

Mathematical Problems in Engineering

Theory, Methods, and Applications

Special Issue
Information and Modeling in Complexity

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

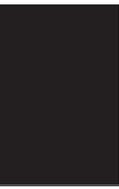


Information and Modeling in Complexity

Mathematical Problems in Engineering

Information and Modeling in Complexity

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



Copyright © 2012 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in "Mathematical Problems in Engineering." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

- E. M. Abdel-Rahman, Canada
Hamdy Nabih Agiza, Egypt
Sebastian Anita, Romania
W. Assawinchaichote, Thailand
K. V. Avramov, Ukraine
Erwei Bai, USA
Ezzat G. Bakhom, USA
J. M. Balthazar, Brazil
Shankar P. Bhattacharyya, USA
Stefano Boccaletti, Spain
Michael J. Brennan, UK
John Burns, USA
Piermarco Cannarsa, Italy
Carlo Cattani, Italy
Miguel Cerrolaza, Venezuela
Michael J. Chappell, UK
Kue-Hong Chen, Taiwan
Kui Fu Chen, China
Michele Chiumenti, Spain
Jyh Horng Chou, Taiwan
Slim Choura, Tunisia
Biswa N. Datta, USA
M. do R. de Pinho, Portugal
Joao B. R. Do Val, Brazil
Horst Ecker, Austria
Alex Elias-Zuniga, Mexico
Anders Eriksson, Sweden
Qi Fan, USA
Moez Feki, Tunisia
Ricardo Femat, Mexico
Rolf Findeisen, Germany
R. A. Fontes Valente, Portugal
Zoran Gajic, USA
Ugo Galvanetto, Italy
Furong Gao, Hong Kong
Oleg V. Gendelman, Israel
Paulo Batista Gonçalves, Brazil
Oded Gottlieb, Israel
Muhammad R. Hajj, USA
Thomas Hanne, Switzerland
Katica R. Hedrih, Serbia
M. I. Herreros, Spain
Wei-Chiang Hong, Taiwan
J. Horacek, Czech Republic
Gordon Huang, Canada
Anuar Ishak, Malaysia
Reza Jazar, Australia
Zhijian Ji, China
J. Jiang, China
J. J. Judice, Portugal
Tadeusz Kaczorek, Poland
Tamas Kalmar-Nagy, USA
Tomasz Kapitaniak, Poland
Hamid R. Karimi, Norway
Nikolaos Kazantzis, USA
Jurgen Kurths, Germany
Claude Lamarque, France
Stefano Lenci, Italy
Ming Li, China
Shanling Li, Canada
Teh-Lu Liao, Taiwan
P. Liatsis, UK
Jui-Sheng Lin, Taiwan
Wanquan Liu, Australia
Bin Liu, Australia
Angelo Luongo, Italy
Alexei Mailybaev, Brazil
R. Martinez-Guerra, Mexico
B. Maslowski, Czech Republic
Yuri V. Mikhlin, Ukraine
G. V. Milovanović, Serbia
Trung Nguyen Thoi, Vietnam
Hung Nguyen-Xuan, Vietnam
Ben T. Nohara, Japan
Anthony Nouy, France
Francesco Pellicano, Italy
F. Lobo Pereira, USA
A. Pogromsky, The Netherlands
Seppo Pohjolainen, Finland
Stanislav Potapenko, Canada
Sergio Preidikman, USA
Serge Prudhomme, USA
Dane Quinn, USA
K. R. Rajagopal, USA
Sivaguru Ravindran, USA
G. Rega, Italy
Pedro Ribeiro, Portugal
J. Rodellar, Spain
Rosana Rodríguez-López, Spain
A. J. Rodriguez-Luis, Spain
Rubén Ruiz García, Spain
Miguel A. F. Sanjuán, Spain
Ilmar Ferreira Santos, Denmark
Nickolas S. Sapidis, Greece
Andrey V. Savkin, Australia
Massimo Scalia, Italy
Alexander P. Seyranian, Russia
Tony Sheu, Taiwan
Zhan Shu, UK
Christos H. Skiadas, Greece
Davide Spinello, Canada
Sri Sridharan, USA
Jitao Sun, China
Xi-Ming Sun, China
Andrzej Swierniak, Poland
Allen Tannenbaum, USA
Cristian Toma, Romania
Irina N. Trendafilova, UK
Kuppapalle Vajravelu, USA
Victoria Vampa, Argentina
Blas M. Vinagre, Spain
Yongqi Wang, Germany
Moran Wang, USA
Gerhard-Wilhelm Weber, Turkey
Peter R. Wolenski, USA
Kwok W. Wong, HongKong
Christine Qiong Wu, Canada
Ligang Wu, China
Zheng-Guang Wu, China
Xuping Xu, USA
Xing-Gang Yan, UK
Jun-Juh Yan, Taiwan
Xing-Gang Yan, UK
Mahmoud T. Yassen, Egypt
Mohammad I. Younis, USA
Ion Zaballa, Spain
Xu Zhang, China

Contents

Information and Modeling in Complexity, Carlo Cattani, Shengyong Chