J. Zhang, H. Zhang, N. M. Kwok, and Y. F. Li, "Intelligent lighting control for vision-based robotic manipulation," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 8, pp. 3254–33263, 2012.
[24] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, vol. 2010, Article ID 408418, 22 pages, 2010.
[25] J. Y. Zhang, Y. H. Guo, and J. Li, "Research of multi-objective fuzzy vehicle scheduling problem based on satisfaction of customers," *Journal of the China Railway Society*, vol. 25, no. 2, pp. 15–17, 2003 (Chinese).
[26] Y. J. Jia, *Optimal algorithm research of vehicle scheduling problem [Ph.D. thesis]*, Shanghai Jiaotong University, 2004.
[27] B. Wu, *Particle swarm optimization for velaicle routing problem and its application [Ph.D. thesis]*, Zhejiang University of Technology, 2008.
[28] J.-J. Lin, "A GA-based multi-objective decision making for optimal vehicle transportation," *Journal of Information Science and Engineering*, vol. 24, no. 1, pp. 237–260, 2008.
[29] C. H. Wang and C. H. Li, "Optimization of an established multi-objective delivering problem by an improved hybrid algorithm," *Expert Systems with Applications*, vol. 38, no. 4, pp. 4361–4367, 2011.
[30] K.-H. Han and J.-H. Kim, "Quantum-inspired evolutionary algorithm for a class of combinatorial optimization," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 6, pp. 580–593, 2002.
[31] M. Fonseca Carlos and J. Peter Fleming, "Genetic algorithm for multiobjecetive optimization: formulation, dicussion and generalization," in *Proceeding of the 5th International Conference on Genetic Algorithm*, pp. 416–423, 1993.

[32] S. Chen, Y. Zheng, C. Cattani, and W. Wang, "Modeling of biological intelligence for SCM system optimization," *Computational and Mathematical Methods in Medicine*, Article ID 769702, 10 pages, 2012.
[33] S. Y. Chen and Y. F. Li, "Automatic Sensor Placement for Model-Based Robot Vision," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 34, no. 1, pp. 393–408, 2004.
[34] Z. Jinhua, *Multi-Objective Evolutionary Algorithm and Its Application*, Science Press, Beijing, China, 2007.
[35] J. Knowles and D. W. Corne, "The pareto archived evolution strategy: a new baseline algorithm foe pareto multiobjective optimisation," in *Proceeding of the Congress of Evolutionary Computation*, pp. 98–105, 1999.
[36] D. W. Corne, D. K. Joshua, and J. O. Martin, "The pareto envelope-based selection algorithm for multiobjective optimization," in *Proceeding of the Parallel Problem Solving from Nature VI Conference*, pp. 839–848, Springer, 2000.
[37] A. Narayanan and M. Moore, "Quantum-inspired genetic algorithms," in *Proceedings of IEEE International Conference on Evolutionary Computation (ICEC '96)*, pp. 61–66, May 1996.
[38] J. C. Bean, "Genetic alogrithms and random keys for sequedcing and optimization," *ORSA Journal on Computing*, vol. 6, pp. 154–160, 1994.
[39] Z. Jingling, Z. Yanwei et al., "A hybrid quantum-inspired evolutionary algorithm for capacitated vehicle routing problem," in *Proceedings Advanced Intelligent Computing Theories and Applications*, vol. 5226, pp. 31–38, Springer, Berlin, Germany, 2008.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 183906, 13 pages
doi:10.1155/2012/183906

Research Article

The Impact of the HMCFRP Ratio on the Strengthening of Steel Composite I-Beams

E. Agcakoca and M. Aktas

Faculty of Engineering, Department of Civil Engineering, Sakarya University, Sakarya, Turkey

Correspondence should be addressed to M. Aktas, muharrema@sakarya.edu.tr

Received 5 October 2012; Accepted 23 October 2012

Academic Editor: Carlo Cattani

Carbon fiber-reinforced polymer materials have become popular in the construction industry during the last decade for their ability to strengthen and retrofit concrete structures. The recent availability of high-modulus carbon fiber-reinforced polymer strips (HMCFRP) has opened up the possibility of using this material in strengthening steel structures as well. The strips can be used in steel bridge girders and structures that are at risk of corrosion-induced cross-sectional losses, structural deterioration from aging, or changes in function. In this study, a set of bending experiments was performed on three types of steel beams reinforced with HMCFRP. The results were used to enhance a nonlinear finite element model built with ABAQUS software. The accuracy of the mathematical models for HMCFRP, epoxy, and steel profiles was compared with the experimental results, and the ability of HMCFRP to continue carrying load from the steel beams during rupture and postrupture scenarios was observed using numerical analysis. Using these verified finite element models, a parametric analysis was performed on the HMCFRP failure modes and the quantity to be used with IPE profile steel beams. The maximum amount of HMCFRP needed for strengthening was determined, and an upper limit for its use was calculated to avoid any debonding failure of the fiber material.

1. Introduction

Steel structural elements may require strengthening due to changes in functionality, increases in load-bearing capacity due to heavier traffic loads and corrosion occurring over time. The replacement of such elements and the installation of additional layer sections are typically proposed and applied in practice to solve these problems. However, these proposed solutions are problematic due to the increase in traffic delays they incur for bridge repairs and their cost inefficiencies [1]. Carbon fiber-reinforced polymer (CFRP) materials are heavily used in civil engineering, especially for repairing and strengthening concrete structural elements [2–5]. Such material was not usually preferred for steel structural repairs and strengthening due

to its low elasticity modulus, but it is now being used in steel structural elements with the emergence of high-modulus carbon fiber-reinforced polymer (HMCFRP). The properties of CFRP, including corrosion resistivity, low weight, and prevention of postfatigue cracks, are the reason for this implementation [6–8]. A number of research studies exist on strengthening steel structures with HMCFRP, including studies on the CFRP end dimensions that will bond to steel most effectively [9, 10], extension lengths [11, 12], the impact of its thickness on its load-carrying capacity [8, 13], and its use in welded joints [14, 15]. It was concluded that the most important issues remaining in the application of fiber materials to steel and concrete structures are bonding and development length. Aktas et al. [16] studied the impact of development length on load capacity and failure modes and determined the minimum development lengths to avoid debonding of HMCFRP. Other research also exists on the fatigue behavior of strengthened steel beams [7, 10]. Addressing the mechanical properties of fiber composites is essential in numerical modeling. The amount of volume fraction of fibers in composite matrix plays an important role for ignoring inhomogeneity [17].

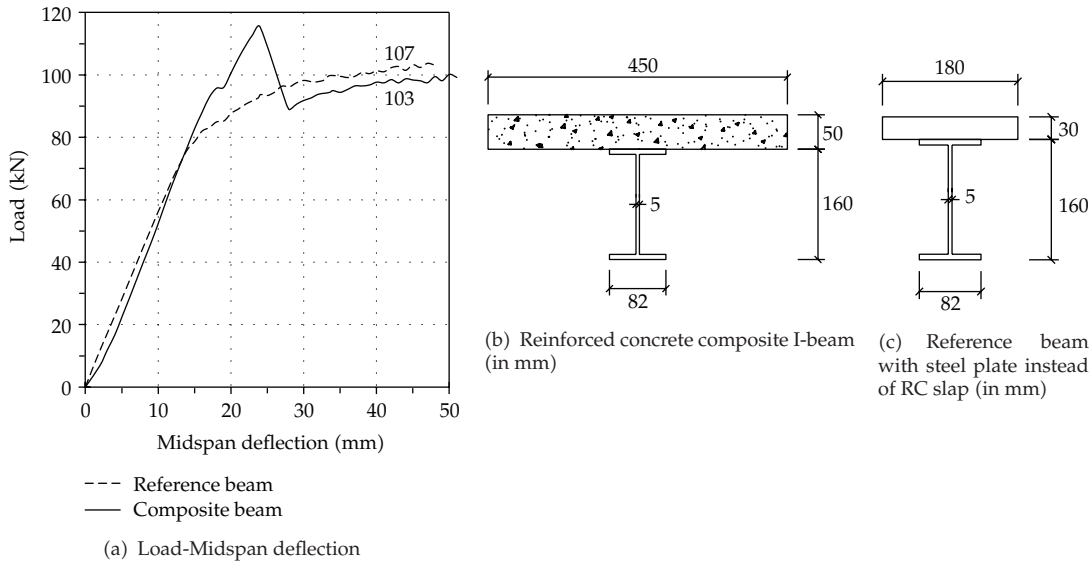
The experiments found in the literature survey are limited and do not include a broad parametric study. Questions remain about whether there is an upper limit to the quantity of CFRP material that can be used in these scenarios, whether there is a relationship between an increase in load-carrying capacity and the use of HMCFRP and whether CFRP debonding occurs at epoxy interfaces such as in concrete structures. The goal of this study was to answer these questions as they apply to BA-composite IPE steel profiles strengthened with HMCFRP. Parametric studies were conducted using the nonlinear finite element method, and the developed finite element model was verified with three different experiments.

2. Testing and Numerical Verification

A two-step experiment was performed in this part of the study. In the first step, a steel I-beam specimen was constructed and tested that matched the load-carrying capacity of a concrete composite steel I-beam. The concrete slab components were replaced by their steel counterparts and used for further studies, avoiding problems associated with manufacturing concrete parts and issues with numerical modeling during this parametric study on concrete composite I-beams. The load deflection behavior from the bending experiment validating this assumption is shown in Figure 1(a). Based on the similarity of the results, we continued to use the beam structure with a steel plate in Figure 1(c) in the rest of the study instead of the reinforced concrete composite I-beam shown in Figure 1(b). The second step consisted of strengthening experiments.

2.1. Verification Experiments: Strengthened Beam

Three identical steel composite beam plates were identically prepared and strengthened with HMCFRP to verify and test their nonlinear finite element model components. HMCFRP with dimensions of $1.4 \times 50 \times 2400$ mm was bonded to steel tension flange with epoxy material. Prior to bonding, the bottom flanges were mechanically cleaned of dust and dirt. The epoxy was carefully prepared at the rate recommended by its manufacturer and applied directly to the bottom flange, creating a very thin layer. The HM-CFRP was then clamped on the epoxy with wooden clamps. After the manufacturer's recommended time had passed, the specimens were ready for bending experiments.



(a) Load-Midspan deflection

(b) Reinforced concrete composite I-beam (in mm)

(c) Reference beam with steel plate instead of RC slap (in mm)

Figure 1: Comparison of beams with a reinforced concrete slab and a steel plate slab.

The beam specimens had a length of 3,000 mm, and the distance between the supports was 2,900 mm. The distance between the loading point and the end of the HM-CFRP material was 1,050 mm, the distance between loading points was 800 mm, and the load was applied using a 400 kN piston. A 200 × 200 × 1000 mm spreader beam was used to apply equal loads to all of the beam's loading points. Two 20 × 30 × 900 mm steel bars were utilized to ensure load transfer from the spreader to the beam. Rubber pads were used to prevent the bars from causing damage to the specimen, and three displacement meters were employed at midspan and on the bottom flange where loads were applied to measure displacements. The test setup is presented graphically in Figure 2(a). A four-point load test layout that provided a fixed moment was applied to observe the strengthened beam's bending behavior.

The load deflection figures for the three strengthened beams and the unstrengthened reference beam are shown in Figure 2(b). To clearly illustrate the behavior of the structural elements, we define four regions (a, b, c, and d) in the figures: (a) corresponds to the steel yield, (b) corresponds to the start of HMCFRP rupture, (c) shows loading of the steel without HMCFRP, and (d) shows the conclusion of the experiment. The added load-carrying capacity due to HMCFRP installation is an average of 39%.

2.2. Numerical Verification

This section discusses the steps involved in nonlinear finite element modeling. Instead of costly lab experiments, nonlinear finite element models can be used to obtain approximate results with numerical experiments. However, the accuracy of the results may depend on the type of finite elements, material mathematical models, interface models and mesh densities used. The finite element model in this study was continuously improved until its results matched the actual experimental results.

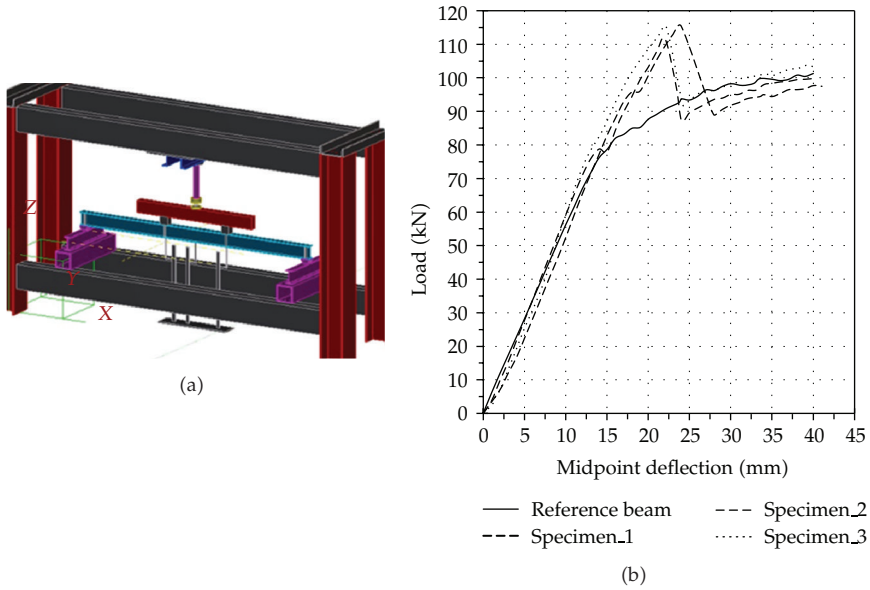


Figure 2: Test setup and results for strengthened beams.

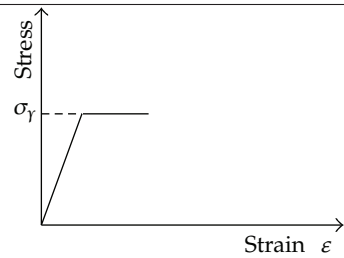
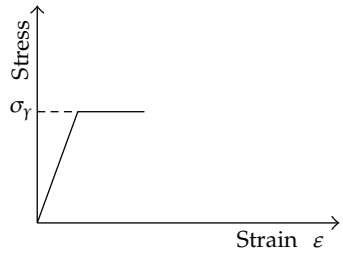
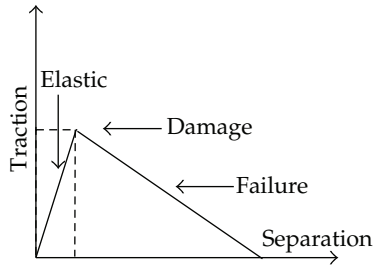
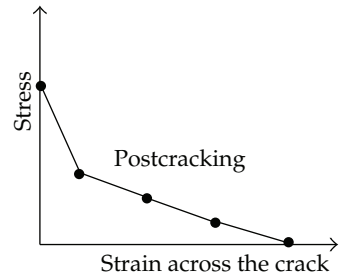
Table 1: Mechanical properties of the specimen.

HMCFRP		Steel (I-beam)			Steel plate		Epoxy	
Ult. strength (MPa)	1200	Flange	Yield strength (MPa)	275	Yield strength (MPa)	255	G, modulus (GPa)	10720
E. modulus (GPa)	440		E. modulus (GPa)	179	E. modulus (GPa)	210	E. modulus (GPa)	29770
Width (mm)	50	Web	Yield strength (MPa)	265			Nom. stress (MPa)	25
Thickness (mm)	1.4		E. modulus (GPa)	161				
Ult. long strain (mm/mm)	0.2-0.3%		Poisson's ratio	0.3				

ABAQUS-Explicit software was used for the numerical analysis. Numerical difficulties arise when solving nonlinear problems in the implicit finite element method since the iterative approach which needs to achieve convergence to enforce equilibrium may fail in highly nonlinear problems. The nonlinearities in the material and geometries were accounted for in the analysis, and the nonlinear finite element model was derived by considering sensitive mesh dependency, using reduced integration in its formulation. This reduced computing time and minimized potential inaccuracies. The mechanical properties of the steel, epoxy, and HMCFRP used in the numerical model were experimentally measured and are provided in Table 1. The finite element types used in the numerical modeling and mathematical models from the analysis are listed in Table 2.

The steel components were modeled using mathematical models of an elastoplastic material (Table 2(a)). Work hardening of the steel was not taken into account. The Von Mises yield criterion was chosen as the failure of the steel. Based on the satisfactory accuracy of the finite element model for 4-node shell element shown in a study conducted by Aktas et al. [18],

Table 2: Finite element properties of the specimen.

Member	Finite element name in ABAQUS	Description	Material model	Depiction of behavior
Steel I-beam	S4R	Shell, 4-noded, reduced integration	Elastoplastic	 (a)
Steel plate	C3D8R	Solid, 8-noded, reduced integration	Elastoplastic	 (b)
Epoxy	COH3D8	Cohesive, 8-noded	Traction-separation	 (c)
HMCFRP	C3D8R	Solid, 8-noded, reduced integration	Brittle cracking	 (d)

this paper also used a 4-node shell element in constructing I-beam. Because the steel plate being substituted for reinforced concrete had minor deformations, it was modeled as a solid 8-node element (Table 2(b)).

It is very important to have accurate and robust models for the epoxy material named as a cohesive layer in the analysis because the epoxy provides the medium for stress transfer

between the steel and HMCFRP. It was modeled as a solid 8-node cohesive layer element and defined with a traction-separation mathematical model because it creates only a very thin layer and has a separation failure mode (Table 2(c)). In other words, the behavior of the epoxy defined as cohesive material was defined with a progressive damage and failure definition. Damage initiation of the epoxy occurs when the relationship shown in (2.1) is satisfied, where the tensile stress perpendicular to surface of the element is t_n and that parallel to the surface is t_s and t_t, and their maximum values are t_n^0, t_s^0, and t_t^0, respectively. Consider

$$\left\{\frac{t_n}{t_n^0}\right\}^2 + \left\{\frac{t_s}{t_s^0}\right\}^2 + \left\{\frac{t_t}{t_t^0}\right\}^2 = 1. \tag{2.1}$$

Equations (2.2) are used in damage evolution behavior. These equations show the $\overline{t_n}$, $\overline{t_s}$, and $\overline{t_t}$ tensile strengths calculated from the elastic behavior of the material. D is a scalar between 0 and 1 and is used as the damage parameter. The behavior between the maximum value and zero-tensile stress rupture moment is modeled linearly as follows:

$$\begin{aligned} t_n &= (1-D)\overline{t_n}, \\ t_s &= (1-D)\overline{t_s}, \\ t_t &= (1-D)\overline{t_t}. \end{aligned} \tag{2.2}$$

Epoxy layer which is modeled by employing cohesive elements is tied to the neighboring surfaces both on steel bottom flange and HM-CFRP top surface. Thus, the degree of freedom at the intersecting nodes is same for epoxy, steel, and HM-CFRP. Traction separation properties are assigned to epoxy material which is modeled as a film material with a thickness of 0.1 mm.

HMCFRP is defined by the Brittle-cracking mathematical model because it is exposed to higher tensile stresses and will not yield after rupture. In this analysis, the material was modeled as linear-elastic up to its maximum tensile stress capacity, beyond which it loses all load-carrying capacity. It can be very hard to identify this postcracking phase experimentally because it occurs so quickly. This study modeled postcracking as the 1% strain level using the linear model (Table 2(d)). Shear retention was also accounted for to consider variations in post-cracked shear stiffness. The fiber material was modeled as 8-node solid elements.

The model developed for the experimental test beam using the modeling details provided above is depicted in Figure 3. The boundary conditions and the loading scenario are modeled as they were in the actual experiment. The aspect ratio of one is ensured when constructing finite elements. Also the mesh sensitivity analysis is carried out by seeding the model with 5 mm, 10 mm, 20 mm, and 40 mm sized finite elements, respectively. The model is found to be sensitive to mesh density thus 10 mm mesh size is selected for further studies reported herein.

The results obtained from the numerical model analyses were compared with the actual experimental results to validate the accuracy of the models. The comparisons of the beam's behavior under loading are shown in Figure 4, and the maximum load-carrying capacities obtained from the experiments and finite element modeling (FEM) analysis are listed in Table 3 for both strengthened and unstrengthened beams.

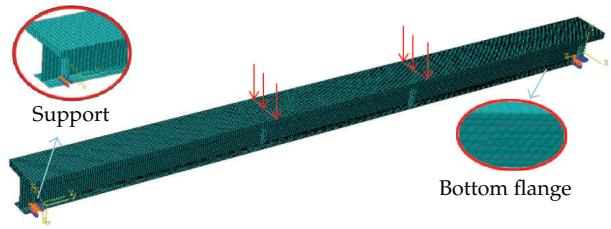
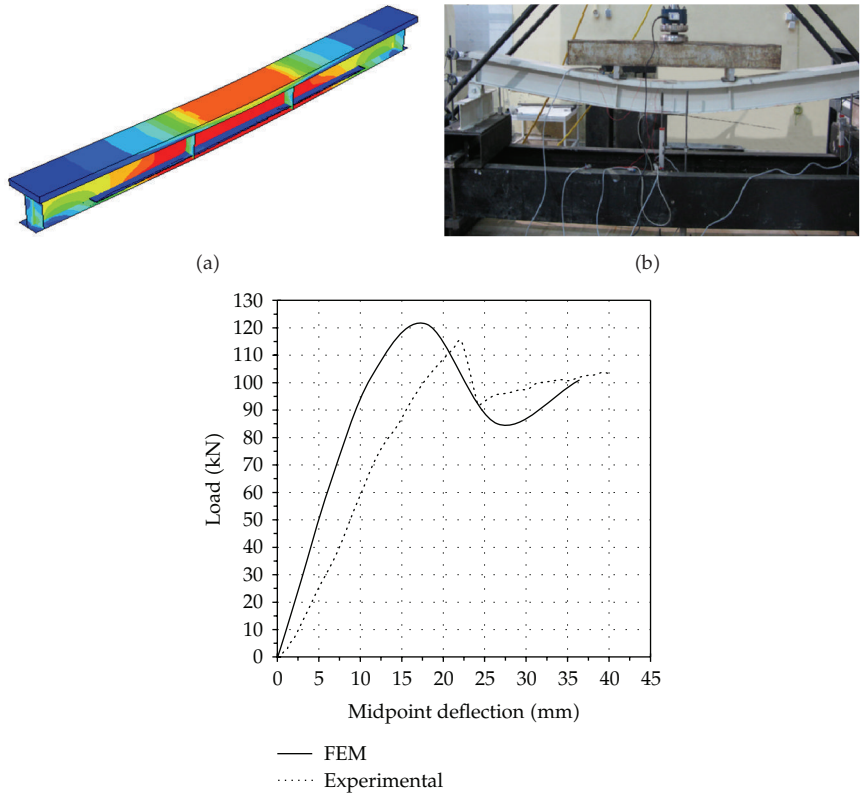


Figure 3: Finite element model of experimental specimen.



(c) Comparison of numerical and experimental test results

Figure 4: Numerical verification of experimental results.

Table 3: Comparison of strengthened and reference beams during numerical analysis and experimental testing.

		Experiment	FEM	% Difference
Reference beam	Max load (kN)	83	84	1.2
	Max moment (kN-m)	43.6	44.13	1.2
Strengthened beam	Max load (kN)	115.4	121	4.85
	Max moment (kN-m)	60.58	63.52	4.85

The experimental results and numerical models for the unstrengthened beam show a difference of 1.2% in maximum load-carrying capacity. This difference increased to 4.85% for strengthened beams. Both of these values are within acceptable ranges. The higher error of the strengthened beams can be explained by the complex behavior of the interface element. The difference between deflection points corresponding to the maximum loads is approximately 24% and could be explained as a consequence of the approximations used in the material models and ignoring geometric imperfections for the numerical model. Hence, the material is more inclined to display rigid behavior in the numerical model. Because the goal of this study was to determine the increase in load-carrying capacity during the strengthening process, the difference in the deflection estimates can be considered negligible for parametric studies.

The most important result obtained from the numerical model is the rupture behavior of the HMCFRP and the initiation of the postrupture load bearing of the steel. Analysis of the figures shows a sudden drop in the specimen's load-carrying ability right after the peak point, which corresponds to the rupture of the HMCFRP. When this rupture occurs, the steel continues to carry load because it has undergone a strain hardening process. This experimentally observed behavior was also achieved using numerical analysis under similar conditions.

3. Parametric Study

Upon further development and testing, the experimentally validated numerical models provided insight into the specific questions asked at the beginning of the study, including the maximum quantity of HMCFRP that can be used, failure modes, and the relationship between HMCFRP usage and load-carrying capacity increases. A total of 149 models for numerical specimens were developed using various quantities of HMCFRP with 7 commonly used industry-wide IPE profile sections of different lengths. Table 4 shows the properties of the numerical specimens. Based on the study reported by Aktas et al. [16], the development lengths were set at 900, 1800, and 3600 mm for beams with lengths of 3000, 6000, and 12000 mm, respectively. Variations in the specimens' load-carrying capacities were investigated by using various quantities of HMCFRP.

3.1. Parametric Study Results and Discussion

Individual load deflection figures were created for all numerical specimens after the parametric study was completed. The results are discussed for an example in more detail, while the results for all beams are tabulated.

To illustrate the results, the load deflection figures obtained for an IPE 270 profile with different HMCFRP ratios are shown in Figure 5. The percentage increases in load-carrying capacity calculated with respect to the reference beam are compared and plotted in the figure. Figure 6 summarizes the relation between HMCFRP usage and the percentage increase in capacity calculated for a single beam. As shown in Figure 6, the maximum increase in load-carrying capacity for an IPE 270 was determined for the beam strengthened with 8.9% HMCFRP. The increase in capacity began to decline when the HMCFRP usage reached 11.8%.

To explain this behavior, one must analyze the stresses of all materials at midspan where the maximum stresses have been observed. Figure 7 shows the stress contour of each material at midspan for the IPE 270 beam strengthened with 11.8% HMCFRP.

Investigation of the stress contour maps for the beam strengthened with 11.8% HMCFRP shows that the steel did not yield around the boundary of the elastic region (a),

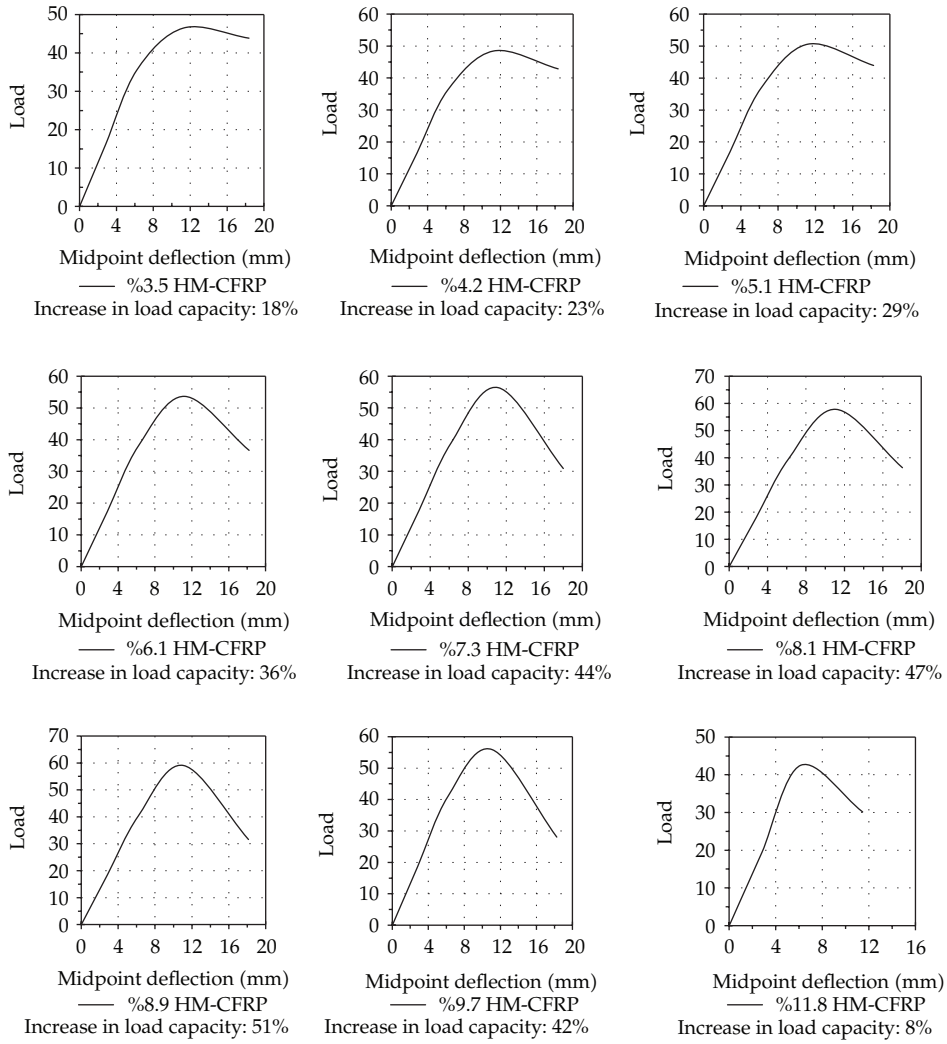


Figure 5: Load deflection plots for IPE 270, $L = 3000$ mm.

while the epoxy and HMCFRP seemed to approach their respective maximum stress levels. The region following the ultimate capacity (b) shows the point where epoxy broke along with the associated decay in the stress of the HMCFRP. (b) was also the point at which the steel stresses started to increase. This behavior can also be explained by the debonding of the HMCFRP without having reached its maximum deformation. In other words, although the HMCFRP stress level did not achieve failure, the epoxy material reached its ultimate capacity.

The maximum HMCFRP quantities are listed for each numerical specimen in Table 5 individually based on the procedures explained in Figures 5 and 6.

The values in the table above show a declining trend in the maximum HMCFRP usage ratio as the beam section area increases. This is an expected outcome because there is a higher chance of interface element failure due to extra loading on the HMCFRP installation.

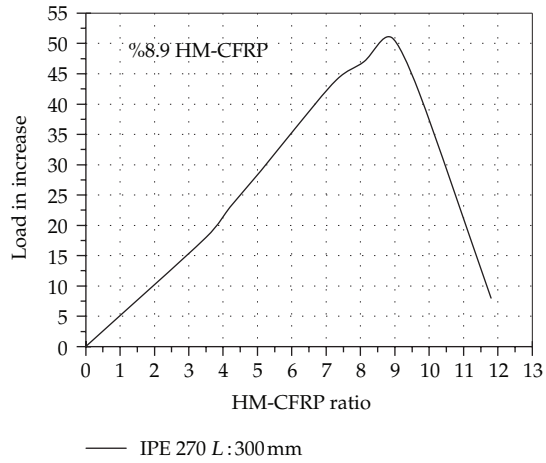


Figure 6: Relationship between HMCFRP ratio and load percentage increase.

Table 4: Numerical specimens in the parametric study.

Beam section	Beam length (mm)	Development length (mm)	HMCFRP length (mm)	$\text{Area}_{\text{HM-CFRP}}/\text{Area}_{\text{Section}}$
IPE 120	3000	900	2600	3.5–16.3
IPE 160	3000	900	2600	3.5–17.5
IPE 220	3000	900	2600	3.5–18.0
	6000	1800	5200	3.5–11.7
IPE 270	3000	900	2600	3.5–11.8
	6000	1800	5200	3.5–22.4
IPE 330	6000	900	5200	3.5–22.4
IPE 400	6000	900	5200	3.5–14.1
	12000	1800	10400	1.7–7.10
IPE 500	6000	900	5200	3.5–11.7
	12000	1800	10400	1.8–8.80

Table 5: Maximum HMCFRP ratio for each beam.

Section	Beam length (mm)	HMCFRP ratio (%)	Increase in P_u (%)	Number of specimens
IPE 120	3000	14	97	16
IPE 160	3000	11	67	17
IPE 220	3000	15	69	15
	6000	8	109	12
IPE 270	3000	8.9	51	9
	6000	7.3	54	15
IPE 330	6000	8.8	61	16
IPE 400	6000	6.7	31	13
	12000	3.5	62	9
IPE 500	6000	5.1	39	12
	12000	4.2	40	11

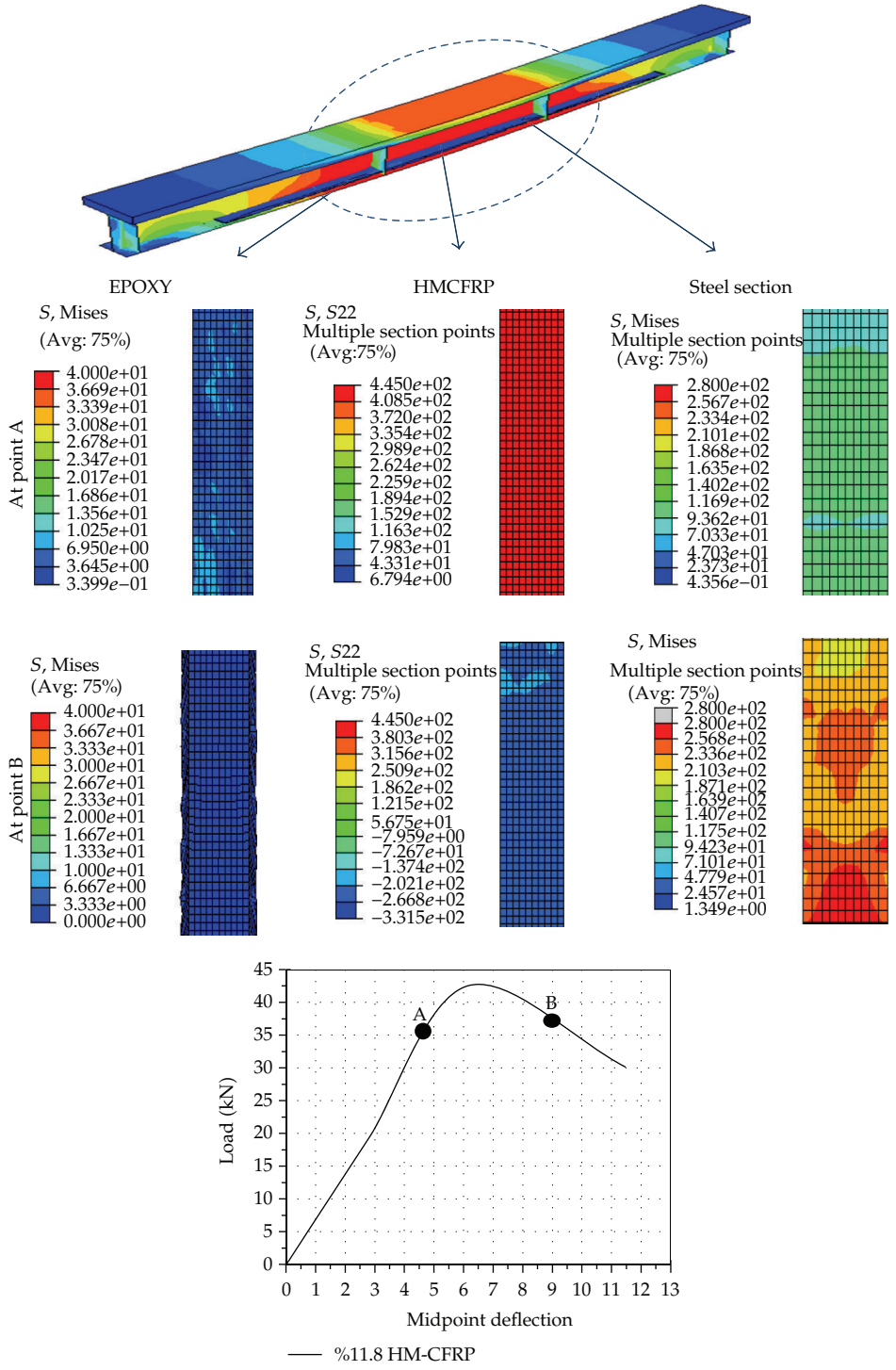


Figure 7: Stress contour of each material at beam midspan.

4. Conclusions

The load-carrying capacity of steel beams can be increased up to approximately 100% by strengthening the beams with HMCFRP. The ability to increase load-carrying capacity without increasing the structure's dead weight is a significant improvement over previous strengthening techniques. This improvement also reduces deformations due to dead weight loading. The amount of HMCFRP used in such a scenario affects the behavior of the beam significantly. To achieve effective strengthening, HMCFRP must reach high stress levels close to rupture without debonding at the epoxy interface. It is important to prevent such cases from occurring. Although greater HMCFRP usage is expected to progressively increase the load-carrying capacity, it has been shown that this is not generally the case. Instead, there is an upper limit to the quantity of HMCFRP that can be used. The most effective method of determining this upper limit is to develop a nonlinear finite element model of the beam type and analyze this model. This study discussed a methodology for developing such finite element models in detail. Using this methodology, steel beams can be repaired and strengthened when environmental effects such as corrosion and ever-increasing load requirements necessitate the addition of material to a structure.

References

[1] D. Schnerch and S. Rizkalla, "Strengthening of scaled steel-concrete composite girders and steel monopole towers with CFRP," in *Proceedings of the 2nd International Conference on FRP Composites in Civil Engineering (CICE '04)*, Adelaide, Australia, December 2004.

[2] O. Benjeddou, M. B. Ouezdou, and A. Bedday, "Damaged RC beams repaired by bonding of CFRP laminates," *Construction and Building Materials*, vol. 21, no. 6, pp. 1301–1310, 2007.

[3] A. Sharif, G. J. Al-Sulaimani, I. A. Basunbul, M. H. Baluch, and B. N. Ghaleb, "Strengthening of initially loaded reinforced concrete beams using FRP plates," *ACI Structural Journal*, vol. 91, no. 2, pp. 160–168, 1994.

[4] M. Arduini and A. Nanni, "Behavior of precracked RC beams strengthened with carbon FRP sheets," *Journal of Composites for Construction*, vol. 1, no. 2, pp. 63–70, 1997.

[5] D. S. Lunn and S. H. Rizkalla, "Strengthening of infill masonry walls with FRP materials," *Journal of Composites for Construction*, vol. 15, no. 2, pp. 206–214, 2011.

[6] A. H. Al-Saidy, F. W. Klaiber, and T. J. Wipf, "Repair of steel composite beams with carbon fiber-reinforced polymer plates," *Journal of Composites for Construction*, vol. 8, no. 2, pp. 163–172, 2004.

[7] J. A. García, A. Chiminelli, B. García, M. Lizaranzu, and M. A. Jiménez, "Characterization and material model definition of toughened adhesives for finite element analysis," *International Journal of Adhesion and Adhesives*, vol. 31, no. 4, pp. 182–192, 2011.

[8] M. Tavakkolizadeh and H. Saadatmanesh, "Strengthening of steel-concrete composite girders using carbon fiber reinforced polymers sheets," *Journal of Structural Engineering*, vol. 129, no. 1, pp. 30–40, 2003.

[9] S. Rizkalla, M. Dawood, and D. Schnerch, "Development of a carbon fiber reinforced polymer system for strengthening steel structures," *Composites A*, vol. 39, no. 2, pp. 388–397, 2008.

[10] M. Dawood and S. Rizkalla, "Bond and splice behavior of high modulus CFRP materials boned to steel structures," in *Proceedings of the 3rd International Conference on FRP Composites in Civil Engineering Miami*, 2006.

[11] D. Schnerch, K. Stanford, E. A. Sumner, and S. Rizkalla, "Bond behavior of CFRP strengthened steel bridges and structures," in *Proceeding of International Symposium on Bond Behavior of FRP in structures*, 2005.

[12] D. Schnerch, M. Dawood, S. Rizkalla, and E. Sumner, "Proposed design guidelines for strengthening of steel bridges with FRP materials," *Construction and Building Materials*, vol. 21, no. 5, pp. 1001–1010, 2007.

[13] R. Sen, L. Liby, and G. Mullins, "Strengthening steel bridge sections using CFRP laminates," *Composites B*, vol. 32, no. 4, pp. 309–322, 2001.

[14] T. Chen, X. Gu, and X. Zhao, "Fatigue behavior of CFRP repair non-load carrying cruciform welded Joint," in *Proceedings of the 5th International Conference on FRP Composites in Civil Engineering*, pp. 907–910, 2010.

[15] T. Chen, Q. Q. Yu, X. L. Gu, and X. L. Zhao, "Study on fatigue behavior of strengthened non-load-carrying cruciform welded joints using carbon fiber sheets," *International Journal of Structural Stability and Dynamics*, vol. 12, no. 1, pp. 179–194, 2012.

[16] M. Aktas and E. Agcakoca, "Defining development length Of HM-CFRP in composite I-section strengthened with HM-CFRP," *New World Sciences Academy*, vol. 7, no. 2, pp. 47–59, 2012.

[17] C. Cattani, J. J. Rushchitsky, and S. V. Sinchilo, "Physical constants for one type of nonlinearly elastic fibrous micro-and nanocomposites with hard and soft nonlinearities," *International Applied Mechanics*, vol. 41, no. 12, pp. 1368–1377, 2005.

[18] M. Aktas and C. J. Earls, "Minor axis moment-thrust response behavior in steel I-shaped members," *Journal of Structural Engineering*, vol. 132, no. 7, pp. 1079–1086, 2006.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 673049, 5 pages
doi:10.1155/2012/673049

Research Article

Representing Smoothed Spectrum Estimate with the Cauchy Integral

Ming Li[1,2]

[1] *School of Information Science & Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China*
[2] *Department of Computer and Information Science, University of Macau, Padre Tomas Pereira Avenue, Taipa, Macau, China*

Correspondence should be addressed to Ming Li, ming_lihk@yahoo.com

Received 1 October 2012; Accepted 20 November 2012

Academic Editor: Sheng-yong Chen

Estimating power spectrum density (PSD) is essential in signal processing. This short paper gives a theorem to represent a smoothed PSD estimate with the Cauchy integral. It may be used for the approximation of the smoothed PSD estimate.

1. Introduction

Estimating power spectrum density (PSD) of signals plays a role in signal processing. It has applications to many issues in engineering [1–21]. Examples include those in biomedical signal processing, see, for example, [1–3, 6, 12, 13]. Smoothing an estimate of PSD is commonly utilized for the purpose of reducing the estimate variance, see, for example, [22–29]. By smoothing a PSD estimate, one means that a smoothed estimate of PSD of a signal is the PSD estimate convoluted by a smoother function [30, 31]. This short paper aims at providing a representation of a smoothed PSD estimate based on the Cauchy's integral.

2. Cauchy Representation of Smoothed PSD Estimate

Let $x(t)$ be a signal for $-\infty < t < \infty$. Let $S_{xx}(\omega)$ be its PSD, where $\omega = 2\pi f$ is radian frequency and f is frequency. Then, by using the Fourier transform, $S_{xx}(\omega)$ is computed by

$$S_{xx}(\omega) = \left| \int_{-\infty}^{\infty} x(t) e^{-j\omega t}\, dt \right|^2, \quad j = \sqrt{-1}. \tag{2.1}$$

In practical terms, if $x(t)$ is a random signal, $S_{xx}(\omega)$ may never be achieved exactly because a PSD is digitally computed only in a finite interval, say, (T_1, T_2) for $T_1 \neq T_2$. Therefore, one can only attain an estimate of $S_{xx}(\omega)$.

Denote by $\widehat{S}_{xx}(\omega)$ an estimate of $S_{xx}(\omega)$. Then,

$$\widehat{S}_{xx}(\omega) = \left| \int_{T_1}^{T_2} x(t) e^{-j\omega t}\, dt \right|^2. \tag{2.2}$$

Without generality losing, we assume $T_1 = 0$ and $T_2 = T$. Thus, the above becomes

$$\widehat{S}_{xx}(\omega) = \left| \int_{0}^{T} x(t) e^{-j\omega t}\, dt \right|^2. \tag{2.3}$$

In the discrete case, one has the following for a discrete signal $x(n)$ [21–23]:

$$\widehat{S}_{xx}(\omega) = \left| \sum_{n=0}^{N-1} x(n) e^{-j\omega n} \right|^2. \tag{2.4}$$

Because

$$\left| \sum_{n=L}^{N+L-1} x(n) e^{-j\omega n} \right|^2 \neq \left| \sum_{n=M}^{N+M-1} x(n) e^{-j\omega n} \right|^2 \quad \text{for } L \neq M, \tag{2.5}$$

$\widehat{S}_{xx}(\omega)$ is usually a random variable. One way of reducing the variance of $\widehat{S}_{xx}(\omega)$ is to smooth $\widehat{S}_{xx}(\omega)$ by a smoother function denoted by $G(\omega)$. Denote by $\widetilde{S}_{xx}(\omega)$ the smoothed PSD estimate. Let $*$ imply the operation of convolution. Then, $\widetilde{S}_{xx}(\omega)$ is given by

$$\widetilde{S}_{xx}(\omega) = \widehat{S}_{xx}(\omega) * G(\omega). \tag{2.6}$$

Assume that $\widetilde{S}_{xx}(\omega)$ is differentiable any time for $-\infty < \omega < \infty$. Then, by using the Taylor series at $\omega = \omega_0$, $\widehat{S}_{xx}(\omega)$ is expressed by

$$\widehat{S}_{xx}(\omega) = \sum_{l=0}^{\infty} \frac{\widehat{S}_{xx}^{(l)}(\omega_0)}{l!} (\omega - \omega_0)^n. \tag{2.7}$$

Therefore,

$$\widetilde{S}_{xx}(\omega) = \sum_{l=0}^{\infty} \frac{\widehat{S}_{xx}^{(l)}(\omega_0)}{l!} (\omega - \omega_0)^n * G(\omega). \tag{2.8}$$

Let $\omega - \omega_0 = \omega_1$. Then,

$$(\omega - \omega_0)^n * G(\omega) = \omega_1^n * G(\omega_1 + \omega_0). \tag{2.9}$$

Thus, we have a theorem to represent $\widetilde{S}_{xx}(\omega)$ based on the Cauchy integral.

Theorem 2.1. *Suppose $\widehat{S}_{xx}(\omega)$ is differentiable any time at ω_0. Then, the smoothed PSD, that is, $\widetilde{S}_{xx}(\omega)$, may be expressed by*

$$\widetilde{S}_{xx}(\omega) = \sum_{l=0}^{\infty} \frac{\widehat{S}_{xx}^{(l)}(\omega_0)}{l!} \omega_1^l * G(\omega_1 + \omega_0) = \sum_{l=0}^{\infty} \widehat{S}_{xx}^{(l)}(\omega_0) \int_0^{\omega_1} \frac{(\omega_1 - \omega_\tau)^l}{l!} G(\omega_\tau + \omega_0)\, d\omega_\tau. \tag{2.10}$$

Proof. The Cauchy integral in terms of $G(\omega_\tau + \omega_0)$ is in the form

$$\int_0^{\omega_1} \frac{(\omega_1 - \omega_\tau)^l}{l!} G(\omega_\tau + \omega_0)\, d\omega_\tau = \underbrace{\int_0^{\omega_\tau} d\omega_\tau \cdots \int_0^{\omega_\tau} G(\omega_\tau + \omega_0)\, d\omega_\tau}_{l+1}. \tag{2.11}$$

That may be taken as the convolution between $\omega_1^l / l!$ and $G(\omega_1 + \omega_0)$. Thus,

$$\frac{\omega_1^l}{l!} * G(\omega_1 + \omega_0) = \int_0^{\omega_1} \frac{(\omega_1 - \omega_\tau)^l}{l!} G(\omega_\tau + \omega_0)\, d\omega_\tau. \tag{2.12}$$

Therefore, (2.10) holds. This completes the proof. □

The present theorem is a theoretic representation of a smoothed PSD estimate. It may yet be a method to be used in the approximation of a smoothed PSD estimate. As a matter of fact, we may approximate $\widetilde{S}_{xx}(\omega)$ by a finite series given by

$$\widetilde{S}_{xx}(\omega) \approx \sum_{l=0}^{L} \widehat{S}_{xx}^{(l)}(\omega_0) \int_0^{\omega_1} \frac{(\omega_1 - \omega_\tau)^l}{l!} G(\omega_\tau + \omega_0)\, d\omega_\tau. \tag{2.13}$$

From the above theorem, we have the following corollary.

Corollary 2.2. *Suppose $\widehat{S}_{xx}(\omega)$ is differentiable any time at $\omega = 0$. Then, $\widetilde{S}_{xx}(\omega)$ may be expressed by*

$$\widetilde{S}_{xx}(\omega) = \sum_{l=0}^{\infty} \frac{\widehat{S}_{xx}^{(l)}(0)}{l!} \omega^l * G(\omega) = \sum_{l=0}^{\infty} \widehat{S}_{xx}^{(l)}(0) \int_0^{\omega} \frac{(\omega - \omega_\tau)^l}{l!} G(\omega_\tau)\, d\omega_\tau. \tag{2.14}$$

The proof is omitted since it is straightforward when one takes into account the proof of theorem.

3. Conclusions

We have presented a theorem with respect to a representation of a smoothed PSD estimate of signals based on the Cauchy integral. The theorem constructively implies that the design of a smoother function $G(\omega)$ may consider the approximation described by the Cauchy integral with the finite Taylor series (2.13). In addition, the smoother function $G(\omega)$ can also be taken as a solution to the integral equation (2.14), which is worth being investigated in the future.

Acknowledgments

This work was supported in part by the 973 plan under the Project Grant no. 2011CB302800 and by the National Natural Science Foundation of China under the Project Grant nos. 61272402, 61070214, and 60873264.

References

[1] R. Kramme, K. P. Hoffmann, and R. S. Pozos, *Springer Handbook of Medical Technology*, Springer, New York, NY, USA, 2012.

[2] Wang and S. P. Patrick, Eds., *Pattern Recognition, Machine Intelligence and Biometrics*, Springer, Berlin, Germany; Higher Education Press, Beijing, China, 2011.

[3] V. Capasso and D. Bakstein, *An Introduction to Continuous-Time Stochastic Processes: Theory, Models, and Applications to Finance, Biology, and Medicine*, Birkhauser, Berlin, Germany, 2005.

[4] N. Rosario, H. Mantegna, and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, New York, NY, USA, 2000.

[5] L. Anselin and S. J. Rey, Eds., *Perspectives on Spatial Data Analysis*, Springer, New York, NY, USA, 2010.

[6] C. Cattani and J. Rushchitsky, *Wavelet and Wave Analysis as Applied to Materials with Micro or Nanostructure*, World Scientific, Singapore, 2007.

[7] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.

[8] C. Cattani, "On the existence of wavelet symmetries in Archaea DNA," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 673934, 21 pages, 2012.

[9] C. Cattani, E. Laserra, and I. Bochicchio, "Simplicial approach to fractal structures," *Mathematical Problems in Engineering*, vol. 2012, Article ID 958101, 21 pages, 2012.

[10] C. Cattani, "Fractional calculus and Shannon wavelet," *Mathematical Problems in Engineering*, vol. 2012, Article ID 502812, 26 pages, 2012.

[11] C. Toma, "Advanced signal processing and command synthesis for memory-limited complex systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 927821, 13 pages, 2012.

[12] Z. Liao, S. Hu, D. Sun, and W. Chen, "Enclosed Laplacian operator of nonlinear anisotropic diffusion to preserve singularities and delete isolated points in image smoothing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 749456, 15 pages, 2011.

[13] S. Hu, Z. Liao, D. Sun, and W. Chen, "A numerical method for preserving curve edges in nonlinear anisotropic smoothing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 186507, 14 pages, 2011.

[14] E. G. Bakhoum and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.

[15] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, San Diego, Calif, USA, 1999.

[16] P. Flandrin and P. Borgnat, "Time-frequency energy distributions meet compressed sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 2974–2982, 2010.

[17] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: a time-frequency approach," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3459–3470, 2010.

[18] M. Orini, R. Bailon, L. Mainardi, P. Laguna, and P. Flandrin, "Characterization of dynamic interactions between cardiovascular signals by time-frequency coherence," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 663–673, 2011.

[19] J. Xiao and P. Flandrin, "Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2851–2860, 2007.
[20] G. Rilling and P. Flandrin, "Sampling effects on the empirical mode decomposition," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 43–59, 2009.
[21] E. A. Robinson, "A historical perspective of spectrum estimation," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 885–907, 1982.
[22] S. K. Mitra and J. F. Kaiser, *Handbook For Digital Signal Processing*, Wiley, Chichester, UK, 1993.
[23] A. Papoulis, *Signal Analysis*, McGraw-Hill, New York, NY, USA, 1977.
[24] M. S. Bartlett, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, no. 1-2, pp. 1–16, 1950.
[25] J. Niu, K. Li, W. D. Jiang, X. Li, G. Y. Kuang et al., "A new method of micro-motion parameters estimation based on cyclic autocorrelation function," *Science China Information Sciences*. In press.
[26] P. C. Mu, D. Li, Q. Y. Yin, and W. Guo, "Robust MVDR beamforming based on covariance matrix reconstruction," *Science China Information Sciences*. In press.
[27] D. Wang and Y. Wu, "Effects of finite samples on the resolution performance of the rank reduction estimator," *Science China Information Sciences*. In press.
[28] X. Zhang, Y. Zhang, J. Zhang, S. Y. Chen, D. Chen, and X. Li, "Unsupervised clustering for logo images using singular values region covariance matrices on Lie groups," *Optical Engineering*, vol. 51, no. 4, article 047005, 2012.
[29] N. M. Kwok, X. Jia, D. Wang, S. Y. Chen, Q. P. Ha, and G. Fang, "Visual impact enhancement via image histogram smoothing and continuous intensity relocation," *Computers & Electrical Engineering*, vol. 37, no. 5, pp. 681–694, 2011.
[30] A. Papoulis, "Minimum-bias windows for high-resolution spectral estimates," *IEEE Transactions on Information Theory*, vol. 19, no. 1, pp. 9–12, 1973.
[31] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 275138, 10 pages
doi:10.1155/2012/275138

Research Article

Airborne Infrared and Visible Image Fusion for Target Perception Based on Target Region Segmentation and Discrete Wavelet Transform

Yifeng Niu,[1,2] Shengtao Xu,[3] Lizhen Wu,[1] and Weidong Hu[2]

[1] *College of Mechatronics and Automation, National University of Defense Technology, Changsha 410073, China*
[2] *ATR State Key Laboratory, National University of Defense Technology, Changsha 410073, China*
[3] *Jiuquan Satellite Launch Centre, Lanzhou 732750, China*

Correspondence should be addressed to Yifeng Niu, niuyifeng@nudt.edu.cn

Received 18 September 2012; Accepted 22 October 2012

Academic Editor: Carlo Cattani

Infrared and visible image fusion is an important precondition of realizing target perception for unmanned aerial vehicles (UAVs), then UAV can perform various given missions. Information of texture and color in visible images are abundant, while target information in infrared images is more outstanding. The conventional fusion methods are mostly based on region segmentation; as a result, the fused image for target recognition could not be actually acquired. In this paper, a novel fusion method of airborne infrared and visible image based on target region segmentation and discrete wavelet transform (DWT) is proposed, which can gain more target information and preserve more background information. The fusion experiments are done on condition that the target is unmoving and observable both in visible and infrared images, targets are moving and observable both in visible and infrared images, and the target is observable only in an infrared image. Experimental results show that the proposed method can generate better fused image for airborne target perception.

1. Introduction

Unmanned aerial vehicles (UAVs) are aircrafts which have the capability of flight without an onboard pilot. UAV can be remotely controlled, semiautonomous, autonomous, or have a combination of these capabilities. UAV can execute given missions in a whole lot of domains [1]. In order to complete these various missions, UAV firstly needs to be equipped with sensor payloads to acquire images of the mission area and realize environment perception. The sensor payloads include infrared sensor, and visible light sensor. Infrared and visible image fusion from the same scene is the basis of target detection and recognition for UAV.

However, the images shot by airborne sensors are dynamic, which increases more difficulties for visible and infrared image fusion. In order to acquire the situation assessment, it is very important to extract the target information. In fact, information of texture and color in visible images are very abundant, while the target information, especially an artificial target, is more outstanding in infrared images. According to this, we can divide the image regions based on target regions, which can utilize the target information more effectively.

Image fusion can be performed at four levels of the information representation, which are signal, pixel, feature, and symbolic levels [2]. An infrared and visible image fusion method based on region segmentation is proposed in [3, 4] and adopted the nonsubsampled contourlet transform (NSCT) to fuse the regions given and optimize the quality of the fused image, while the method did not capture the target information effectively. Yin et al. proposed an infrared and visible image fusion method based on color contrast enhancement which can be propitious to target detection and improve the observer performance [5]. Hong et al. presented a fusion framework for infrared and visible images based on data assimilation and genetic algorithm (GA) [6]. Shao et al. introduced fast discrete curvelet transform (FDCT) and focus measure operators to realize the fusion of infrared and visible images [7]. Other fusion methods for the infrared and color visible images have been compared in [8–10], in which the fusion method based on discrete wavelet transform (DWT) performed well. Though NSCT, FDCT, and other novel transforms are superior to DWT in some performances, the transform should not be the key and core problem for infrared and visible image fusion [11]. In this paper, we only use DWT as a meaning to research the fusion concerning target detection and perception, for DWT has less computational complexity [12]. Yao et al. proposed a new approach to airborne visible and infrared image fusion [13, 14] and researched target fusion detection [15], which used the relevant information in different frames that could meet the demand for low real-time fusion.

Dynamic image fusion has its own characteristics, which require that the fusion method is consentaneous and robust to both time and space [16–18]. In order to utilize the different region features and get more effective target and background information, a method of visible and infrared image fusion in DWT domain based on dynamic target detection and target region segmentation is proposed. First, the image segmentation is done based on the detected candidate target regions, and then the information between the frames can be used to attain the stability and consensus to time. Finally, the different fusion rules are designed according to the characteristics of target regions to complete visible and infrared image fusion.

2. Image Segmentation Based on Target Regions

According to the requirements of airborne image processing for the accuracy and the speed, the moving target in the airborne images can be detected using the frame difference method based on background motion compensation [19]. The algorithm flow of target region detection is shown in Figure 1.

2.1. Motion Information Extraction Based on Frame Difference

On the basis of motion compensation for the image background, target detection can be done by applying the frame difference method to the image sequences. The regions whose pixel values are constant can be regarded as background regions, while the regions whose pixel

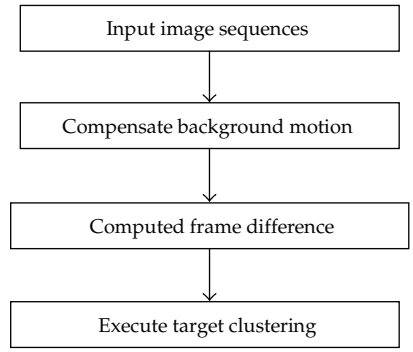


Figure 1: Flow of target region detection.

values change with different points are moving target regions, which include the motion information of targets. Using the converse transform parameters, we can make motion compensation for frame $n + 1$ and compute the difference between the former frame and the current.

$$D_n(x, y) = \left| I_n(x, y) - I_{n+1}(x, y) \right|, \tag{2.1}$$

where $D_n(x, y)$ denotes the frame difference of frame $n+1$ at point (x, y) and $I_n(x, y)$ denotes the pixel value of frame n at point (x, y). The change of pixel values in unmoving regions should be zero; however, because of some random noises, luminance change, and weather change, the differences are fairly small and belong to the salt and pepper noise. In order to extract the moving target regions, we need to select a proper threshold to segment the source images. Thus, we can acquire the moving target regions and relevant motion information. This method is easy to realize and robust to illumination, which could be a proper method for airborne image processing.

2.2. *Target Clustering and Image Segmentation*

After getting motion information from the neighbor frames, each target region has not been differentiated, and some unpredicted noises may appear in the target regions. In order to distinguish the different moving regions and filter noisy points in the target regions, the target clustering algorithm is proposed, which is shown in Figure 2.

The criterion for judging whether the points belong to a certain cluster is distance. According to the transcendental knowledge, the distances among different clusters can be established. The threshold can be set beyond which a new cluster can be added. When each target is distinguished, the number and the range of points in each cluster need to be computed. The points whose quantity is less than a certain number can be regarded as noisy points or false points and then can be eliminated. For the confirmed target regions, the region size and boundary can be computed, and then each target region can be marked out in the source images.

Suppose the visible image is E_1 and the infrared image is I_1, each image is to be divided into two parts at least, target region of R and background region of B, shown in Figure 3. There exist several target regions in source images. As shown in Figure 3, two source images

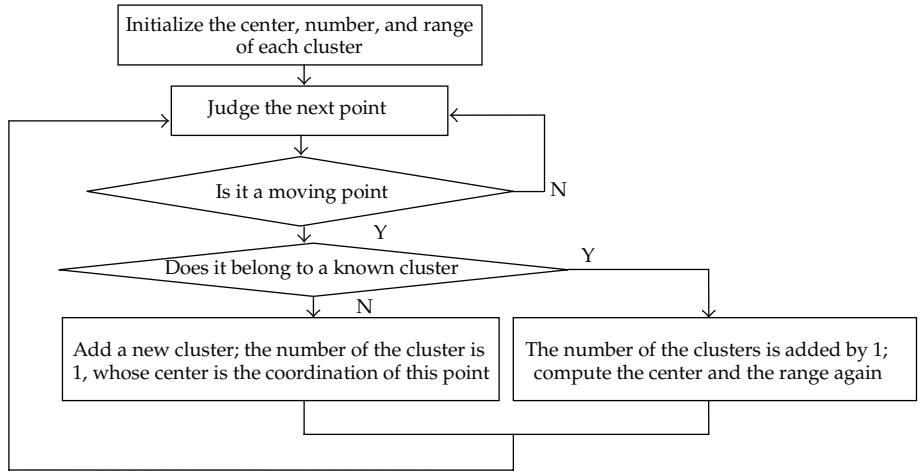
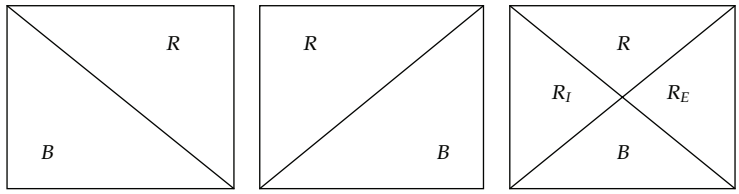


Figure 2: Flow of target clustering.



(a) Visible image (b) Infrared image (c) Region segmentation

Figure 3: Illustration of region segmentation based on target regions.

are superposed, and four regions come into being, including common target region of R, visible target region of R_E, infrared target region of R_I, and common background region of B. For different regions, the different fusion strategies and methods are adapted to gain better fusion effect.

3. Visible and Infrared Image Fusion Based on Target Regions in DWT Domain

First, input the registered source visible image E_1 and infrared image I_1. Find the DWT of each E_1 and I_1 to a specified number of decomposition levels of N. Suppose (x, y) denotes the coordination of any pixel point, $g_i^E(x, y)$ and $f_N^E(x, y)$ denote high-frequency and low-frequency subband coefficients of the visible image, respectively, $g_i^I(x, y)$ and $f_N^I(x, y)$ denote the high-frequency and low-frequency sub band coefficients of the infrared image, respectively, F denotes the fused image, and $g_i^F(x, y)$ and $f_N^F(x, y)$ denote the high-frequency and low-frequency sub band coefficients of the fused image. The fusion rules of different regions are as follow.

Step 1. For the target region that only exists in visible image or infrared image, the high-frequency and low-frequency subband coefficients of F can be computed as follow:

$$g_i^F(x,y) = \begin{cases} g_i^E(x,y) & \text{if } (x,y) \in R_E, \\ g_i^I(x,y) & \text{if } (x,y) \in R_I, \end{cases}$$
$$f_i^F(x,y) = \begin{cases} f_N^E(x,y) & \text{if } (x,y) \in R_E, \\ f_N^I(x,y) & \text{if } (x,y) \in R_I. \end{cases} \tag{3.1}$$

Step 2. For the common target region R, according to similarity measurement of two images, we use the selective or the weighted average rule to fuse the source images. The similarity of the two images (M_{EI}) in region R is defined as

$$M_{EI}(R) = \frac{2\sum_{(x,y)\in R} I_E(x,y)\cdot I_I(x,y)}{\sum_{(x,y)\in R}\left[I_E(x,y)\right]^2 + \sum_{(x,y)\in R}\left[I_I(x,y)\right]^2}. \tag{3.2}$$

Then compute the energy of all high-frequency sub band coefficients in region R of the two images, and use it as the fusion measure

$$S^P(R) = \sum_{(x,y)\in R}\sum_{i}^{N} g_i^P(x,y)^2, \quad P = E, I. \tag{3.3}$$

If $M_{EI}(R) < \alpha$, where α is a similarity threshold, such as 0.8, then apply the selective fusion rule to the two image

$$g_i^F(x,y) = \begin{cases} g_i^E(x,y), & \text{if } S^E(R) \geq S^I(R), \\ g_i^I(x,y), & \text{if } S^E(R) < S^I(R), \end{cases}$$
$$f_N^F(x,y) = \begin{cases} f_N^E(x,y), & \text{if } S^E(R) \geq S^I(R), \\ f_N^I(x,y), & \text{if } S^E(R) < S^I(R). \end{cases} \tag{3.4}$$

If $M_{EI}(R) \geq \alpha$, then apply the weighted average fusion rule to the two images and define the weighted coefficients

$$\omega_{\min} = \frac{1}{2}\left(1 - \frac{1 - M_{EI}(R)}{1-\alpha}\right),$$
$$\omega_{\max} = 1 - \omega_{\min}. \tag{3.5}$$

Then fusion process in the common target region R can be described as follows:

$$g_i^F(x,y) = \begin{cases} \omega_{\max} g_i^E(x,y) + \omega_{\min} g_i^I(x,y), & \text{if } S^E(R) \geq S^I(R), \\ \omega_{\max} g_i^I(x,y) + \omega_{\min} g_i^E(x,y), & \text{if } S^E(R) < S^I(R), \end{cases}$$
$$f_N^F(x,y) = \begin{cases} \omega_{\max} f_N^E(x,y) + \omega_{\min} f_N^I(x,y), & \text{if } S^E(R) \geq S^I(R), \\ \omega_{\max} f_N^I(x,y) + \omega_{\min} f_N^E(x,y), & \text{if } S^E(R) < S^I(R). \end{cases} \tag{3.6}$$

Step 3. For the common background region B, take the different fusion strategies for the high-frequency and low-frequency sub band, respectively. For low-frequency sub band, we use the average method for fusion directly

$$f_N^F(x,y) = 0.5 f_N^E(x,y) + 0.5 f_N^I(x,y). \tag{3.7}$$

For high-frequency sub band, we adopt the fusion rule based on windows proposed by Burt and Kolczynski [20]. First, compute the energy of high-frequency sub band coefficients in local window, and use it as the fusion measure

$$S^P(x,y) = \sum_{(m,n)\in N(x,y)} \omega(m,n) \cdot g_i^P(m,n)^2, \quad P = E, I, \tag{3.8}$$

where $N(x,y)$ is the window whose center is (x,y), $\omega(m,n)$ is a weighted coefficient, and the sum is 1. According to the similarity fusion method in a local window, the similarity is defined as

$$M_i(x,y) = \frac{2 \cdot \sum_{(m,n)\in N(x,y)} \omega(m,n) \cdot \left| g_i^E(m,n) \right| \cdot \left| g_i^I(m,n) \right|}{S_i^P(x,y) + S_i^I(x,y)}. \tag{3.9}$$

If $M_i(x,y) < \alpha_i$, where α_i is the similarity threshold, then apply the selected fusion rule for the two image

$$g_i^F(x,y) = \begin{cases} g_i^E(x,y), & \text{if } S^E(x,y) \geq S^I(x,y), \\ g_i^I(x,y), & \text{if } S^E(x,y) < S^I(x,y). \end{cases} \tag{3.10}$$

If $M_i(x,y) \geq \alpha_i$, then apply the selected fusion rule to the two image, take the weighted average fusion rule for the two images, and define the weighted coefficients

$$\omega_{\min} = \frac{1}{2}\left(1 - \frac{1 - M_i(x,y)}{1 - \alpha_i}\right),$$
$$\omega_{\max} = 1 - \omega_{\min}. \tag{3.11}$$

Table 1: Evaluation metrics of different image fusion methods.

Fusion method	Entropy	Mutual information	Edge preservation
Weighted average	4.4058	0.9521	0.4152
Principal component analysis	6.1254	1.3214	0.7124
Laplacian pyramid	6.3547	1.4521	0.7457
Gradient pyramid	5.9687	1.3542	0.7012
Contrast pyramid	6.3142	1.4217	0.7315
Discrete wavelet transform	6.5214	1.4754	0.7462
Our fusion method	6.5321	1.5531	0.8213

The fusion process in the common background region B can be described as follows:

$$g_i^F(x,y) = \begin{cases} \omega_{\max} g_i^E(x,y) + \omega_{\min} g_i^I(x,y), & \text{if } S^E(x,y) \geq S^I(x,y), \\ \omega_{\max} g_i^I(x,y) + \omega_{\min} g_i^F(x,y), & \text{if } S^E(x,y) < S^I(x,y). \end{cases} \tag{3.12}$$

4. Results and Analysis of Fusion Experiments

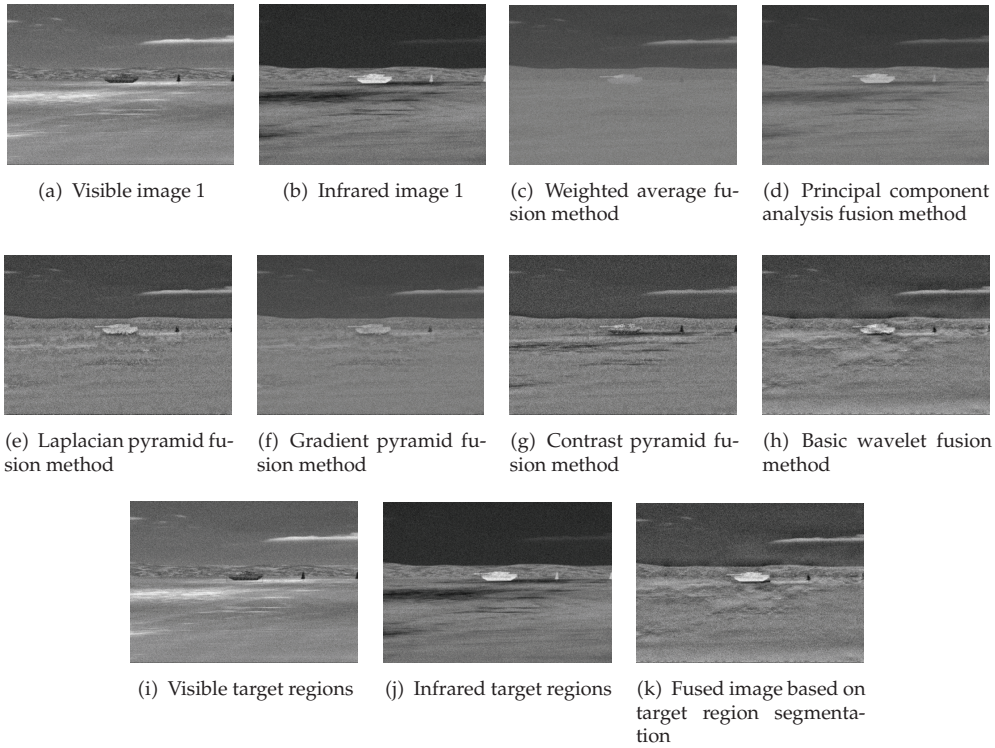
4.1. Target Is Unmoving and Observable Both in Visible and Infrared Images

First, the experiment to fuse the visible and infrared images in which the target is unmoving and observable in both of the two images is conducted. There exists a static tank target in source images. We can get the small rectangle region including the tank-based target region segmentation, as shown in Figures 4(i) and 4(j), other part is the background region, thus, the source image can be divided into different parts, and different fusion rules are adopted to produce the fused image. As the source images are fused using different pyramid methods, the decomposition level is 3, while the fusion method based on discrete wavelet transform adopt the decomposition level of 2.

The evaluation metrics of different fusion method are given in Table 1. The definition of these metrics can be found in [21, 22]. It can be seen that our method not only preserves abundant texture information, but also image edges maintain well and the fusion effect is better than other methods. In fact, result of weighted average method is blurring in background, and target is not clear, which shows that the source information has not been utilized efficiently. The results of pyramid decomposition methods are abundant in background, but the target information has not been preserved. The basic wavelet method preserves both the details and the target information. However, all of above methods are based on one rule, and the whole image use the same fusion rule, which has not used different fusion rules according to different region characteristics and inevitably result in losing information. The result of our method which adopts different fusion rules according to the different regions shows that the background information is more abundant, and the target is more outstanding.

4.2. Target Is Moving and Observable Both in Visible and Infrared Images

Second, the experiment to fuse the visible and infrared images in which the target is moving and observable both in the two images is conducted (note: the source images are from [16]).



(a) Visible image 1 (b) Infrared image 1 (c) Weighted average fusion method (d) Principal component analysis fusion method

(e) Laplacian pyramid fusion method (f) Gradient pyramid fusion method (g) Contrast pyramid fusion method (h) Basic wavelet fusion method

(i) Visible target regions (j) Infrared target regions (k) Fused image based on target region segmentation

Figure 4: Results on condition that the target is observable both in visible and infrared image.

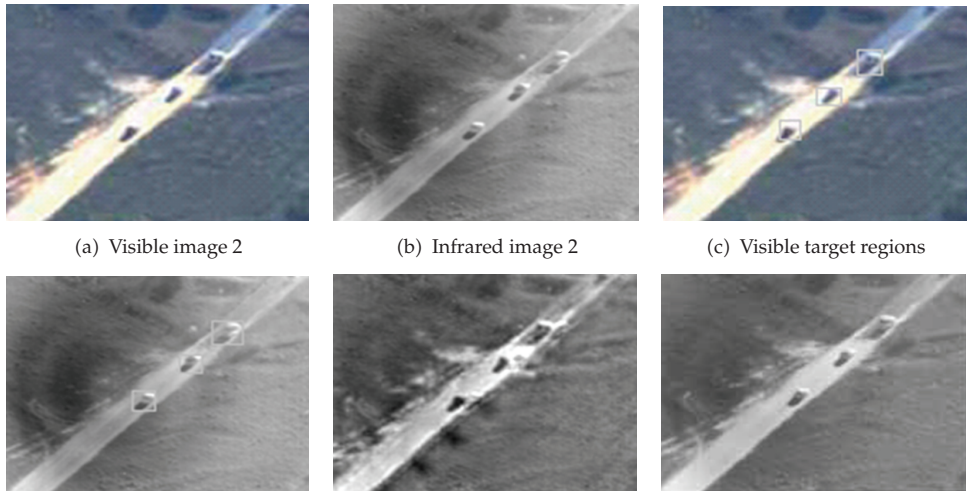
There exist several moving vehicle targets in source images which increases more difficulties for multiple source image fusion. Through target detection, the source images can be divided into four different regions, that are target R_1, R_2, R_3, and R_4 and the background, which be shown in Figures 5(c) and 5(d). We adopt different fusion rules according to the different regions and make a contrast research to the basic wavelet method. The results in Figures 5(e) and 5(f) show that the background information is more abundant, and the target is more outstanding than those of the basic wavelet method.

4.3. Target Is Observable Only in Infrared Image

In the actual environment, the target can be sheltered from other objects, which can be observed by the infrared sensor, such as Figures 6(a) and 6(b), in which the person target is unobservable in the visible image, while it is observable in the infrared image. From Figures 6(c) and 6(d), it can be seen that the target information changes using the basic wavelet method, which becomes dark inside and blur along the edge, while the target is clear using our method which means that the target information and background information have been mostly preserved and can meet demand of the target extraction.

5. Conclusions

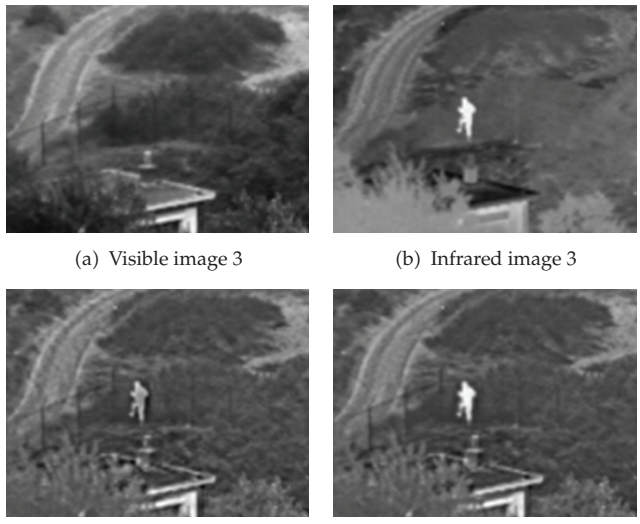
In this paper, we proposed a new approach to infrared and visible image fusion based on detected target regions in DWT domain, which can help UAV to realize environment



(a) Visible image 2 (b) Infrared image 2 (c) Visible target regions

(d) Infrared target regions (e) Fused image by basic wavelet method (f) Fused image based on target region segmentation

Figure 5: Results on condition that targets are moving both in visible and infrared image.



(a) Visible image 3 (b) Infrared image 3

(c) Fused image by basic wavelet method (d) Fused image based on target region segmentation

Figure 6: Results on condition that the target is observable only in infrared image.

perception. Other than the conventional fusion methods based on region segmentation, we proposed a frame difference method for target detection, which can be used to segment the source images, and then we design different fusion rules based on target regions to fuse the visible and infrared images, which can gain more target information and preserve more background information in the source images. In the future, the method can be spread to other source image for fusion, and the time performance of our method can be improved based on GPU (Graphics Processing Unit) and other hardware.

Acknowledgment

This work is supported by a special financial grant from the China Postdoctoral Science Foundation (Grant no. 20100481512, 201104765).

References

[1] R. R. Pitre, X. R. Li, and R. Delbalzo, "UAV route planning for joint search and track missions-an information-value approach," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2551–2565, 2012.

[2] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, no. 4, pp. 259–280, 2003.

[3] B. L. Guo, Q. Zhang, and Y. Hou, "Region-based fusion of infrared and visible images using nonsubsampled contourlet transform," *Chinese Optics Letters*, vol. 6, no. 5, pp. 338–341, 2008.

[4] L. Kun, G. Lei, L. Huihui, and C. Jingsong, "Fusion of infrared and visible light images based on region segmentation," *Chinese Journal of Aeronautics*, vol. 22, no. 1, pp. 75–80, 2009.

[5] S. F. Yin, L. C. Cao, Y. S. Ling, and G. Jin, "One color contrast enhanced infrared and visible image fusion method," *Infrared Physics and Technology*, vol. 53, no. 2, pp. 146–150, 2010.

[6] L. Hong, Z. H. He, J. Xiang, and S. Li, "Fusion of infrared and visible image based on genetic algorithm and data assimilation," in *Proceedings of International Workshop on Intelligent Systems and Applications, (ISA '09)*, pp. 1–5, May 2009.

[7] Z. F. Shao, J. Liu, and Q. M. Cheng, "Fusion of infrared and visible images based on focus measure operators in the curvelet domain," *Applied Optics*, vol. 51, no. 12, pp. 1910–1921, 2012.

[8] J. Y. Liu, Q. Wang, and Y. Shen, "Comparisons of several pixel-level image fusion schemes for infrared and visible light images," in *Proceedings of the IEEE Instrumentation and Measurement Technology Conference, (IMTC '05)*, vol. 3, pp. 2024–2027, May 2005.

[9] X. Q. Zhang, Q. L. Chen, and T. Men, "Comparison of fusion methods for the infrared and color visible images," in *Proceedings of 2nd IEEE International Conference on Computer Science and Information Technology, (ICCSIT '09)*, pp. 421–424, August 2009.

[10] J. Saeedi and K. Faez, "Infrared and visible image fusion using fuzzy logic and population-based optimization," *Applied Soft Computing*, vol. 12, no. 3, pp. 1041–1054, 2012.

[11] J. Zhang, S. Y. Chen, S. Liu et al., "Normalized weighted shape context and its application in feature-based matching," *Optical Engineering*, vol. 47, no. 9, Article ID 097201, 2008.

[12] C. Cattani and A. Kudreyko, "On the discrete harmonic wavelet transform," *Mathematical Problems in Engineering*, vol. 2008, Article ID 687318, 2008.

[13] F. H. Yao and A. Sekmen, "Multi-source airborne IR and optical image fusion and its application to target detection," *Lecture Notes in Computer Science*, vol. 5359, no. 2, pp. 651–660, 2008.

[14] F. H. Yao, A. Sekmen, and M. Malkani, "A novel method for real-time multiple moving targets detection from moving IR camera," in *Proceedings of 19th International Conference on Pattern Recognition (ICPR '08)*, pp. 1–4, December 2008.

[15] F. H. Yao, G. F. Shao, A. Sekmen, and M. Malkani, "Real-time multiple moving targets detection from airborne IR imagery by dynamic Gabor filter and dynamic gaussian detector," *Eurasip Journal on Image and Video Processing*, vol. 2010, no. 9, Article ID 124681, pp. 1–22, 2010.

[16] C. Cattani, S. Y. Chen, and G. Aldashev, "Information and modeling in complexity," *Mathematical Problems in Engineering*, vol. 2012, Article ID 868413, 4 pages, 2012.

[17] S. Y. Chen, Y. H. Wang, and C. Cattani, "Key issues in modeling of complex 3D structures from video sequences," *Mathematical Problems in Engineering*, vol. 2012, Article ID 856523, 17 pages, 2012.

[18] S. Y. Chen, J. H. Zhang, Y. F. Li et al., "A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 118–127, 2012.

[19] S. T. Xu, Y. F. Niu, L. C. Shen et al., "Real-time moving target detection method form unmanned airborne camera image," *Computer Engineering*, vol. 36, pp. 271–273, 2010 (Chinese).

[20] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Proceeding of the 4th International Conference on Computer Vision (ICCV '93)*,, pp. 173–182, May 1993.

[21] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 493–500, 2007.

[22] Y. H. Yuan, J. L. Zhang, B. K. Chang, and Y. Han, "Objective quality evaluation of visible and infrared color fusion image," *Optical Engineering*, vol. 50, no. 3, Article ID 033202, 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 582126, 11 pages
doi:10.1155/2012/582126

Research Article

The Johnson Noise in Biological Matter

Massimo Scalia,[1] Massimo Sperini,[2] and Fabrizio Guidi[3]

[1] *Department of Mathematics, Istituto "G. Castelnuovo", University of Rome "La Sapienza", Piazzale Aldo Moro, 5, 00185 Rome, Italy*
[2] *Technical Institute "R. Rossellini", Via della Vasca Navale, 58 CP., 00146 Rome, Italy*
[3] *Naval Technical Institute "M. Colonna", Via Salvatore Pincherle, 201 CP., 00146 Rome, Italy*

Correspondence should be addressed to Massimo Scalia, massimo.scalia@uniroma1.it

Received 28 September 2012; Accepted 6 October 2012

Academic Editor: Carlo Cattani

Can a very low intensity signal overcome a disturbance, the power density of which is much higher than the signal one, and yield some observable effects? The Johnson noise seems to be a disturbance so high as to cause a negative answer to that question, when one studies the effects on the cell level due to the external ELF fields generated by electric power lines (Adair, 1990, 1991). About this subject, we show that the masking effect due to the Johnson noise, known as "Adair's constraint" and still present in the scientific debate, can be significantly weakened. The values provided by the Johnson noise formula, that is an approximate expression, can be affected by a significant deviation with respect to the correct ones, depending on the frequency and the kind of the cells, human or not human, that one is dealing with. We will give some examples. Eventually, we remark that the so-called Zhadin effect, although born and studied in a different context, could be viewed as an experimental test that gives an affirmative answer to the initial question, when the signal is an extremely weak electromagnetic field and the disturbance is a Johnson noise.

1. Introduction

Much attention has been devoted over many decades to a problem that arose in the electrocommunications field [1–3], but such that its consequences extend beyond telecommunications, electroengineering, or other specific problems to a general question, that we try to summarize in this way: can a very low intensity signal overcome a disturbance, the power density of which is much higher than the signal one, and yield some observable effects?

A problem of this kind has become a matter of some lively discussions inside the scientific community, in the case of the biological or health effects due to electromagnetic fields of low intensity in the *whole region* of nonionizing radiations (NIR). Just in this context a relevant influence has been played along all these years, in the scientific debate,

by the statement of the Council of the APS (American Physical Society), that excluded the promotion of cancers by power line fields [4]. Actually, the directions of the APS statement disregarded not only cancer risks but also almost all risks associated to the exposure to the fields generated by power lines. That statement has been reaffirmed by the Council of the APS in a more recent brief note, with a further consideration: "*... In addition no biophysical mechanism for the initiation or promotion of cancer by electric or magnetic fields from power lines have been identified*" [5].

The aim of the present paper is not to enter this debate. But, among the scientific bases of the APS position paper, recalled in [6], we look at the one that has played a so relevant role up today in the debate concerning physicists and all researchers in bioelectromagnetism, such to merit the name of "Adair's constraint". Really, Adair proposed the concept of "thermal noise electric field" in his papers [7, 8], and his "constraint" extends far beyond the possible cancer effects: "*... any effects on the cell level of fields in the body generated by weak external ELF fields will be masked by thermal noise effects and, hence, such fields cannot be expected to have* any *significant effect on the biological activities of the cells*" [8]. This statement was very strong, despite of important contemporary review papers about the interaction of ELF fields with humans [9], in which hundreds of references had described laboratory and clinical studies of the effects at the cellular level of an exposure to 50–60 Hz electric and magnetic fields: "*a substantial amount of experimental evidence obtained with in vitro cell and organ cultures indicates that pericellular currents produced by ELF (extremely low frequency) fields lead to structural and functional alterations in components of the cell membrane*" [10]. Since the "Adair's constraint" is still working in the scientific discussion (see, e.g., [11]), we mean to show that it can be significantly weakened and, at the same time, to give a contribution to avert a possible generalization of the masking effect to all the NIR spectrum.

In the current literature many articles about the thermal noise in biological cells recur to the approximate expression of the Johnson noise, as the Adair's ones do. We note that a bit of caution is necessary when dealing with that formula, in order to avoid that the mean square values of the noise tension differ significantly from the correct ones. This is the case of the estimates of exposure of a human cell membrane to a VLF antenna, that radiates in the range (14–30) kHz. On the contrary, the Johnson's formula works well when applied to the irradiation of eggs of Salmo lacustris from an oscillating magnetic field in the range (450–1350) kHz, a historical experiment performed by Italian researchers in the Twenties.

Eventually, we remark that the so-called Zhadin effect, although born and studied in a different context, could be viewed as an experimental test proving that an affirmative answer to the initial question can be given, in the particular case when the very low signal is an extremely weak electromagnetic field and the disturbance has just the character of a Johnson noise.

2. The Johnson Noise

Many of the reasoning and the models applied to describe interactions between electromagnetic fields and the biological matter rely on the phenomenon of Johnson noise, when the thermal noise has to be taken into account. And over this case we will focus, even though there are other kinds of noise, in order to compare the results of this paper with those presented in [7, 8]. Thus, it's necessary to recall, shortly, the experiment performed in electronics by Johnson and the corresponding formula.

Let us consider a conductor, in the interior of which there is a very large number of free to move electrons. At a fixed temperature T (°K), the thermal agitation of electrons implies

that the stochastic motion of the charged particles and their collisions produce at some time an accumulation of charge at one end of the conductor, while in a successive instant there will be an excess of electrons on the other end. Therefore, we could find a tension between the two terminals, that is a stochastic variable of the time with a well-defined mean square value: the thermal noise tension. This tension would not depend on a possible continuous electric current flowing in the conductor, since the thermal velocity of electrons is much higher than the drift velocity (~10^3 times).

Johnson was the first [1] who observed the fluctuations of tension at the ends of a conductor with a resistance R, realizing that the mean square value $\langle V^2 \rangle$ of the instantaneous noise tension $V(t)$ was proportional to R and to the absolute temperature T. Besides, he found that the ratio $\langle V^2 \rangle / R$ does not depend on the nature or the shape of conductors and assumed that this tension was due to the thermal agitation of the electrons inside the material [1]: the "Johnson noise" tension. Testing different kinds of conductors, he also found $\langle V^2 \rangle / R = kT$, where $k = 1.38 \cdot 10^{-23}$ J/k is the Boltzmann's constant, in a good agreement—within the 8%—with the experimental data [2]. Harry Nyquist had been immediately requested by Johnson of an explanation of the results of his experiment—they were at that time colleagues in the Bell Telephone Laboratories—and answered with an easy and fine conceptual experiment, a fundamental element of which is a "bipole" [3].

A bipole is any electric component with two terminals, characterized by a complex impedance

$$Z(\nu) = R(\nu) + jX(\nu), \tag{2.1}$$

where ν is the frequency (of the signal), j is the imaginary unit and the functions of frequency, R and X, depend on the capacity C and the inductance L of the bipole. Actually, since a pure resistor does not exist as a physical object, we are obliged to schematize a conductor as a bipole, also in a conceptual experiment. For T different from zero, the instantaneous noise tension fluctuates, also in the absence of an electric external field; but, for a pure resistor, its mean value cannot be other than 0:

$$V_n(t) : \langle V_n \rangle = 0, \tag{2.2}$$

where n stands for noise. Thus, over an interval of time we cannot have any electric field nor any electric current inside a pure resistor; only the instantaneous ones, led by the fluctuating tension, are permitted, otherwise we should have created a perpetual motion. And it is not by chance that in his experiment Johnson recurred, substantially, to the measure of the effective value of the noise tension as the available observable, that is the square root of $\langle {V_n}^2 \rangle$. The mean square value, $\langle {V_n}^2 \rangle$, at the ends of a bipole, is given by (see, e.g., [2, 3])

$$\left\langle {V_n}^2 \right\rangle_{\mathrm{TOT}} = 4kT \int_0^\infty R(\nu) d\nu, \tag{2.3}$$

where ν is the frequency of the noise spectrum. The measure devices do not allow to perform measures over the range of all frequencies as requested in (2.3), thus it is more frequently

used as the expression of the noise in a definite frequency band $[\nu_1, \nu_2]$:

$$\left\langle V_n{}^2 \right\rangle_{\text{PART}} = 4kT \int_{\nu 1}^{\nu 2} R(\nu) d\nu = 4kT \int_{\nu 1}^{\nu 2} \left\{ \frac{R}{\left[1 + (2\pi\nu RC)^2\right]} \right\} d\nu, \tag{2.4}$$

where the last equality holds only if the impedance Z reduces to that of an R-C conductor. When $R(\nu) = R$, one has

$$\left\langle V_n{}^2 \right\rangle_{\text{PART}} = 4kTR\Delta\nu, \tag{2.5}$$

with $\Delta\nu = [\nu_1, \nu_2]$. Expression (2.5) is just the Johnson noise formula: it holds when in the last term of (2.4) the condition $(2\pi\Delta\nu RC)^2 \ll 1$ is satisfied (and the integration domain is limited).

Thus, the limits of applicability of (2.5) are clear and it could be useful to emphasize that the thermal noise generated in a conductor for a given frequency band can be analyzed in terms of a pure resistor only if that condition is fulfilled.

3. The "Thermal Noise Electric Field" and the "Adair's Constraint"

Then, let us consider an R-C parallel circuit, that is often used to model biological tissues at thermal equilibrium (T = cost). Really, in most of the models the impedance circuit reduces to an R-C one, because the contribution of inductance results, in general, is experimentally negligible. In this case C is the capacity of the model circuit, and the fluctuation of the tension at the ends of the bipole is given by (2.4).

Adair introduced the concept of "electric noise field", in a paper [7], and confirmed it in a subsequent work [8], in order to describe and quantify the observed phenomena. The aim was to take into account that "*in any material the charge density fluctuates thermally according to thermodynamics imperatives generating fluctuating electric fields*" [8]. After having named the mean square value of the noise tension: $\langle V_{kT}{}^2 \rangle$, the main assumption which he applies to the model is that the mean square value of the tension is given, also for biological tissues or cellular membrane, by the expression (2.5).

Therefore, in the case "*of a hypothetical measurement of the voltage across the plates of a parallel plate capacitor where a cube of tissue of length d on a side is held between the plates ... The time-average noise voltage* $\langle V_{kT} \rangle$ *can then be expressed as* $\langle V_{kT}{}^2 \rangle = 4TkR\Delta\nu$, $E_{kT} = \langle V_{kT} \rangle / d$" [8].

For a cubic volume of the tissue, d^3, the time-average noise voltage is $\langle V_{kT} \rangle = (4\rho kT\Delta\nu/d)^{1/2}$, correspondingly $E_{kT} = (4\rho kT\Delta\nu/d^3)^{1/2}$. Assuming $d = 20\,\mu$m, $T = 310°$K, $\rho = 2\,\Omega \cdot$ m, and for a frequency span $\Delta\nu = 100$ Hz, one gets

$$E_{kT} = 20.6 \cdot 10^{-3} = 0.02\,\text{V/m}, \tag{3.1}$$

"which is about 3.000 times larger than the field induced by a 300 V/m external field" [8].

For a cubic section (this choice undergoes a criticism [12]) of a cellular membrane with the same temperature T and values $\rho = 10^6\,\Omega \cdot$ m, $d = 5 \cdot 10^{-9}$ m, "... *the thermal noise electric field is then* $E_{kT} \approx 3.7 \cdot 10^6$ V/m, *which is about* $2 \cdot 10^8$ *times that from a 300 V/m external field*" [8].

For the sake of commodity we report the latter figure as [8]

$$E_{kT} \approx 3.7 \cdot 10^6 \,\mathrm{V/m} \tag{3.2}$$

and observe that following these theoretical estimates, mainly (3.2), the only one conclusion to be drawn is that we have already quoted [8] in our introduction about the masking of all biological effects on the cell level by the thermal noise effects (consequently, the weak ELF fields will not have observable effects on the cells).

Even if one could agree with a model that represents a biological object as a bipole of electronics, one can obtain (3.1) and (3.2) only if $R(\nu) = R$, that is, when a tissue of the human body or the cell membrane could be considered as a *pure resistor*. But there is no reason why this occurrence, denied to the materials of electronics at a point that Johnson and Nyquist were obliged to make use of a bipole (i.e., an impedance) for their experiments, also the conceptual ones, can take place in biology.

Besides, it is easy to check that the value of R used to get (3.2) does not fulfill condition $(2\pi\Delta\nu RC)^2 \ll 1$, that allows the use of (2.5) instead of (2.4).

Thus, since the literature does not provide any reason why a simple electronics model should behave differently if it deals with biology, it remains true that "*the time-average noise voltage* $\langle V_{kT} \rangle$" is 0 for a pure resistor (as we noticed at the beginning of this paper). Consequently, the electric field, that is, the mean value $\langle V_{kT} \rangle$ divided by the length d of the conductor, does the same thing.

The Adair's theoretical estimates were incorporated in the background paper [6]; further, Adair himself has repeated his arguments and kept on his "constraint" in more recent papers [13–15]; and those estimates have been taken into serious account for many years, standing as "Adair's theoretical exposure limits" also in recent papers (see, e.g., [11]).

4. A Result and Some Observations

Replying to an opponent of his theses [12], Adair says: "*my discussion was not original but taken largely from the paper of Weaver and Astumian*" [13], referring to the paper [16]. In that work, Weaver and Astumian deal, among other things, with the problem of calculating the mean square of the noise tension of a cell membrane and the corresponding electric field, using the same data underlying the expression (3.2), but obtaining different figures.

Let us try to give an answer, accepting to represent a cell membrane as an R-C circuit. The relationship between the tension at the terminals of the capacitor, V_C, and its stored energy, E, is given by

$$E = \left(\frac{1}{2}\right) C {V_C}^2. \tag{4.1}$$

For the equipartition theorem, the probability dP of finding the system in the voltage interval $(V, V + dV)$ is proportional to $e^{-E/kT}$,

$$dP = P_0 \cdot \exp\left[-\frac{(1/2)C{V_C}^2}{kT}\right] dV_C. \tag{4.2}$$

After normalizing to fix the value of the constant P_0, one gets

$$\left\langle V_C{}^2 \right\rangle = \left(\frac{C}{2\pi kT}\right)^{1/2} \int V_C \cdot \exp\left[-\frac{(1/2)C\left\langle V_C{}^2\right\rangle}{kT}\right] dV_C \tag{4.3}$$

and performing the integration on $(-\infty, \infty)$ one obtains

$$\left\langle V_C{}^2 \right\rangle = \frac{kT}{C}. \tag{4.4}$$

At the balance, the expression (4.4) leads to a value equal to that given by (2.3).

In order to obtain the value of C for a cellular membrane the thickness of which is d, it is almost natural to think of a capacitor made of two concentric spheres with radii r and $r + d$, respectively, instead of referring to the situation of two parallel planar plates (in this case, in fact, the capacity could depend on the number N of squares, in which the plate can be divided to perform the calculation; and the corresponding expression of C will depend on N^2 [17]). Then, the value of C is

$$C = \frac{\varepsilon_r \varepsilon_0 4\pi r^2}{d}, \tag{4.5}$$

where ε_0 and ε_r are, respectively, the values of dielectric constant in vacuum and in a specified matter. From (4.4) it follows for the electric field

$$E = \frac{\left(\left\langle V_C{}^2 \right\rangle\right)^{1/2}}{d} = \frac{(kT/C)^{1/2}}{d}; \tag{4.6}$$

if the same values as in the expression (3.2) are assumed for the parameters, that is $\varepsilon_r = 2.5$, a value drawn from literature, and $r = 10^{-5}$ m, $d = 5 \cdot 10^{-9}$ m, $T = 310°$K, one obtains

$$E \approx 5.5 \cdot 10^3 \text{ V/m}. \tag{4.7}$$

The value given in (4.7) is implicitly referred to the whole spectrum of frequencies; if only ELF effects are investigated, then one has to calculate the finite integral of the expression (2.4) in the interval $[\nu_1 = 0\,\text{Hz}, \nu_2 = 100\,\text{Hz}]$. Performing the integration on the interval $[\nu_1, \nu_2]$ one obtains

$$\begin{aligned}\left\langle V_n{}^2 \right\rangle_{\text{PART}} &= 4kT \int_{\nu 1}^{\nu 2} \left\{ \frac{R}{\left[1 + (2\pi \nu RC)^2\right]} \right\} d\nu \\ &= \left(\frac{kT}{C}\right)\left(\frac{2}{\pi}\right) \text{arctg} \left\{ \frac{[2\pi(\nu_2 - \nu_1)RC]}{\left[1 + (2\pi RC\sqrt{\nu_1 \nu_2})^2\right]} \right\}.\end{aligned} \tag{4.8}$$

If we take for resistivity the same value as in [8, 16], $\rho = 10^6\,\Omega \cdot \mathrm{m}$, in the assumed frequency band the expression (4.8) gives

$$E \approx 260\,\mathrm{V/m}; \tag{4.9}$$

that is a value, for the electric noise field, 10^{-4} times minor than (3.2) (and $10^{-4} \div 10^{-5}$ times minor of the cell membrane electric field—the measures give about $14 \cdot 10^6\,\mathrm{V/m}$—as it is reasonable to expect).

In conclusion, it is not arbitrary to argue that if a cell membrane is represented not as a pure resistor but as a conductor, then the value for the noise electric field is much lower than that given by (3.2); correspondingly, the "Adair's constraint" results significantly weakened.

Furthermore, one could reasonably expect that also the mean value of the noise tension at the ends of a "bipole" be zero, just as that at the ends of a resistor. Thus, the figures (4.7), (4.9) would constitute more an estimation to determine possible peak values than theoretical data to compare with the experimental ones.

5. Some Examples about the Johnson's Formula in Biology

We have just observed in the previous formulae that it is crucial to use a correct value of R, thus one could think that, if this request $(2\pi\Delta\nu RC)^2 \ll 1$ is satisfied, not a relevant gap between the estimates from (2.5) and (4.8) can occur not only in the ELF region but also in other bands of the spectrum of frequencies. That is, our criticism about the representation of a conductor as a pure resistor could be right in principle but poor of effects on the results of experiments. On the contrary, it is not true when one deals with human cell membranes, *the parameters of which have values around those used by Adair and Weaver*, as it is immediate to check. In fact the recalled condition is not well fulfilled already in the region of tens kHz, where the Johnson's formula (2.5) gives a significant deviation from the correct value of the square mean tension as given by (4.8), as we will see just after; and for larger frequencies obviously only the formula (4.8) can be applied.

Let us consider a high power transmitter operating in the band (14–30) kHz; a VLF antenna such that: "…induces currents and fields in people living in the urban area within 2 km of the antenna that are greater than those in people living very close to high voltage power lines…" [18]. It follows that an intensity of electromagnetic field equal to that of power lines is reached at much shorter distance, but the effects will be the same in correspondence to the same value of the irradiated field. In any case, these different behaviours, significant for the health impact, do not affect the calculus of $\langle {V_n}^2 \rangle_{\mathrm{PART}}$: it depends only on the electrical properties of the cellular membrane—we'll keep the values of the parameters of the previous example—and on the frequency of the emitted signal. We take $\Delta\nu$ equal, as usual, to one tenth of the range of operating band. Thus, just because the condition $(2\pi\Delta\nu RC)^2 \ll 1$ is not well satisfied, in fact $(2\pi\Delta\nu RC)^2 = 0.045$, from (2.5) one can obtain the hypothetical estimate

$$\left\langle {V_n}^2 \right\rangle_{\mathrm{PART}}^{(2.5)} = 109 \cdot 10^{-12}\,\mathrm{V}, \tag{5.1}$$

while the calculation performed by (4.8) gives

$$\left\langle V_n^2 \right\rangle_{\text{PART}}^{(4.8)} = 21 \cdot 10^{-12}\,\text{V}. \tag{5.2}$$

Consequently, the corresponding values of electric noise field differ about 130%:

$$E^{(2.5)} = 2087 \cdot 10^3\,\text{V/m}, \qquad E^{(4.8)} = 912 \cdot 10^3\,\text{V/m}; \tag{5.3}$$

obviously, only the minor one being correct.

It is easy to verify that for such a cell membrane, the electric noise field decreases, as one could expect, when frequency grows. Thus in the region $(0.5 \div 1.5)$ MHz, frequently used in laboratory experiments, one can find $E^{(4.8)} \approx 137\,\text{V/m}$; in the region $(50 \div 60)$ MHz, where are operating a lot of broadcasting and telecasting devices, it will be $E^{(4.8)} \approx 6.8\,\text{V/m}$, and for the frequencies of mobile phone or radar bridge $(2 \div 3)$ GHz, $E^{(4.8)} \approx 1.5\,\text{V/m}$.

For nonhuman cell membranes, it can happen that also at high frequencies the parameters satisfy $(2\pi\Delta\nu RC)^2 \ll 1$, so that the deviation will be negligible and it will be easier to compute the square mean tension by the Johnson's formula.

In an experiment realized indeed many years ago to test the effects of electromagnetic fields on the embryonic development of eggs of Salmo lacustris [19] the authors formulated the hypothesis that the deformations revealed on the irradiated eggs and on the embryonic development of Salmo lacustris with respect to those not exposed should be caused by the magnetic field. In fact, they took care of measuring a constant temperature during the time of exposition of the eggs, so that an alteration of the embryonic development due to the heat released by the irradiating field could be excluded.

Let us look at the figures that one could have obtained for the thermal noise tension in that experiment. The eggs of Salmo lacustris were exposed to an oscillating magnetic field in the range (450–1350) kHz and the values reported in literature for those eggs give $r = 2.5 \cdot 10^{-3}\,\text{m}$ [20, 21], $d = 50 \cdot 10^{-6}\,\text{m}$ [22, 23] and $\rho = 5 \cdot 10^3\,\Omega \cdot \text{cm}^2$ [21]. Now, the condition $(2\pi\Delta\nu RC)^2 \ll 1$ is fulfilled for $\Delta\nu = 100\,\text{kHz}$, thus the approximated formula (2.5) can be applied obtaining

$$\begin{aligned} \left\langle V_n^2 \right\rangle_{\text{PART}}^{(2.5)} &= 2.18 \cdot 10^{-12}\,\text{V}, \\ E^{(2.5)} &= 29.5\,\text{mV/m}; \end{aligned} \tag{5.4}$$

out of curiosity, the values given by the formula (4.8) are, respectively,

$$\left\langle V_n^2 \right\rangle_{\text{PART}}^{(4.8)} = 2.062 \cdot 10^{-12}\,\text{V}, \qquad E^{(4.8)} = 28.7\,\text{mV/m}. \tag{5.5}$$

This occurrence is not amazing why the variations of resistance and capacity, in function of frequency, of cell membranes have been investigated for a long time [24] and it is well known that for the cell membrane of an egg of Salmo lacustris the resistance is much lower than that of a spherical human one; and its conductance increases rapidly at high frequencies [21]. Therefore, for a nonhuman cell this example shows that when it is exposed to high

frequencies, the electric noise field, if does exist, is not a barrier to mask such effects at the cell level of an external field; since, as a matter of fact, the alterations of eggs and their embryonic development were experimentally revealed not as caused by the release of thermal energy.

6. Concluding Remarks

One of the key problems in bioelectromagnetism is to explain the mechanism of the influence of weak electromagnetic fields on biological objects; it remains unclear, in spite of numerous experimental data. In particular, it is not clear how low frequencies or static fields, magnetic or electric, can lead to the "resonance" of biochemical reactions, even when the energy of such fields is very small in comparison with the energy kT of the process. The lack of a theoretical explanation, that is satisfying or shared among researchers, is now called "kT problem" or "kT paradox" [25].

On this subject, after a very long-time referees' action Michail Zhadin reported the alteration of electric properties of a nonbiological system, made up by an aqueous diluted solution of amino acids (glutamic acid) [26], in correspondence of the frequencies of an ion cyclotron resonance.

A direct current voltage of 80 mV was applied to the solution contained in a electrolytic cell, near to the value of the cell membrane potential. The solution was exposed to the combined action of two parallel magnetic fields, one weak and static (B_0 = 20–40 μT = micro Tesla), the other extremely weak and alternating (B_c = 10–80 nT = nano Tesla), both applied orthogonally to the electrolytic current direction.

A very narrow intensity peak in the electric current can be measured, when the frequency ν of the alternating magnetic field B_c matches the ion cyclotron resonance of the ionized solution; this frequency ν is given by the well-known formula

$$\nu = \frac{qB_0}{2\pi m}, \tag{6.1}$$

where m, q are, respectively, the mass and the charge of the electrolytic ion. The frequency windows found by Zhadin were at 4 Hz for $B_0 = 40\,\mu$T and B_c = 10, 20, 30 nT; in the interval [2, 4] Hz, spaced by 0.5 Hz, when B_0 = 20, 25, 30, 40 μT and B_c = 25 nT.

Many authors refer to this result as the *Zhadin effect*, that has been successfully replicated in Italy [27–29] and in Germany [30]. Several attempts to give a theoretical explanation of the physical mechanism underlying that effect have been made [27, 31] in the framework of the quantum electrodynamics of condensed matter proposed by [32], also by Zhadin himself [33], after a previous analysis performed in terms of the semiclassical resonance theory [34].

The Zhadin effect, born and actually studied in another context, is, in the limits of that experiment, a positive answer to our opening question: an extremely weak magnetic field—the very low intensity signal—can overcome, in correspondence of a frequency window, the noise tension—the disturbance—the energy density of which is very lager than that of the field. This eventuality suggests that a similar effect could take place also when biological cells are irradiated by very weak electromagnetic fields; that is, the masking effect on the cell level by thermal noise could have a break down.

This suggestion could be the reason why Adair has criticized the previously quoted paper [34], going on with his argument: the equivalent electric field acting on the ion thermal

motion is many times minor than the electric noise field [35]. But, in his indications does not appear the estimate (3.2). An analogue reasoning is developed in [25], whose criticism versus the theoretical analysis performed in [34] is focused on the energy of a particle motion at the cyclotron resonance frequency, directly compared with the thermal agitation energy kT.

Thus, the problem is leaded to the properness of the models that try to explain the Zhadin effect. In fact, Zhadin asserts about the attempts to interpret his experiment in terms of resonance: "*... Unfortunately, for free ions such sort of effects are absolutely impossible because dimensions of an ion rotation radius should be measured by meters at room temperature and at very low static magnetic fields used in all the before experiments. Even for bound ions these effects should be absolutely impossible for the positions of classic physics because of rather high viscosity of biological liquid media...*" [34]. But on another side, the recalled attempts to bring that experiment within the conceptual framework of the theory formulated by Preparata have not yet met a general sharing among the insiders.

In few words, a very complex and up to now open problem.

References

[1] J. B. Johnson, "Thermal agitation of electricity in conductors," *Nature*, vol. 119, no. 2984, pp. 50–51, 1927.

[2] J. B. Johnson, "Thermal agitation of electricity in conductors," *Physical Review*, vol. 32, no. 1, pp. 97–109, 1928.

[3] H. Nyquist, "Thermal agitation of electric charge in conductors," *Physical Review*, vol. 32, no. 1, pp. 110–113, 1928.

[4] APS, "Power-Line Fields and Public Health. Statement of the Council of the American Physical Society," 1995, http://www.aip.org/fyi/1995/fyi95.069.htm.

[5] APS, "National Policy. Electric and Magnetic Fields and Public Health," 2005, http://www.aps.org/policy/statements/05_3.cfm.

[6] D. Hafemeister, "Background Paper on 'Power-Line Fields and Public Health'," 1996, http://www.calpoly.edu/~dhafemei/background2.html.

[7] R. K. Adair, "Are biological effects of weak ELF fields possible?" in *Proceedings of the 12th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 12, pp. 1559–1561, November 1990.

[8] R. K. Adair, "Constraints on biological effects of weak extremely-low-frequency electromagnetic fields," *Physical Review A*, vol. 43, no. 2, pp. 1039–1048, 1991.

[9] T. S. Tenforde and W. T. Kaune, "Interaction of extremely low frequency electric and magnetic fields with humans," *Health Physics*, vol. 53, no. 6, pp. 585–606, 1987.

[10] T. S. Tenforde, "Biological interactions of extremely-low-frequency electric and magnetic fields," *Bioelectrochemistry and Bioenergetics*, vol. 25, no. 1, pp. 1–17, 1991.

[11] C. D. Abeyrathne, P. M. Farrell, and M. N. Halgamuge, "Analysis of biological effects and limits of exposure to weak magnetic fields," in *Proceedings of the 5th International Conference on Information and Automation for Sustainability (ICIAfS '10)*, pp. 415–420, Colombo, Sri Lanka, December 2010.

[12] R. W. P. King, "The interaction of power-line electromagnetic fields with the human body," *IEEE Engineering in Medicine and Biology Magazine*, vol. 17, no. 6, pp. 67–73, 1998.

[13] R. K. Adair, "Comments regarding the Article of R.W.P. King," *IEEE Engineering in Medicine and Biology*, vol. 17, pp. 73–75, 1998.

[14] R. K. Adair, "Static and low-frequency magnetic field effects: health risks and therapies," *Reports on Progress in Physics*, vol. 63, no. 3, pp. 415–454, 2000.

[15] R. K. Adair, "Noise and stochastic resonance in voltage-gated ion channels," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, pp. 12099–12104, 2003.

[16] J. C. Weaver and R. D. Astumian, "The response of living cells to very weak electric fields: the thermal noise limit," *Science*, vol. 247, no. 4941, pp. 459–462, 1990.

[17] R. W. P. King, "King's rebuttal to Adair's comments," *IEEE Engineering in Medicine and Biology*, vol. 17, pp. 76–78, 1998.

[18] R. W. P. King and C. W. Harrison, "Electromagnetic field in human body due to VLF transmitter," in *Proceedings of the IEEE 21st Annual Northeast Bioengineering Conference*, pp. 121–123, May 1995.

[19] L. Gianferrari and E. Pugno Vanoni, "Sull'azione di campi elettrici ad alta frequenza sullo sviluppo embrionale: I. Esperienze su Salmo Lacustris," Rendiconti R. Accademia Lincei, Classe Scienze Fisiche Matematiche e Naturali Serie VIII (S.5, I Sem.): 576–578, 1923.
[20] R. Bartel, B. Falowska, K. Bieniarz, and P. Epler, "Dependence of egg diameter on the size and age of cultivated female lake trout," *Archives of Polish Fisheries*, vol. 13, no. 1, pp. 121–126, 2005.
[21] L. Rothschild, "The theory of alternating current measurements in biology and its application to the investigation of the biophysical properties of the trout egg," *Journal of Experimental Medicine*, vol. 23, pp. 77–99, 1946.
[22] A. I. ZOTIN, "The mechanism of hardening of the salmonid egg membrane after fertilization or spontaneous activation," *Journal of Embryology and Experimental Morphology*, vol. 6, no. 4, pp. 546–568, 1958.
[23] C. M. Stehr and J. W. Hawkes, "The comparative ultrastructure of the egg membrane and associated pore structures in the Starry Flounder, Platichthys stellatus (Pallas), and pink salmon, Oncorhynchus gorbuscha (Walbaum)," *Cell and Tissue Research*, vol. 202, no. 3, pp. 347–356, 1979.
[24] K. S. Cole, "Electric phase angle of cell membranes," *The Journal of General Physiology*, vol. 15, no. 6, pp. 641–649, 1932.
[25] V. N. Binhi and A. B. Rubin, "Magnetobiology: the kT paradox and possible solutions," *Electromagnetic Biology and Medicine*, vol. 26, no. 1, pp. 45–62, 2007.
[26] M. N. Zhadin, V. V. Novikov, F. S. Barnes, and N. F. Pergola, "Combined action of static and alternating magnetic fields on ionic current in aqueous glutamic acid solution," *Bioelectromagnetics*, vol. 19, no. 1, pp. 41–45, 1998.
[27] E. Del Gludice, M. Fleischmann, G. Preparata, and G. Talpo, "On the "unreasonable" effects of ELF magnetic fields on upon a system of ions," *Bioelectromagnetics*, vol. 23, no. 7, pp. 522–530, 2002.
[28] N. Comisso, E. Del Giudice, A. De Ninno et al., "Dynamics of the ion cyclotron resonance effect on amino acids adsorbed at the interfaces," *Bioelectromagnetics*, vol. 27, no. 1, pp. 16–25, 2006.
[29] L. Giuliani, S. Grimaldi, A. Lisi, E. D'Emilia, N. Bobkova, and M. Zhadin, "Action of combined magnetic fields on aqueous solution of glutamic acid: the further development of investigations," *BioMagnetic Research and Technology*, vol. 6, article 1, 2008.
[30] A. Pazur, "Characterisation of weak magnetic field effects in an aqueous glutamic acid solution by nonlinear dielectric spectroscopy and voltammetry," *BioMagnetic Research and Technology*, vol. 2, article 8, 2004.
[31] E. Del Giudice, G. Preparata, and M. Fleischmann, "QED coherence and electrolyte solutions," *Journal of Electroanalytical Chemistry*, vol. 482, no. 2, pp. 110–116, 2000.
[32] G. Preparata, "QED Coherence in Matter," World Scientific, 1995.
[33] M. N. Zhadin, "On mechanism of combined extremely weak magnetic field action on aqueous solution of amino acid," in *Non-Thermal Effects and Mechanisms between Electromagnetic Fields and Living Matter*, L. Giuliani and M. Soffritti, Eds., vol. 5 of *European Journal of Oncology*, ICEMS Monograph, 2010.
[34] M. Zhadin and F. Barnes, "Frequency and amplitude windows in the combined action of DC and low frequency AC magnetic fields on ion thermal motion in a macromolecule: theoretical analysis," *Bioelectromagnetics*, vol. 26, no. 4, pp. 323–330, 2005.
[35] R. K. Adair, "Comment: analyses of models of ion actions under the combined action of AC and DC magnetic fields," *Bioelectromagnetics*, vol. 27, no. 4, pp. 332–334, 2006.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 329489, 18 pages
doi:10.1155/2012/329489

Research Article

Bayesian Analysis of the Survival Function and Failure Rate of Weibull Distribution with Censored Data

Chris Bambey Guure[1] and Noor Akma Ibrahim[1,2]

[1] *Institute for Mathematical Research, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*
[2] *Department of Mathematics, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*

Correspondence should be addressed to Chris Bambey Guure, bam.bey1@hotmail.com

Received 17 August 2012; Accepted 18 September 2012

Academic Editor: Carlo Cattani

The survival function of the Weibull distribution determines the probability that a unit or an individual will survive beyond a certain specified time while the failure rate is the rate at which a randomly selected individual known to be alive at time $(t-1)$ will die at time (t). The classical approach for estimating the survival function and the failure rate is the maximum likelihood method. In this study, we strive to determine the best method, by comparing the classical maximum likelihood against the Bayesian estimators using an informative prior and a proposed data-dependent prior known as generalised noninformative prior. The Bayesian estimation is considered under three loss functions. Due to the complexity in dealing with the integrals using the Bayesian estimator, Lindley's approximation procedure is employed to reduce the ratio of the integrals. For the purpose of comparison, the mean squared error (MSE) and the absolute bias are obtained. This study is conducted via simulation by utilising different sample sizes. We observed from the study that the generalised prior we assumed performed better than the others under linear exponential loss function with respect to MSE and under general entropy loss function with respect to absolute bias.

1. Introduction

As a result of the adaptability in fitting time-to-failure of a very widespread multiplicity to multifaceted mechanisms, the Weibull distribution has assumed the centre stage especially in the field of life-testing and reliability/survival analysis. It has shown to be very useful for modelling and analysing life time data in medical, biological and engineering sciences, Lawless [1]. Much of the attractiveness of the Weibull distribution is due to the wide variety of shapes it can assume by altering its parameters.

Censoring is a feature that is recurrent in lifetime and reliability data analysis, it occurs when exact lifetimes or run-outs can only be collected for a portion of the inspection units.

According to Rinne [2], "A data sample is said to be censored when, either by accident or design, the value of the variables under investigation is unobserved for some of the items in the sample."

Maximum likelihood estimator (MLE) is quiet efficient and very popular both in literature and practice. Bayesian approach has been employed for estimating parameters. Some research have been done to compare MLE and that of the Bayesian approach in estimating the survival function and the parameters of the Weibull distribution. Sinha [3] determined the Bayes estimates of the reliability function and the hazard rate of the Weibull failure time distribution by employing squared error loss function, Abdel-Wahid and Winterbottom [4] studied the approximate Bayesian estimates for the Weibull reliability function and hazard rate from censored data by employing a new method that has the potential of reducing the number of terms in Lindley procedure, and Al Omari and Ibrahim [5] conducted a study on Bayesian survival estimator for Weibull distribution with censored data using squared error loss function with Jeffreys prior amongst others.

Guure et al. [6] applied Bayesian estimation, for the two-parameter Weibull distribution using extension of Jeffreys' prior information with three loss functions, Syuan-Rong and Shuo-Jye [7] considered Bayesian estimation and prediction for Weibull model with progressive censoring. Other recent papers employing different models can be seen in Cattani and Ciancio [8], Wang et al. [9], Cattani [10, 11]. Similar work can be seen in Gupta and Kundu [12] Zellner [13], Al-Aboud [14], Al-Athari [15], Cattani et al. [16] Pandey et al. [17], and a work on generalized exponential distribution: Bayesian estimations, Kundu and Gupta [18] which is somehow similar to the Weibull distribution.

The aim of this paper is twofold. First is to consider the maximum likelihood estimator of the survival function and the failure rate. In order to obtain the estimate of the survival function and the failure rate, we first need the MLE of the Weibull two parameters under consideration here. It is observed that the MLE cannot be obtained in analytical form, we therefore assumed the Newton-Raphson method to compute the MLE, and the method works quite well.

The second aim of this paper is to consider the Bayesian approach for the survival function and the failure rate. It is remarkable that most of the Bayesian inference procedures have been developed with the usual squared error loss function, which is symmetrical and associates equal importance to the losses due to overestimation and underestimation of equal magnitude according to Vasile et al. [19]. However, such applications may be restrictive in most situations of practical importance. For example, in the estimation of the failure rate, an overestimation is usually much more serious than an underestimation. In this case, the use of a symmetric loss function might be inappropriate as stated by Basu and Ebrahimi [20]. In this paper, we obtain the Bayes estimates under the LINEX loss function, general entropy loss function, and squared error loss function using Lindley's approximation technique via simulation study with informative prior and a generalised noninformative prior.

The rest of the paper is arranged as follows: Section 2 contains the derivative of the parameters based on which the survival function and the failure rate are determined under maximum likelihood estimator, Section 3 is the Bayesian estimation approach. This is followed by simulation study in Section 4. Results and discussion are in Sections 5 and 6 is the conclusion.

2. Maximum Likelihood Estimation

Let $(t_i, \ldots, t_n)$ be the set of n random lifetimes with respect to the Weibull distribution with α and β as the parameters, where α is the scale parameter and β the shape parameter.

The probability density function (pdf), the cumulative distribution function (cdf), and the survival function $S(\cdot)$ of Weibull are given, respectively, as

$$\begin{aligned} f(t;\alpha,\beta) &= \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \quad \forall t > 0, \\ F(t;\alpha,\beta) &= 1-\exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right], \\ S(t;\alpha,\beta) &= \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]. \end{aligned} \tag{2.1}$$

The likelihood function is

$$L(t_i;\alpha,\beta,\delta) = \prod_{i=1}^{n}\left[f(t_i,\alpha,\beta)\right]^{\delta_i}\left[s(t_i,\alpha,\beta)\right]^{1-\delta_i}, \tag{2.2}$$

where $\delta_i = 1$ is the failure time and $\delta_i = 0$ is for censored observations. $S(\cdot)$ is the survival function.

The log-likelihood function of (2.2) is

$$\ell = \sum_{i=1}^{n}\left[\delta_i\left[\ln(\beta)-\beta\ln(\alpha)+(\beta-1)\ln(t_i)\right]-\left(\frac{t_i}{\alpha}\right)^{\beta}\right]. \tag{2.3}$$

We differentiate (2.3) with respect to the unknown parameters and equal the resulting equation to zero as follows:

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= -\frac{\beta\sum_{i=1}^{n}\delta_i}{\alpha}+\frac{\beta\sum_{i=1}^{n}(t_i/\alpha)^{\beta}}{\alpha} = 0, \\ \frac{\partial \ell}{\partial \beta} &= \frac{\sum_{i=1}^{n}\delta_i}{\beta}-\sum_{i=1}^{n}\delta_i\ln(\alpha)+\sum_{i=1}^{n}\delta_i\ln(t_i)-\sum_{i=1}^{n}\left(\frac{t_i}{\alpha}\right)^{\beta}\ln\left(\frac{t_i}{\alpha}\right) = 0. \end{aligned} \tag{2.4}$$

The maximum likelihood estimator of α is

$$\widehat{\alpha}_{ML} = \left(\frac{\sum_{i=1}^{n}(t_i)^{\widehat{\beta}}}{\sum_{i=1}^{n}\delta_i}\right)^{1/\widehat{\beta}}. \tag{2.5}$$

The shape parameter $\widehat{\beta}$ is obtained by the method of Newton-Raphson since it cannot be solved analytically.

The estimates of the survival function and the failure rate of the Weibull distribution under MLE are

$$\widehat{S}_{\mathrm{ML}}(t) = \exp\left[-\left(\frac{t}{\widehat{\alpha}_{\mathrm{ML}}}\right)^{\widehat{\beta}_{\mathrm{ML}}}\right],$$
$$\widehat{F}_{\mathrm{ML}}(t) = \frac{\widehat{\beta}_{\mathrm{ML}}}{\widehat{\alpha}_{\mathrm{ML}}}\left(\frac{t}{\widehat{\alpha}_{\mathrm{ML}}}\right)^{\widehat{\beta}_{\mathrm{ML}}-1}. \tag{2.6}$$

3. Bayesian Inference

In the field of statistics, Bayesian inference is viewed as a method that employs the Bayes' rule in other to update the probability estimate of a hypothesis taking into consideration new evidence as they become available. Bayesian updating is one of the essential techniques used in modern statistics, more importantly in mathematical statistics. Bayesian updating is particularly important in analysing sequence of data. Bayesian inference is applicable in many fields today, for example, engineering, medicine, accounting, and others. Bayes makes use of our prior beliefs of the parameters which is known as the prior distribution. It is the distribution of the parameters before any data is observed and is given as $p(\theta)$. It also takes into consideration the observed data which is viewed as the likelihood function and given as $L(t \mid \theta)$.

Bayesian inference is based on the posterior distribution which is simply the ratio of the joint density function to the marginal distribution.

The posterior distribution is

$$\pi^*(\theta \mid t) \propto \frac{p(\theta)L(t \mid \theta)}{\int_0^\infty p(\theta)L(t \mid \theta)d\theta}, \tag{3.1}$$

where t is the data and θ is the parameter of interest.

The Bayes estimator is considered under three loss functions which is also indisputable in Bayesian estimation. They are asymmetric (LINEX and general entropy) loss functions and symmetric (squared error) loss function.

Prior distribution of the parameters need to be assumed for the Bayesian inference. As discussed by Berger and Sun [21] and subsequently by Banerjee and Kundu [22], we let α take on a Gamma (a, b) prior with $a > 0$ and $b > 0$. We assume that the prior of β is independent of the prior of α and is in the neighbourhood of $(0, \infty)$. Let $v(\beta)$ represent the prior of β and $v_1(\alpha)$ for α, where

$$\text{Gamma}\ (\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\beta t). \tag{3.2}$$

Let the likelihood equation which is $L(t_i \mid \alpha, \beta)$ be the same as (2.2).

The joint posterior density function of (α, β) is given by

$$\pi(\alpha, \beta \mid t) \propto L(t \mid \alpha, \beta) v(\beta) v_1(\alpha). \tag{3.3}$$

The posterior probability density function of α and β given the data $(t_1, t_2, \ldots, t_n)$ is obtained by dividing the joint posterior density function over the marginal distribution function as

$$\pi^*(\alpha, \beta \mid t) = \frac{L(t \mid \alpha, \beta) v(\beta) v_1(\alpha)}{\int\int_0^\infty L(t \mid \alpha, \beta) v(\beta) v_1(\alpha) d\alpha \, d\beta}. \tag{3.4}$$

3.1. Generalised Noninformative Prior

Our proposition of the generalised noninformative prior for the parameters α and β is given such that

$$v_2(\alpha, \beta) = \left(\frac{1}{\widehat{\alpha}\widehat{\beta}}\right)^a, \quad a > 0. \tag{3.5}$$

We are assuming there is no or little knowledge on the parameters being estimated, where $\widehat{\alpha}$ and $\widehat{\beta}$ are the estimates from maximum likelihood with respect to the available data obtained by the researcher and a is a constant that can assume any value in order to minimise the prior effect on the posterior distribution. We refer to this approach as the data-dependent prior and it is more or less an empirical or objective Bayes prior. This is an interesting new developing theory of objective priors and, while data dependent, it does not involve an inappropriate double use of the data, Berger [23], unless the sample size is fairly small. The generalised noninformative prior must not be misconstrued to imply a joint prior for the two parameters Weibull distribution. In other words, the scale and shape parameters are independent a priori.

The likelihood function from (2.2) is

$$L(\mathrm{t_i} \mid \alpha, \beta) = \prod_{i=1}^{n} \left\{ \left(\frac{\beta}{\alpha}\right)\left(\frac{t_i}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t_i}{\alpha}\right)^{\beta}\right] \right\}^{\delta_i} \left\{ \exp\left[-\left(\frac{t_i}{\alpha}\right)^{\beta}\right] \right\}^{1-\delta_i}. \tag{3.6}$$

With Bayes theorem the joint posterior distribution of the parameters α and β is

$$\begin{aligned} \pi(\alpha, \beta \mid t) &\propto L(t \mid \alpha, \beta) v_2(\alpha, \beta), \\ &= k v_2(\alpha, \beta) \prod_{i=1}^{n} \left\{ \left(\frac{\beta}{\alpha}\right)\left(\frac{t_i}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t_i}{\alpha}\right)^{\beta}\right] \right\}^{\delta_i} \left\{ \exp\left[-\left(\frac{t_i}{\alpha}\right)^{\beta}\right] \right\}^{1-\delta_i}, \end{aligned} \tag{3.7}$$

where k is the normalizing constant that makes π a proper pdf.

The posterior density function is obtained by using (3.4).

3.2. Linear Exponential Loss Function

This loss function according to Soliman et al. [24] rises approximately exponentially on one side of zero and approximately linearly on the other side.

The posterior expectation of the LINEX loss function according to Zellner [13] is

$$E_\theta\left[L\left(\widehat{\theta}-\theta\right)\right] \propto \exp\left(c\widehat{\theta}\right)E_\theta\left[\exp(-c\theta)\right] - c\left[\widehat{\theta}-E_\theta(\theta)\right] - 1, \tag{3.8}$$

with $E_\theta(\cdot)$ denoting the posterior expectation with respect to the posterior density of θ. Therefore, the Bayes estimator of θ, which is denoted by $\widehat{\theta}_{\mathrm{BL}}$ under LINEX loss function is the value of $\widehat{\theta}$ which minimizes (3.8) and is

$$\widehat{\theta}_{\mathrm{BL}} = -\frac{1}{c}\ln\{E_\theta\left[\exp(-c\theta)\right]\}, \tag{3.9}$$

provided that $E_\theta[\exp(-c\theta)]$ exist and is finite.

The posterior density of the survival function and the failure rate under this loss function are given as

$$\begin{aligned}
S(t)_{\mathrm{BL}} &= E\left\{\exp\left[-c\exp\left[-\left(\frac{t_i}{\alpha}\right)^{\beta}\right]\right] \mid t\right\} \\
&= \frac{\int_0^\infty\int_0^\infty v(\beta)v_1(\alpha)\exp\left[-c\exp\left[-(t_i/\alpha)^{\beta}\right]\right]L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}{\int_0^\infty\int_0^\infty v(\beta)v_1(\alpha)L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}, \\
F(t)_{\mathrm{BL}} &= E\left\{\left[-c\exp\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}\right] \mid t\right\} \\
&= \frac{\int_0^\infty\int_0^\infty v(\beta)v_1(\alpha)\exp\left[-c(\beta/\alpha)(t/\alpha)^{\beta-1}\right]L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}{\int_0^\infty\int_0^\infty v(\beta)v_1(\alpha)L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}.
\end{aligned} \tag{3.10}$$

It can be observed that (3.10) contain ratio of integrals which cannot be obtained analytically and as a result we make use of Lindley approximation procedure to evaluate the integrals involved.

3.3. Lindley Approximation

A prior of β need to be specified here so as to calculate the approximate Bayes estimates of α and β. Having specified a prior for α as Gamma (a,b), it is similarly assumed that $v(\beta)$ also takes on a Gamma (c,d) prior.

Lindley [25] proposed a ratio of integral of the form

$$\frac{\int \omega(\theta)\exp\{L(\theta)\}d\theta}{\int v(\theta)\exp\{L(\theta)\}d\theta}, \tag{3.11}$$

where $L(\theta)$ is the log-likelihood and $\omega(\theta),\ v(\theta)$ are arbitrary functions of θ. Assuming that $v(\theta)$ is the prior distribution for θ and $\omega(\theta) = u(\theta) \cdot v(\theta)$ with $u(\theta)$ being some function of interest.

The posterior expectation according to Sinha [3] is

$$E\{u(\theta) \mid t\} = \frac{\int v(\theta) \exp\{L(\theta) + \rho(\theta)\} d(\theta)}{\int \exp\{L(\theta) + \rho(\theta)\} d(\theta)}, \tag{3.12}$$

where $\rho(\theta) = \log\{v(\theta)\}$.

This can be approximated asymptotically by

$$E\{u(\theta) \mid x\} = \left[u + \frac{1}{2} \sum_i \sum_j (u_{ij} + 2u_i \cdot \rho_j) \cdot \sigma_{ij} + \frac{1}{2} \sum_i \sum_j \sum_k \sum_l L_{ijk} \cdot \sigma_{ij} \cdot \sigma_{kl} \cdot u_l \right], \tag{3.13}$$

where $i, j, k, l = 1, 2, \ldots, n;\ \theta = (\theta_1, \theta_2, \ldots, \theta_m)$.

Taking the two parameters into consideration, (3.13) reduces to

$$\widehat{\theta} = u + \frac{1}{2}[(u_{11}\sigma_{11}) + (u_{22}\sigma_{22})] + u_1\rho_1\sigma_{11} + u_2\rho_2\sigma_{22} + \frac{1}{2}\left[\left(L_{30}u_1\sigma_{11}^2\right) + \left(L_{03}u_2\sigma_{22}^2\right)\right], \tag{3.14}$$

where L is the log-likelihood equation in (2.3).

Taking the survival function, where

$$\begin{aligned} u &= \exp\left\{-c\exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]\right\}, \qquad q = \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right], \\ u_1 &= \frac{\partial u}{\partial \alpha} = c\left(\frac{\beta}{\alpha}\right)\left(\frac{-t}{\alpha}\right)^{\beta} qu, \end{aligned} \tag{3.15}$$

$$\begin{aligned} u_{11} = \frac{\partial^2 u}{\partial \alpha^2} &= -c\left(\frac{\beta^2}{\alpha^2}\right)\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] qu - c\left(\frac{\beta}{\alpha^2}\right)\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] qu \\ &\quad - c\left(\frac{\beta^2}{\alpha^2}\right)\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]^2 qu + c^2\left(\frac{\beta^2}{\alpha^2}\right)\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]^2 q^2 u. \end{aligned} \tag{3.16}$$

In a similar approach $u_2 = \partial u/\partial\beta$ and $u_{22} = \partial^2 u/\partial\beta^2$ can be obtained.

For the failure rate,

$$
\begin{aligned}
u &= \exp\left[-c\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}\right], \qquad p = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}, \\
u_2 &= \frac{\partial u}{\partial \beta} = \left[-cp\,\ln\left(\frac{t}{\alpha}\right) - \left(\frac{c}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}\right]u, \\
u_{22} &= \frac{\partial^2 u}{\partial \beta^2} = \left[-cp\ln^2\left(\frac{t}{\alpha}\right) - 2\left(\frac{c}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}\ln\left(\frac{t}{\alpha}\right)\right]u \\
&\quad + \left[-cp\ln\left(\frac{t}{\alpha}\right) - \left(\frac{c}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}\right]^2 u,
\end{aligned}
\tag{3.17}
$$

$u_1 = \partial u/\partial \alpha$ and $u_{11} = \partial^2 u/\partial \alpha^2$ follow in like manner.

Let

$$
\begin{aligned}
\rho &= \ln v_1(\alpha) + \ln v(\beta), \\
\rho_1 &= \frac{c-1}{\alpha} - d, \qquad \rho_2 = \frac{a-1}{\beta} - b, \\
L_{20} &= \frac{\partial^2 L}{\partial \alpha^2} = \frac{\beta \sum_{i=1}^n \delta_i}{\alpha^2} - \frac{\beta^2 \sum_{i=1}^n (t_i/\alpha)^\beta}{\alpha^2} - \frac{\beta \sum_{i=1}^n (t_i/\alpha)^\beta}{\alpha^2}, \\
\sigma_{11} &= (-L_{20})^{-1}, \\
L_{30} &= -\frac{2\beta \sum_{i=1}^n \delta_i}{\alpha^3} + \frac{\beta^3 \sum_{i=1}^n (t_i/\alpha)^\beta}{\alpha^3} + \frac{3\beta^2 \sum_{i=1}^n (t_i/\alpha)^\beta}{\alpha^3} + \frac{2\beta \sum_{i=1}^n (t_i/\alpha)^\beta}{\alpha^3}, \\
L_{02} &= -\frac{\sum_{i=1}^n \delta_i}{\beta^2} - \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \ln\left(\frac{t_i}{\alpha}\right)^2, \\
\sigma_{22} &= (-L_{02})^{-1}, \\
L_{03} &= \frac{\partial^3 L}{\partial \beta^3} = \frac{2\sum_{i=1}^n \delta_i}{\beta^3} - \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \ln\left(\frac{t_i}{\alpha}\right)^3.
\end{aligned}
\tag{3.18}
$$

3.4. General Entropy Loss Function

This is used to determine the degree of overestimation and underestimation of the parameters. It is simply a generalization of the entropy loss function.

The Bayes estimator of θ, denoted by $\widehat{\theta}_{\mathrm{BG}}$ is given as

$$
\widehat{\theta}_{\mathrm{BG}} = \left[E_\theta(\theta)^{-k}\right]^{-1/k} \tag{3.19}
$$

provided $E_\theta(\theta^{-k})$ exist and is finite.

The posterior density function of the survival function and the failure rate under general entropy loss are given, respectively, as

$$S(t)_{\mathrm{BG}} = E\left\{\left[\exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]\right]^{-k} \mid t\right\}$$

$$= \frac{\int\int_0^\infty v(\beta)v_1(\alpha)\left[\exp\left[-(t/\alpha)^{\beta}\right]\right]^{-k} L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}{\int\int_0^\infty v(\beta)v_1(\alpha)L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta},$$

$$F(t)_{\mathrm{BG}} = E\left\{\left[\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}\right]^{-k} \mid t\right\}$$

$$= \frac{\int\int_0^\infty v(\beta)v_1(\alpha)\left[(\beta/\alpha)(t/\alpha)^{\beta-1}\right]^{-k} L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}{\int\int_0^\infty v(\beta)v_1(\alpha)L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}. \tag{3.20}$$

By making use of Lindley procedure as in (3.14), where u_1, u_{11}, and u_2, u_{22} represent the first and second derivatives of the survival function and the failure rate, the following equations are obtained:

$$u = \left\{\exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]\right\}^{-k}, \qquad e = \left[-\left(\frac{t}{\alpha}\right)^{\beta}\right],$$

$$u_1 = \frac{\partial u}{\partial \alpha} = uk\left(\frac{\beta}{\alpha}\right)e, \tag{3.21}$$

$$u_{11} = \frac{\partial^2 u}{\partial \alpha^2} = uk^2\left(\frac{\beta^2}{\alpha^2}\right)e^2 - k\left(\frac{\beta^2}{\alpha^2}\right)eu - k\left(\frac{\beta}{\alpha^2}\right)eu.$$

Hence, $u_2 = \partial u/\partial\beta$ and $u_{22} = \partial^2 u/\partial\beta^2$ follows.

For the failure rate,

$$u = \left[\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}\right]^{-k}, \qquad r = \left(\frac{t}{\alpha}\right)^{\beta-1},$$

$$u_1 = \frac{\partial u}{\partial \alpha} = -\frac{uk\left[-(\beta/\alpha^2\ )r - (\beta-1)(\beta/\alpha^2\ )r\right]\alpha}{\beta r},$$

$$u_{11} = \frac{\partial^2 u}{\partial \alpha^2} = \frac{uk^2\left[-(\beta/\alpha^2)r - (\beta-1)(\beta/\alpha^2)r\right]^2\alpha^2}{\beta^2 r^2} - \frac{uk\left[2(\beta/\alpha^3)r + 3(\beta/\alpha^3)(\beta-1)r + (\beta/\alpha^3)r(\beta-1)^2\right]\alpha}{\beta r} - \frac{uk\left[-(\beta/\alpha^2)r - (\beta-1)(\beta/\alpha^2)r\right]}{\beta r} - \frac{uk\left[-(\beta/\alpha^2)r - (\beta-1)(\beta/\alpha^2)r\right](\beta-1)}{\beta r}. \tag{3.22}$$

With the same approach as given above $u_{22} = \partial u/\partial\beta$ and $u_{22} = \partial^2 u/\partial\beta^2$ are obtained.

3.5. Squared Error Loss Function

The most common loss function used for Bayesian estimation is the squared error (SE), also called quadratic loss. The square error loss denotes the punishment in using $\widehat{\theta}$ to estimate θ and is given as $E(t \mid \theta) = [\widehat{\theta}(t) - \theta]^2$, where the expectation is taken over the joint distribution of θ and (t). The posterior density function of the survival function and the failure rate under the Symmetric loss function are given as

$$\begin{aligned} S(t)_{\mathrm{BS}} &= E\left\{\exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \mid t\right\} \\ &= \frac{\int\int_0^\infty v(\beta)v_1(\alpha)\left[\exp\left[-(t/\alpha)^{\beta}\right]\right]L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}{\int\int_0^\infty v(\beta)v_1(\alpha)L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}, \\ F(t)_{\mathrm{BS}} &= E\left[\left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1} \mid t\right] \\ &= \frac{\int\int_0^\infty v(\beta)v_1(\alpha)\left[(\beta/\alpha)(t/\alpha)^{\beta-1}\right]L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}{\int\int_0^\infty v(\beta)v_1(\alpha)L(t_i \mid \alpha,\beta)\,d\alpha\,d\beta}. \end{aligned} \tag{3.23}$$

Applying the same Lindley approach here as in (3.14), we have for the Survival Function

$$\begin{aligned} u &= \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right], \qquad u_1 = \frac{\partial u}{\partial \alpha} = -\left(\frac{\beta}{\alpha}\right)eu, \\ u_{11} &= \frac{\partial^2 u}{\partial \alpha^2} = \left(\frac{\beta}{\alpha^2}\right)eu + \left(\frac{\beta^2}{\alpha^2}\right)eu + \left(\frac{\beta^2}{\alpha^2}\right)e^2u. \end{aligned} \tag{3.24}$$

In a similar approach $u_2 = \partial u/\partial\beta$ and $u_{22} = \partial^2 u/\partial\beta^2$ can be obtained.
For the Failure Rate,

$$u = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}, \qquad d = \ln\left(\frac{t}{\alpha}\right),$$
$$u_2 = \frac{\partial u}{\partial\beta} = \left(\frac{1}{\alpha}\right)r + ud, \qquad u_{22} = \frac{\partial^2 u}{\partial\beta^2} = \left(\frac{2}{\alpha}\right)rd^2 + ud. \tag{3.25}$$

$u_1 = \partial u/\partial\alpha$ and $u_{11} = \partial^2 u/\partial\alpha^2$ follow in like manner.

With respect to the generalised noninformative prior, the same procedures as above are also employed but $\rho = \ln v_1(\alpha) + \ln v(\beta)$ is substituted by $\rho = \ln[v_2(\alpha,\beta)]$.

4. Simulation Study

We have considered in this simulation study a sample size of $n = 25,\ 50$ and 100, which is representative of small, moderate, and large data sets. The following steps were employed to generate the data.

A lifetime T is generated from the sample sizes indicated above from the Weibull distribution which represents failure of the product or unit. The values of the assumed actual parameters of the Weibull distribution were taken to be $\alpha = 0.5$ and 1.5 and that of $\beta = 0.8$ and 1.2. The same sample size is generated from the uniform distribution for the censored time C with $(0, b)$, where the value of b depends solely on the proportion of the observations that are censored. In our study, we consider the percentage of censoring to be 30. $T = \min(T, C)$ is taken as the minimum of the failure time and the censored time of the observed time T, where

$$T = \begin{cases} \delta_i = 1, & \text{if } X \le C, \\ \delta_i = 0, & \text{if } X > C. \end{cases} \tag{4.1}$$

To compute the Bayes estimates, an assumption is made such that α and β take, respectively, Gamma (a, b) and Gamma (c, d) priors. We set the hyperparameters to zero, that is, $a = b = c = d = 0$ in order to obtain noninformative priors. Note that at this point, the priors become nonproper but the results do not have any significant difference with the implementation of proper priors as stated by Banerjee and Kundu [22].

The values for the loss parameters of both the LINEX and general entropy were $c = k = \pm 1$ and ± 2. For problems on how to choose the loss parameter values, see Calabria and Pulcini [26]. We have also considered the generalised noninformative prior to be $a = 3$ and 5 without loss of generality. These were iterated (R) 1000 times. The mean squared error and the absolute bias values are determined and presented below for the purpose of comparison. Consider the following:

$$\text{MSE}\left(\widehat{\theta}\right) = \frac{\sum_{r=1}^{R}\left(\widehat{\theta}^r - \theta\right)^2}{R-1}, \qquad \text{Abs. Bias}\left(\widehat{\theta}\right) = \frac{\sum_{r=1}^{R}\left|\widehat{\theta}^r - \theta\right|}{R-1}. \tag{4.2}$$

Table 1: MSEs and Abs. Biases (in parenthesis) for the survival function $S(t)$.

n		$\alpha = 0.5$		$\alpha = 1.5$	
		$\beta = 0.8$	$\beta = 1.2$	$\beta = 0.8$	$\beta = 1.2$
25	ML	0.01577 (0.09892)	0.01613 (0.10131)	0.01644 (0.10076)	0.01635 (0.10273)
	BS	0.01581 (0.09909)	0.01618 (0.10149)	0.01648 (0.10092)	0.01639 (0.10290)
	BL ($c = 1$)	0.01681 (0.10213)	0.01731 (0.10556)	0.01564 (0.09971)	0.01707 (0.10299)
	BL ($c = -1$)	0.01613 (0.10035)	0.01732 (0.10361)	0.01625 (0.10014)	0.01687 (0.10428)
	BL ($c = 2$)	0.01604 (0.09968)	0.01742 (0.10428)	0.01629 (0.10079)	0.01711 (0.10437)
	BL ($c = -2$)	0.01623 (0.10091)	0.01753 (0.10573)	0.01578 (0.09839)	0.01708 (0.10266)
	BG ($k = 1$)	0.01568 (0.09809)	0.01770 (0.10566)	0.01520 (0.09737)	0.01667 (0.10269)
	BG ($k = -1$)	0.01565 (0.09855)	0.01774 (0.10563)	0.01565 (0.09976)	0.01721 (0.10453)
	BG ($k = 2$)	0.01495 (0.09495)	0.01613 (0.10047)	0.01549 (0.09792)	0.01722 (0.10326)
	BG ($k = -2$)	0.01589 (0.09983)	0.01686 (0.10303)	0.01549 (0.09711)	0.01812 (0.10809)
50	ML	0.01179 (0.08736)	0.01370 (0.09561)	0.01155 (0.08619)	0.01274 (0.09169)
	BS	0.01180 (0.08741)	0.01371 (0.09566)	0.01156 (0.08624)	0.01276 (0.09174)
	BL ($c = 1$)	0.01160 (0.08666)	0.01354 (0.09606)	0.01164 (0.08681)	0.01299 (0.09248)
	BL ($c = -1$)	0.01137 (0.08563)	0.01246 (0.09134)	0.01144 (0.08683)	0.01299 (0.09263)
	BL ($c = 2$)	0.01180 (0.08794)	0.01275 (0.09056)	0.01199 (0.08865)	0.01282 (0.09201)
	BL ($c = -2$)	0.01192 (0.08840)	0.01353 (0.09489)	0.01197 (0.08823)	0.01239 (0.09009)
	BG ($k = 1$)	0.01179 (0.08768)	0.01312 (0.09299)	0.01139 (0.08609)	0.01263 (0.09120)
	BG ($k = -1$)	0.01799 (0.08722)	0.01259 (0.09168)	0.01109 (0.08408)	0.01274 (0.09257)
	BG ($k = 2$)	0.01143 (0.08651)	0.01320 (0.09432)	0.01136 (0.08635)	0.01273 (0.09260)
	BG ($k = -2$)	0.01159 (0.08686)	0.01229 (0.09049)	0.01133 (0.08472)	0.01306 (0.09325)
100	ML	0.00969 (0.08234)	0.01108 (0.08929)	0.00915 (0.07937)	0.01119 (0.08962)
	BS	0.00969 (0.08235)	0.01108 (0.08931)	0.00715 (0.07938)	0.01119 (0.08963)
	BL ($c = 1$)	0.00953 (0.08105)	0.01109 (0.08969)	0.00932 (0.08059)	0.01136 (0.09084)
	BL ($c = -1$)	0.00938 (0.08148)	0.01062 (0.08779)	0.00952 (0.08159)	0.01056 (0.08705)
	BL ($c = 2$)	0.00945 (0.08189)	0.01087 (0.08877)	0.00746 (0.08212)	0.01067 (0.08779)
	BL ($c = -2$)	0.00970 (0.08226)	0.01111 (0.08950)	0.00943 (0.08102)	0.01078 (0.08804)
	BG ($k = 1$)	0.00951 (0.08202)	0.01132 (0.09110)	0.00916 (0.07991)	0.01124 (0.09032)
	BG ($k = -1$)	0.00958 (0.08190)	0.01085 (0.08838)	0.00933 (0.08117)	0.01071 (0.08810)
	BG ($k = 2$)	0.00905 (0.07915)	0.01118 (0.08959)	0.00845 (0.08126)	0.01068 (0.08739)
	BG ($k = -2$)	0.00893 (0.07912)	0.01071 (0.08751)	0.00946 (0.08176)	0.01124 (0.09051)

ML: maximum likelihood, BG: general entropy loss function, BL: LINEX loss function, BS: squared error loss function.

5. Results and Discussion

From Tables 1 and 3, the most dominant estimator that had the smallest mean squared error is the Bayesian under linear exponential loss function(LINEX). This happened with generalised noninformative prior except at $n = 50$ with $\alpha = 1.5$ and $\beta = 1.2$ that we observed that the noninformative gamma prior gave the smallest MSE. This was followed closely by the general entropy loss function (GELF) with the noninformative gamma prior. What is remarkable is that the smallest absolute bias values occurred mostly with the generalised noninformative prior. GELF was slightly ahead of LINEX but both were better than SELF and that of the MLE.

Table 2: MSEs and Abs. Biases (in parenthesis) for the failure rate $F(t)$.

n		$\alpha = 0.5$		$\alpha = 1.5$	
		$\beta = 0.8$	$\beta = 1.2$	$\beta = 0.8$	$\beta = 1.2$
25	ML	3.68099 (0.54756)	0.67105 (0.70131)	0.32433 (0.22685)	0.08476 (0.24458)
	BS	3.68274 (0.54755)	0.67104 (0.70129)	0.32525 (0.22687)	0.08476 (0.24581)
	BL ($c = 1$)	2.90228 (0.64591)	0.73442 (0.73753)	0.22901 (0.21768)	0.08163 (0.24202)
	BL ($c = -1$)	3.26232 (0.65045)	0.73879 (0.72318)	0.83318 (0.22460)	0.07709 (0.23813)
	BL ($c = 2$)	2.82598 (0.65421)	0.73698 (0.73324)	0.46876 (0.21809)	0.07626 (0.23769)
	BL ($c = -2$)	2.05587 (0.66813)	0.71049 (0.71956)	0.28028 (0.22308)	0.07876 (0.23947)
	BG ($k = 1$)	3.45630 (0.64855)	0.70479 (0.72267)	0.51944 (0.21948)	0.07729 (0.23826)
	BG ($k = -1$)	3.38251 (0.66240)	0.68825 (0.71194)	0.43598 (0.21791)	0.07829 (0.23991)
	BG ($k = 2$)	2.74521 (0.68257)	0.68329 (0.70471)	0.53684 (0.22410)	0.07949 (0.24022)
	BG ($k = -2$)	2.63739 (0.67368)	0.64373 (0.69148)	0.32230 (0.21216)	0.08217 (0.24390)
50	ML	2.74229 (0.54756)	0.55963 (0.66088)	0.10479 (0.18498)	0.05991 (0.21456)
	BS	2.74207 (0.54755)	0.55962 (0.66087)	0.10482 (0.18497)	0.05991 (0.21246)
	BL ($c = 1$)	2.06952 (0.56294)	0.55590 (0.66412)	0.09266 (0.18624)	0.06056 (0.21627)
	BL ($c = -1$)	2.07558 (0.55519)	0.54833 (0.65331)	0.10693 (0.18579)	0.06066 (0.21665)
	BL ($c = 2$)	1.66649 (0.56345)	0.54094 (0.64491)	0.08616 (0.18786)	0.06103 (0.21764)
	BL ($c = -2$)	1.48847 (0.56854)	0.57137 (0.66531)	0.11013 (0.18562)	0.05871 (0.21287)
	BG ($k = 1$)	2.35703 (0.57154)	0.55850 (0.65845)	0.09676 (0.18616)	0.05904 (0.21293)
	BG ($k = -1$)	1.06927 (0.56011)	0.54374 (0.64790)	0.16384 (0.18465)	0.06124 (0.21868)
	BG ($k = 2$)	1.05864 (0.55842)	0.56163 (0.66459)	0.08869 (0.18615)	0.06057 (0.21822)
	BG ($k = -2$)	0.85649 (0.55619)	0.52414 (0.63780)	0.08375 (0.18829)	0.06088 (0.21713)
100	ML	0.67729 (0.53509)	0.50136 (0.64635)	0.05015 (0.17142)	0.05539 (0.21498)
	BS	0.67733 (0.53509)	0.50136 (0.64635)	0.05015 (0.17142)	0.05539 (0.21498)
	BL ($c = 1$)	0.45556 (0.52062)	0.49361 (0.64479)	0.06047 (0.17411)	0.05607 (0.21656)
	BL ($c = -1$)	0.42816 (0.52640)	0.47733 (0.63167)	0.09317 (0.17839)	0.05342 (0.21074)
	BL ($c = 2$)	0.48344 (0.52865)	0.49163 (0.64201)	0.04895 (0.17618)	0.05405 (0.21265)
	BL ($c = -2$)	0.47254 (0.52719)	0.49932 (0.64575)	0.05123 (0.17538)	0.05406 (0.21182)
	BG ($k = 1$)	0.44526 (0.52608)	0.50362 (0.65328)	0.05159 (0.17252)	0.05432 (0.21427)
	BG ($k = -1$)	0.51433 (0.53095)	0.48884 (0.63931)	0.06519 (0.17545)	0.05395 (0.21272)
	BG ($k = 2$)	0.44763 (0.51844)	0.50123 (0.64629)	0.07165 (0.17638)	0.05428 (0.21193)
	BG ($k = -2$)	0.46723 (0.51550)	0.48396 (0.63387)	0.05236 (0.17513)	0.05526 (0.21548)

ML: maximum likelihood, BG: general entropy loss function, BL: LINEX loss function, BS: squared error loss function.

The implication from above is that the generalised noninformative prior worked remarkable well under both LINEX and GELF with respect to the mean squared error and absolute bias. What need to be stated is that the generalised prior ensures that since there is very little knowledge on the function being estimated as a result of which a noninformative prior is assumed and obtained from the available data via MLE, then one should ensure that at the end it plays little role so that the Bayesian inference is based on the available data.

Considering Tables 2 and 4, we noticed again that the generalised prior performed astonishingly well against the noninformative prior under both the GELF and LINEX. Both priors performed equally well under general entropy and linear exponential loss functions. With the absolute bias values, the generalised noninformative prior incredibly performed

Table 3: MSEs and Abs. Biases (in parenthesis) for the survival function $S(t)$ using GNP.

n		$\alpha = 0.5$		$\alpha = 1.5$	
		$\beta = 0.8$	$\beta = 1.2$	$\beta = 0.8$	$\beta = 1.2$
		$a = 5$, [3]		$a = 5$, [3]	
25	BS	0.01671 (0.10237)	0.01731 (0.10550)	0.01553 (0.09804)	0.01740 (0.10389)
		[0.01528 (0.09853)]	[0.01733 (0.10441)]	[0.01632 (0.10123)]	[0.01861 (0.10795)]
	BL ($c = 1$)	0.01736 (0.01025)	0.01608 (0.10052)	0.01531 (0.09805)	0.01694 (0.10227)
		[0.01559 (0.09876)]	[0.01726 (0.10385)]	[0.01585 (0.09796)]	[0.01755 (0.10476)]
	BL ($c = -1$)	0.01756 (0.10315)	0.01655 (0.10308)	0.01614 (0.10075)	0.01748 (0.10470)
		[0.01544 (0.09803)]	[0.01806 (0.10735)]	[0.01680 (0.10254)]	[0.01727 (0.10453)]
	BL ($c = 2$)	0.01797 (0.10635)	0.01649 (0.10116)	0.01467 (0.09588)	0.01718 (0.10317)
		[0.01716 (0.10194)]	[0.01806 (0.10617)]	[0.01506 (0.09667)]	[0.01793 (0.10640)]
	BL ($c = -2$)	0.01756 (0.10499)	0.01637 (0.10090)	0.01571 (0.09796)	0.01748 (0.10551)
		[0.01554 (0.09817)]	[0.01732 (0.10397)]	[0.01602 (0.10003)]	[0.01857 (0.10914)]
	BG ($k = 1$)	0.01645 (0.10157)	0.01659 (0.10271)	0.01587 (0.09919)	0.01719 (0.10355)
		[0.01519 (0.09767)]	[0.01741 (0.10468)]	[0.01649 (0.10107)]	[0.01782 (0.10583)]
	BG ($k = -1$)	0.01660 (0.10067)	0.01734 (0.10515)	0.01621 (0.09978)	0.01808 (0.10817)
		[0.01554 (0.09757)]	[0.01609 (0.10096)]	[0.01539 (0.09644)]	[0.01793 (0.10731)]
	BG ($k = 2$)	0.01794 (0.10643)	0.01773 (0.10477)	0.01653 (0.10009)	0.01688 (0.10393)
		[0.01551 (0.09821)]	[0.01690 (0.10249)]	[0.01518 (0.09757)]	[0.01667 (0.10306)]
	BG ($k = -2$)	0.01623 (0.10158)	0.01730 (0.10519)	0.01562 (0.09759)	0.01748 (0.10465)
		[0.01654 (0.10175)]	[0.01773 (0.10552)]	[0.01636 (0.10043)]	[0.01822 (0.10797)]
50	BS	0.01178 (0.08765)	0.01299 (0.09209)	0.01174 (0.08754)	0.01312 (0.09343)
		[0.01163 (0.08629)]	[0.01255 (0.09154)]	[0.01163 (0.08679)]	[0.01272 (0.09223)]
	BL ($c = 1$)	0.01092 (0.08456)	0.01259 (0.09203)	0.01134 (0.08587)	0.01276 (0.09198)
		[0.01128 (0.08459)]	[0.01284 (0.09326)]	[0.01181 (0.08617)]	[0.01289 (0.09290)]
	BL ($c = -1$)	0.01201 (0.08874)	0.01266 (0.09119)	0.01149 (0.08649)	0.01270 (0.09149)
		[0.01118 (0.08561)]	[0.01274 (0.09196)]	[0.01126 (0.08554)]	[0.01251 (0.09036)]
	BL ($c = 2$)	0.01126 (0.08480)	0.01288 (0.09267)	0.01164 (0.08756)	0.01307 (0.09326)
		[0.01149 (0.08640)]	[0.01261 (0.09249)]	[0.01106 (0.08521)]	[0.01335 (0.09481)]
	BL ($c = -2$)	0.01119 (0.08567)	0.01304 (0.09343)	0.01135 (0.08484)	0.01299 (0.09379)
		[0.01158 (0.08676)]	[0.01234 (0.09038)]	[0.01209 (0.08853)]	[0.01310 (0.09310)]
	BG ($k = 1$)	0.01109 (0.08496)	0.01322 (0.09375)	0.01137 (0.08565)	0.01309 (0.09474)
		[0.01111 (0.08467)]	[0.01287 (0.09549)]	[0.01176 (0.08690)]	[0.01279 (0.09197)]
	BG ($k = -1$)	0.01187 (0.08821)	0.01365 (0.09533)	0.01191 (0.08832)	0.01263 (0.09130)
		[0.01147 (0.08608)]	[0.01289 (0.09229)]	[0.01126 (0.08621)]	[0.01298 (0.09259)]
	BG ($k = 2$)	0.01135 (0.08535)	0.01307 (0.09349)	0.01126 (0.08878)	0.01361 (0.09489)
		[0.01197 (0.08815)]	[0.01257 (0.09113)]	[0.01145 (0.08588)]	[0.01251 (0.09138)]
	BG ($k = -2$)	0.01199 (0.08802)	0.01325 (0.09363)	0.01232 (0.08958)	0.01259 (0.09138)
		[0.01166 (0.08674)]	[0.01299 (0.09296)]	[0.01138 (0.08564)]	[0.01302 (0.09359)]
100	BS	0.00967 (0.08273)	0.01069 (0.08780)	0.00935 (0.08153)	0.01107 (0.08928)
		[0.00936 (0.08143)]	[0.01087 (0.08831)]	[0.00929 (0.08039)]	[0.01084 (0.08928)]
	BL ($c = 1$)	0.00984 (0.08327)	0.01084 (0.08875)	0.00867 (0.07797)	0.01061 (0.08743)
		[0.00961 (0.08267)]	[0.01090 (0.08894)]	[0.00969 (0.08240)]	[0.01093 (0.08894)]
	BL ($c = -1$)	0.00951 (0.08128)	0.01051 (0.08688)	0.00975 (0.08276)	0.01119 (0.08929)
		[0.00925 (0.08078)]	[0.01043 (0.08648)]	[0.00963 (0.08201)]	[0.01110 (0.08923)]
	BL ($c = 2$)	0.00937 (0.08062)	0.01064 (0.08733)	0.00929 (0.08096)	0.01105 (0.08896)
		[0.00973 (0.08233)]	[0.01101 (0.08887)]	[0.00969 (0.08302)]	[0.01118 (0.09015)]

Table 3: Continued.

n	$\alpha = 0.5$		$\alpha = 1.5$	
	$\beta = 0.8$	$\beta = 1.2$	$\beta = 0.8$	$\beta = 1.2$
	$a = 5$, [3]		$a = 5$, [3]	
BL ($c = -2$)	0.00952 (0.08184)	0.01098 (0.08906)	0.00966 (0.08244)	0.01087 (0.08856)
	[0.00994 (0.08372)]	[0.01102 (0.08887)]	[0.00938 (0.08141)]	[0.01094 (0.08877)]
BG ($k = 1$)	0.00895 (0.07874)	0.01091 (0.08838)	0.00946 (0.08088)	0.01059 (0.08729)
	[0.00926 (0.08019)]	[0.01093 (0.08855)]	[0.00955 (0.08214)]	[0.01108 (0.08918)]
BG ($k = -1$)	0.00914 (0.08047)	0.01093 (0.08833)	0.00956 (0.08156)	0.01078 (0.08802)
	[0.00897 (0.07956)]	[0.01091 (0.08844)]	[0.00985 (0.08260)]	[0.01001 (0.08491)]
BG ($k = 2$)	0.00929 (0.08071)	0.01106 (0.08889)	0.00921 (0.07979)	0.01142 (0.09069)
	[0.00963 (0.08220)]	[0.01111 (0.08947)]	[0.00975 (0.08284)]	[0.01092 (0.08865)]
BG ($k = -2$)	0.00936 (0.08122)	0.01128 (0.09042)	0.00990 (0.08379)	0.01111 (0.08941)
	[0.00961 (0.08213)]	[0.01072 (0.08811)]	[0.00947 (0.08152)]	[0.01132 (0.09072)]

GNP: generalised nonInformative prior, BG: general entropy loss function, BL: LINEX loss function, BS: squared error loss function.

better under LINEX loss function but was almost equal with the noninformative prior under the general entropy loss function.

To obtain the MSE values for each estimated value, the MSE is calculated for each of the one thousand estimated values of the survival function and the failure rate that is, from 1 to 1000. At the end, we obtain the average of the MSE, values. Our aim is to find out how close the estimated values of the estimators are to the true value. The absolute bias values are obtained in like manner and of course from the same simulated values as that of the MSE.

6. Conclusion

In this study, we consider the point estimation of the Weibull distribution based on right censoring through simulation. MLE and Bayes estimators are applied to estimate the survival function and the failure rate of this lifetime distribution. The Bayes estimators are obtained using linear exponential, general entropy, and squared error loss functions. We also employed the Bayesian noninformative prior approach in estimating survival function and the failure rate. In order to reduce the complicated integrals that are in the posterior distribution which cannot explicitly be obtained in close form, we employed the Lindley approximation procedure to calculate the Bayes estimators.

Another point worth noting is that we assumed an informative prior for the scale parameter and the shape parameter which led us into obtaining an improper prior. We also made a proposition for a generalised noninformative prior.

From the results and discussions above, it is evident that the proposed generalised noninformative prior performed quite well than the noninformative gamma prior. In all the cases, the Bayesian estimator using the generalised noninformative prior under the linear exponential loss function overall performed better than the other estimators and under the other different loss functions with respect to the mean squared error. It has also been observed that the smallest absolute bias values occurred predominantly with the generalised

Table 4: MSEs and Abs. Biases (in parenthesis) for the failure rate $F(t)$ using GNP.

n		$\alpha = 0.5$		$\alpha = 1.5$	
		$\beta = 0.8$	$\beta = 1.2$	$\beta = 0.8$	$\beta = 1.2$
		$a = 5$, [3]		$a = 5$, [3]	
25	BS	1.60882 (0.60581)	0.75601 (0.74247)	0.26918 (0.22161)	0.08242 (0.24192)
		[3.50808 (0.66123)]	[0.70983 (0.72011)]	[0.21814 (0.22270)]	[0.08227 (0.24568)]
	BL ($c = 1$)	3.10223 (0.66497)	0.66435 (0.70130)	0.24086 (0.21740)	0.07698 (0.23560)
		[0.97145 (0.64334)]	[0.73049 (0.73445)]	[0.46015 (0.22941)]	[0.08183 (0.24475)]
	BL ($c = -1$)	5.07183 (0.70026)	0.72658 (0.71728)	0.17702 (0.22059)	0.07997 (0.24249)
		[1.89290 (0.65885)]	[0.72855 (0.73625)]	[0.28853 (0.22911)]	[0.07864 (0.23852)]
	BL ($c = 2$)	2.57595 (0.67277)	0.70089 (0.71957)	0.19602 (0.21443)	0.08062 (0.24190)
		[1.42447 (0.65551)]	[0.73245 (0.73581)]	[0.26676 (0.22149)]	[0.08261 (0.24604)]
	BL ($c = -2$)	2.47065 (0.68677)	0.68712 (0.70818)	0.29206 (0.22263)	0.07838 (0.24087)
		[2.90974 (0.65568)]	[0.69822 (0.71461)]	[0.31298 (0.23181)]	[0.08179 (0.24609)]
	BG ($k = 1$)	5.09142 (0.63923)	0.68482 (0.71454)	0.32629 (0.21537)	0.07963 (0.24117)
		[1.59876 (0.64264)]	[0.71224 (0.72443)]	[0.39187 (0.22194)]	[0.08104 (0.24269)]
	BG ($k = -1$)	1.11653 (0.60663)	0.73329 (0.72535)	0.28561 (0.21867)	0.07938 (0.24085)
		[2.05799 (0.65800)]	[0.67003 (0.70302)]	[0.28235 (0.28235)]	[0.07783 (0.24017)]
	BG ($k = 2$)	3.26989 (0.63433)	0.70409 (0.72102)	0.14943 (0.21824)	0.08079 (0.24219)
		[3.59616 (0.65300)]	[0.70879 (0.71412)]	[0.20827 (0.22005)]	[0.08084 (0.24227)]
	BG ($k = -2$)	1.56123 (0.63166)	0.68200 (0.71084)	0.22126 (0.21289)	0.07983 (0.24229)
		[1.82844 (0.64736)]	[0.71757 (0.72167)]	[0.16709 (0.21693)]	[0.07978 (0.24356)]
50	BS	0.99284 (0.55903)	0.53895 (0.64125)	0.21602 (0.18856)	0.06088 (0.21759)
		[1.20392 (0.55531)]	[0.53767 (0.64578)]	[0.23509 (0.18781)]	[0.05879 (0.21411)]
	BL ($c = 1$)	0.85858 (0.55004)	0.54626 (0.65313)	0.21497 (0.18676)	0.06021 (0.21590)
		[0.84299 (0.54806)]	[0.54644 (0.65612)]	[0.18170 (0.18565)]	[0.06051 (0.21811)]
	BL ($c = -1$)	0.90621 (0.56868)	0.55652 (0.65539)	0.17717 (0.18777)	0.06052 (0.21565)
		[0.70963 (0.54405)]	[0.53659 (0.64556)]	[0.19359 (0.18475)]	[0.06063 (0.21566)]
	BL ($c = 2$)	0.84116 (0.54422)	0.55124 (0.65304)	0.19584 (0.18896)	0.06132 (0.21808)
		[1.05227 (0.55656)]	[0.55016 (0.65747)]	[0.19113 (0.18416)]	[0.06118 (0.21996)]
	BL ($c = -2$)	0.83699 (0.55410)	0.54021 (0.64881)	0.19281 (0.18479)	0.06111 (0.21948)
		[0.87199 (0.55693)]	[0.52738 (0.63962)]	[0.18451 (0.18884)]	[0.06072 (0.21715)]
	BG ($k = 1$)	1.08699 (0.55039)	0.55381 (0.65826)	0.17302 (0.18223)	0.06119 (0.21882)
		[0.95041 (0.55106)]	[0.54985 (0.65467)]	[0.24233 (0.18548)]	[0.06097 (0.21819)]
	BG ($k = -1$)	1.12507 (0.55889)	0.55956 (0.66129)	0.25057 (0.19005)	0.06023 (0.21461)
		[1.01425 (0.55157)]	[0.55221 (0.65488)]	[0.11878 (0.18655)]	[0.06018 (0.21560)]
	BG ($k = 2$)	1.17354 (0.55312)	0.57334 (0.67120)	0.17699 (0.18661)	0.06244 (0.22011)
		[1.00135 (0.56841)]	[0.54657 (0.64917)]	[0.10609 (0.18484)]	[0.05985 (0.21521)]
	BG ($k = -2$)	1.18804 (0.56841)	0.56341 (0.66432)	0.13249 (0.18605)	0.06024 (0.21510)
		[1.15358 (0.55573)]	[0.54699 (0.65363)]	[0.14781 (0.18359)]	[0.06057 (0.21754)]
100	BS	0.66670 (0.53230)	0.48857 (0.63609)	0.05482 (0.17604)	0.05524 (0.21483)
		[0.44568 (0.52663)]	[0.49065 (0.64062)]	[0.06864 (0.17410)]	[0.05323 (0.21001)]
	BL ($c = 1$)	0.47245 (0.53712)	0.48375 (0.63740)	0.04677 (0.16927)	0.05397 (0.21152)
		[0.48126 (0.52985)]	[0.49223 (0.64333)]	[0.06145 (0.17697)]	[0.05467 (0.21400)]
	BL ($c = -1$)	0.50560 (0.52798)	0.48800 (0.63826)	0.06408 (0.17864)	0.05512 (0.21358)
		[0.48709 (0.52132)]	[0.47258 (0.62645)]	[0.06531 (0.17577)]	[0.05552 (0.21507)]
	BL ($c = 2$)	0.44645 (0.52047)	0.48540 (0.63612)	0.06605 (0.17422)	0.05524 (0.21454)
		[0.80817 (0.52984)]	[0.48774 (0.63795)]	[0.06651 (0.17786)]	[0.05648 (0.21780)]

Table 4: Continued.

n	$\alpha = 0.5$		$\alpha = 1.5$	
	$\beta = 0.8$	$\beta = 1.2$	$\beta = 0.8$	$\beta = 1.2$
	$a = 5$, [3]		$a = 5$, [3]	
BL ($c = -2$)	0.50473 (0.52856) [0.52768 (0.53450)]	0.49342 (0.64368) [0.49421 (0.64037)]	0.05039 (0.17666) [0.07692 (0.17494)]	0.05434 (0.21327) [0.05561 (0.21493)]
BG ($k = 1$)	0.44440 (0.51305) [0.59699 (0.52223)]	0.49572 (0.64148) [0.49308 (0.64049)]	0.07150 (0.17498) [0.05309 (0.17608)]	0.05361 (0.21112) [0.05444 (0.21287)]
BG ($k = -1$)	0.49145 (0.52292) [0.46108 (0.52107)]	0.49925 (0.64372) [0.49309 (0.63951)]	0.05013 (0.17478) [0.04829 (0.17618)]	0.05340 (0.21109) [0.05185 (0.20729)]
BG ($k = 2$)	0.50315 (0.52624) [0.45561 (0.52818)]	0.50124 (0.64500) [0.49235 (0.64167)]	0.05319 (0.17324) [0.05073 (0.17732)]	0.05761 (0.22003) [0.05492 (0.21381)]
BG ($k = -2$)	0.55240 (0.52938) [0.42269 (0.52581)]	0.51051 (0.65583) [0.48926 (0.63954)]	0.04795 (0.17732) [0.06453 (0.17769)]	0.05482 (0.21406) [0.05630 (0.21743)]

GNP: generalised noninformative prior, BG: general entropy loss function, BL: LINEX loss function, BS: squared error loss function.

noninformative prior under general entropy loss function for both the survival function and the failure rate.

Acknowledgments

The authors would like to thank the anonymous referees for their suggestions and comments. The first author would like to personally thank the Editor, Professor Carlo Cattani, for his encouragement and suggestions.

References

[1] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York, NY, USA, 1982.
[2] H. Rinne, *The Weibull Distribution*, CRC Press, Taylor and Francis Group LLC, New York, NY, USA, 2009.
[3] S. K. Sinha, "Bayes estimation of the reliability function and hazard rate of a weibull failure time distribution," *Tranbajos De Estadistica*, vol. 1, pp. 47–56, 1986.
[4] A. A. Abdel-Wahid and A. Winterbottom, "Approximate Bayesian estimates for the Weibull reliability function and hazard rate from censored data," *Journal of Statistical Planning and Inference*, vol. 16, no. 3, pp. 277–283, 1987.
[5] M. A. Al Omari and N. A. Ibrahim, "Bayesian survival estimation estimation for weibull distribution with censored data," *Journal of Applied Sciences*, vol. 11, pp. 393–396, 2011.
[6] C. B. Guure, N. A. Ibrahim, and A. M. Al Omari, "Bayesian estimation of two-parameter weibull distribution using extension of Jeffreys' prior information with three loss functions," *Mathematical Problems in Engineering*, vol. 2012, Article ID 589640, 13 pages, 2012.
[7] H. Syuan-Rong and W. Shuo-Jye, "Bayesian estimation and prediction for weibull model with progressive censoring," *Journal of Statistical Computation and Simulation*, pp. 1–14, 2011.
[8] C. Cattani and A. Ciancio, "Separable transition density in the hybrid model for tumor-immune system competition," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 610124, 6 pages, 2012.
[9] X. Wang, M. Xia, C. Huiwen, Y. Gao, and C. Cattani, "Hidden-markov-models-based dynamic hand gesture recognition," *Mathematical Problems in Engineering*, vol. 2012, Article ID 986134, 11 pages, 2012.
[10] C. Cattani, "On the existence of wavelet symmetries in archaea DNA," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 673934, 21 pages, 2012.
[11] C. Cattani, "Fractional calculus and shannon wavelet," *Mathematical Problems in Engineering*, vol. 2012, Article ID 502812, 26 pages, 2012.
[12] R. D. Gupta and D. Kundu, "Generalized exponential distribution: different method of estimations," *Journal of Statistical Computation and Simulation*, vol. 69, no. 4, pp. 315–337, 2001.

[13] A. Zellner, "Bayesian estimation and prediction using asymmetric loss functions," *Journal of the American Statistical Association*, vol. 81, no. 394, pp. 446–451, 1986.
[14] F. M. Al-Aboud, "Bayesian estimations for the extreme value distribution using progressive censored data and asymmetric loss," *International Mathematical Forum*, vol. 4, no. 33–36, pp. 1603–1622, 2009.
[15] F. M. Al-Athari, "Parameter estimation for the double-pareto distribution," *Journal of Mathematics and Statistics*, vol. 7, no. 4, pp. 289–294, 2011.
[16] C. Cattani, G. Pierro, and G. Altieri, "Entropy and multifractality for the myeloma multiple TET 2 gene," *Mathematical Problems in Engineering*, vol. 2012, Article ID 193761, 14 pages, 2012.
[17] B. N. Pandey, N. Dwividi, and B. Pulastya, "Comparison between Bayesian and maximum likelihood estimation of the scale parameter in weibull distribution with known shape under linex loss function," *Journal of Scientific Research*, vol. 55, pp. 163–172, 2011.
[18] D. Kundu and R. D. Gupta, "Generalized exponential distribution: bayesian estimations," *Computational Statistics & Data Analysis*, vol. 52, no. 4, pp. 1873–1883, 2008.
[19] P. Vasile, P. Eugeine, and C. Alina, "Bayes estimators of modified-Weibull distribution parameters using Lindley's approximation," *WSEAS Transactions on Mathematics*, vol. 9, no. 7, pp. 539–549, 2010.
[20] A. P. Basu and N. Ebrahimi, "Bayesian approach to life testing and reliability estimation using asymmetric loss function," *Journal of Statistical Planning and Inference*, vol. 29, no. 1-2, pp. 21–31, 1991.
[21] J. O. Berger and D. C. Sun, "Bayesian analysis for the poly-Weibull distribution," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1412–1418, 1993.
[22] A. Banerjee and D. Kundu, "Inference based on type-II hybrid censored data from a weibull distribution," *IEEE Transactions on Reliability*, vol. 57, no. 2, pp. 369–378, 2008.
[23] J. O. Berger, "The case for objective Bayesian analysis," *Bayesian Analysis*, vol. 1, no. 3, pp. 385–402, 2006.
[24] A. A. Soliman, A. H. Abd Ellah, and S. B. Ryan, "Comparison of estimates using record statistics from weibull model: bayesian and non-Bayesian approaches," *Computational Statistics & Data Analysis*, vol. 51, pp. 2065–2077, 2006.
[25] D. V. Lindley, "Approximate Bayesian methods," *Trabajos Estadist*, vol. 31, pp. 223–237, 1980.
[26] R. Calabria and G. Pulcini, "Point estimation under asymmetric loss functions for left-truncated exponential samples," *Communications in Statistics. Theory and Methods*, vol. 25, no. 3, pp. 585–600, 1996.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2012, Article ID 857210, 7 pages
doi:10.1155/2012/857210

Research Article

Entropy-Based Maximally Stable Extremal Regions for Robust Feature Detection

Huiwen Cai,[1] Xiaoyan Wang,[2] Ming Xia,[2] and Yangsheng Wang[1]

[1] *Digital Interactive Media Laboratory, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[2] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Huiwen Cai, huiwen.cai@ia.ac.cn

Received 25 August 2012; Accepted 2 October 2012

Academic Editor: Sheng-yong Chen

Maximally stable extremal regions (MSER) is a state-of-the-art method in local feature detection. However, this method is sensitive to blurring because, in blurred images, the intensity values in region boundary will vary more slowly, and this will undermine the stability criterion that the MSER relies on. In this paper, we propose a method to improve MSER, making it more robust to image blurring. To find back the regions missed by MSER in the blurred image, we utilize the fact that the entropy of probability distribution function of intensity values increases rapidly when the local region expands across the boundary, while the entropy in the central part remains small. We use the entropy averaged by the regional area as a measure to reestimate regions missed by MSER. Experiments show that, when dealing with blurred images, the proposed method has better performance than the original MSER, with little extra computational effort.

1. Introduction

Maximally stable extremal regions (MSER) [1] is a popular local invariant feature detection method because it has many highly desirable properties, such as invariance to monotonic intensity transformation, invariance to adjacency preserving (continuous) transformation, and low computational complexity [2–4]. It is shown in a detailed study [5] that MSER has the best repeatability and accuracy, except when dealing with blurred image. For blurred images, some local regions may no longer be detected by MSER, because MSER only detects regions whose shape is globally stable with respect to intensity perturbation. That is to say, the intensity difference should be large between pixels lying on the boundary and the outside. In the blurred image, sharp intensity contrast may be transformed into gradual variation. As

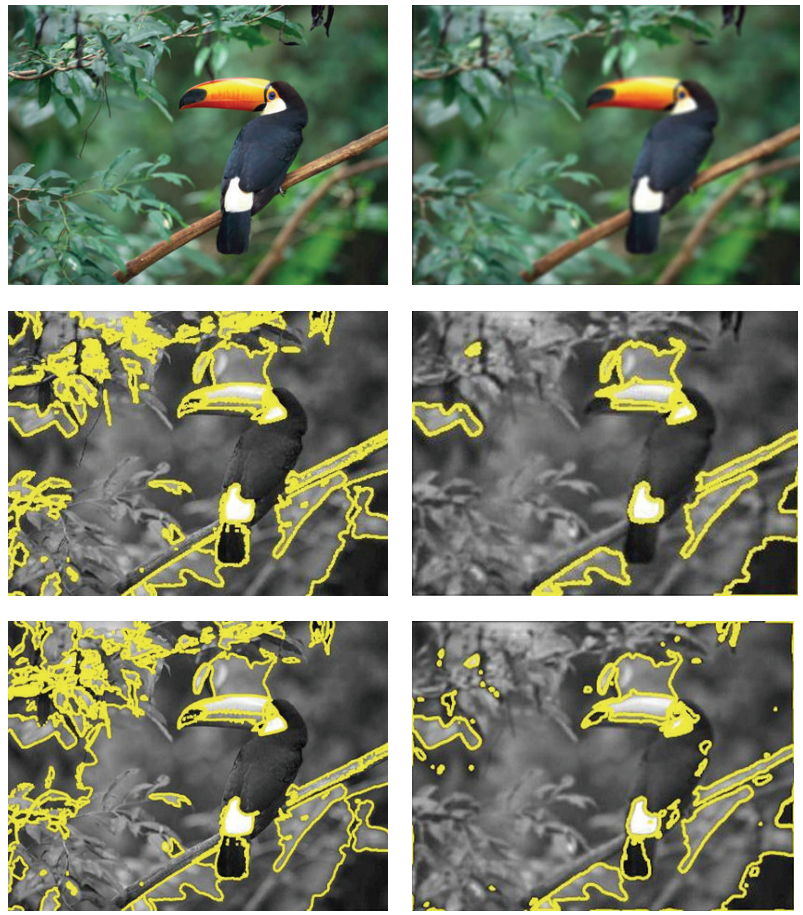


Figure 1: Boundaries of the detected blobs in images. The two on the top are the images for processing. The two in the middle are results of the MSER method. The two in the bottom are results of our method. The left ones are original images, while the right ones are the blurred versions.

demonstrated on the top of Figure 1, with MSER, some regions are lost in the blurred image due to the lack of sharp boundary, like the mouth and tail of the bird.

A blurred image can be seen as an image in a larger scale [6, 7]. Local feature detectors, which are good at dealing with blurring, like the Hessian affine detector and edge-based regions method, all rely on automatic scale selection methods. They construct a measure of local structure within the region and select the local maxima over the scales as the characteristic scale of the region. Even when the image is blurred, proper integration scale and differential scale can still be selected, the same structure can be found, not depending on intensity difference to find the boundary. A multiscale approach is used in [8] to improve MSER's ability to detect features in blurred images, which is a close concept to automatic scale selection. A shortcoming in both methods is that they need to consider many scales in each image position, which greatly increases the computational complexity [9–11]. Moreover, when measuring the structure within the image, a further scale selection may be needed to compute derivatives, such as the Laplacian operator in the Hessian affine detector. Another way to make MSER workable on blurred image is to relax its globally stable criteria, as done in

[12]. Stability is required only locally, for the primitives constituting the three-point frames. Higher number of features and better covered image are achieved, at the expense of more computational effort.

To enhance MSER's ability in the situation of blurring, we can still borrow the idea of using local image structure as a measure to select the proper region scale. In a blurred image, considering MSER's discarded extremal regions which are retained in original image before blurring, we can find a common phenomenon—intensities on the boundary vary more slowly compared to the original image, but the intensities in the central part are kept almost unchanged. In other words, entropy is large on the boundary and small inside. These extremal regions are lost because they are considered not stable with respect to intensity perturbation. To find these regions, we just need to find the extremal regions with minimal entropy. An advantage using entropy is that it does not rely on intensity difference, which may be changed in a blurred image. Another advantage is that entropy measures the region as a whole, instead of just one or a few pixels, which may be easily affected by blurring.

We just use the nested extremal regions extracted in the process of MSER for comparison of entropy, so that there is no additional effort for constructing series of regions of different scales, as what is done in automatic scale selection. Utilizing the nested extremal regions also facilitates the computation of entropy, which will be detailed in Section 2.

In the proposed method, we aim at finding the largest region with the minimal entropy. We use average entropy as our measure and find the region with the lowest average entropy within a local range in the nested extremal regions.

There is another popular local feature detector, "salient region" [13], using the entropy as a criterion. Different from our method, salient region looks for the smallest regions with maximal entropy, because it believes that these regions may contain more information. However, considering blurring, largest region with minimal entropy may be more proper. This is because, after blurring, entropy of region is inclined to increase. Therefore, region scale with maximal entropy will increase but scale with minimal entropy will decrease [14–17]. This could make the salient regions confusable with the background, but regions with minimal entropy are unaffected.

2. Entropy-Based MSER

The proposed method is an extension of the original MSER method, adding entropy as another criterion to find back extremal regions missed by MSER due to blurring. Here is the whole algorithm. Steps (1)–(4) describe the original MSER method, and step (5) is our additional step.

(1) Sort all the pixels by intensity.

(2) Place pixel one by one (in intensity order) in the image, and update the connected component structure, which forms the nested extremal regions.

(3) Compute the area variation for each extremal region:

$$\mathrm{var}_i = \frac{\mathrm{Area}_{i+\Delta} - \mathrm{Area}_i}{\mathrm{Area}_i}. \tag{2.1}$$

Figure 2: Comparison on blurred images. Here boundaries of the detected blobs are shown. The middle six are results of the MSER method. The right six are results of our method.

Here i represents the extremal region with maximal intensity i, and $i + \Delta$ refers to the extremal region expanded from the ith one, with maximal intensity $i + \Delta$. var_i is the relative difference of area when maximal intensity rises from i to $i + \Delta$.

(4) Traverse the nested extremal regions. Find the maximally stable extremal one, which has "var" smaller than its immediate parent and ancestor in the nest.

(5) For the extremal regions not considered as maximally stable, compute their entropy averaged by area. If an extremal region's average entropy is a local minimal in the neighborhood of the nest, we will select it as a complement to the MSERs.

It should be noted that the entropy of an extremal region can be easily calculated, because extremal regions are formed by sorted intensity; thus only subset of intensity is under consideration for each region. Here we show how to compute the average entropy in an incremental way.

Considering R_l as an extremal region with maximal intensity l and its direct parent in the nest R_{l+k}. R_l's intensity histogram can be represented as a_i $(i = 1, 2, \ldots, l)$. Because R_{l+k} contains R_l, besides a_i, there is additional intensity histogram b_i $(i = 1, 2, \ldots, l, l+1, \ldots, l+k)$. Entropy and area can be calculated as follows.

(1) Area of R_l: $\mathrm{S}_l = \sum_{i=1}^{l} a_i$.

(2) Area of R_{l+k}: $\mathrm{S}_{l+k} = \sum_{i=1}^{l} a_i + \sum_{i=1}^{l+k} b_i = S_l + \sum_{i=1}^{l+k} b_i$.

(3) Entropy of R_l: $H_l = -\sum_{i=1}^{l} (a_i / S_l) \log(a_i / S_l)$.

(4) Entropy of R_{l+k}: $H_{l+k} = -\sum_{i=1}^{l} ((a_i + b_i)/S_{l+k}) \log((a_i + b_i)/S_{l+k}) - \sum_{i=l+1}^{l+k} (b_i / S_{l+k}) \log(b_i / S_{l+k})$.

We can see that, with intensity growing up, maximal intensity, area, and entropy all increase. Considering a simple case which often occurs, $b_i = 0$ $(i = 1, 2, \ldots, l)$, which means that the parent region R_{l+k} does not contain other subregions except R_l. Then there is a simple relationship between the average entropy:

$$\frac{H_{l+k}}{S_{l+k}} - \frac{\log S_{l+k}}{S_{l+k}} = \left(\frac{H_l}{S_l} - \frac{\log S_l}{S_l}\right) \times \left(\frac{S_l}{S_{l+k}}\right)^2 - \left(\frac{\sum_{i=l+1}^{l+k} b_i \log b_i}{S_{l+k}^2}\right). \tag{2.2}$$

It is easy to see that we can infer H_{l+k}/S_{l+k} from H_l/S_l in a simple way.

3. Experiments

In order to demonstrate the repeatability and time performance of our method when dealing with blurring, we compare it with the original MSER method, using open-source code provided at [18]. We use the 6 blurred images (bike set) provided by the popular Oxford dataset [19] for test. This blurred sequence is acquired by varying the camera focus [20, 21].

3.1. Repeatability Test

Here we show the results on the images for better comparison, as in Figure 2. We can see from Figure 2 that the main detection difficulties lie on the part where the intensity of the foreground is similar to the background, like the parts of the bike, and blurring makes

Table 1: Time comparison between the original MSER and our method.

Detection method	Computation time (seconds)
Original MSER	0.6476
Entropy-based MSER	1.3799

things worse. Results show that our entropy-based MSER could successfully detect this kind of regions, which will consequentially improve detection repeatability. Furthermore, the number of detected features is greatly increased, which is desirable in the matching and recognition stages.

3.2. Time Test

It is also useful to compare the time consumption between the original method and our method. Table 1 shows the time consumption on the above bike set on a 2.4 GHz Pentium 2 CPU. We can see that no significant extra time is needed in our method.

4. Conclusion

In this paper we have proposed an efficient extension to the maximally stable extremal regions (MSER) which makes it more robust to blurring. We utilize the fact that, in a blurred image, the entropy of probability distribution function of intensity values increases rapidly when the local region expands across the boundary, while the entropy in the central part keeps small, and use entropy averaged by region area as a measure to reestimate region missed by MSER. Experiments show that when dealing with blurred images, our method has better performance than the original MSER, with little extra computational effort.

Acknowledgments

This work has been funded by the National Key Technology R&D Program in the 11th Five Year Plan of China under the program of Research on Disabled People Function Rehabilitation Aids (Grant no. 2009BAI71B07), the Research Project of Department of Education of Zhejiang Province (no. Y201018160), and the State Scholarship Fund from China Scholarship Council (no. 2011833105).

References

[1] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[2] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, vol. 2010, Article ID 408418, 22 pages, 2010.

[3] S. Y. Chen, G. J. Luo, X. Li, S. M. Ji, and B. W. Zhang, "The specular exponent as a criterion for appearance quality assessment of pearl-like objects by artificial vision," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 8, pp. 3264–3272, 2012.

[4] X. Zhang, Y. Zhang, J. Zhang et al., "Unsupervised clustering for logo images using singular values region covariance matrices on Lie groups," *Optical Engineering*, vol. 51, no. 4, Article ID 047005, 8 pages, 2012.

[5] K. Mikolajczyk, T. Tuytelaars, C. Schmid et al., "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.
[6] J. Zhang, S. Y. Chen, S. Liu, and Q. Guan, "Normalized weighted shape context and its application in feature-based matching," *Optical Engineering*, vol. 47, no. 9, Article ID 097201, 2008.
[7] S. Y. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.
[8] P. E. Forssén and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV'07)*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.
[9] C. Cattani, S. Y. Chen, and G. Aldashev, "Information and modeling in complexity," *Mathematical Problems in Engineering*, vol. 2012, Article ID 868413, 4 pages, 2012.
[10] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
[11] M. Li, C. Cattani, and S. Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 13 pages, 2011.
[12] M. Perdoch, J. Matas, and S. Obdrzalek, "Stable affine frames on isophotes," in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV'07)*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.
[13] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.
[14] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007.
[15] D. Nistér H Stewénius, "Linear time maximally stable extremal regions," in *Proceedings of the Computer Vision-ECCV 2008*, vol. 5303, pp. 183–196, 2008.
[16] P. E. Forssén, "Maximally stable colour regions for recognition and matching," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.
[17] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.
[18] http://www.vlfeat.org/.
[19] http://www.robots.ox.ac.uk/~vgg/research/affine/index.html.
[20] S. Y. Chen and Y. F. Li, "Determination of stripe edge blurring for depth sensing," *IEEE Sensors Journal*, vol. 11, no. 2, pp. 389–390, 2011.
[21] H. Liu, S. Chen, and N. Kubota, "Special section on intelligent video systems and analytics," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, article 90, 2012.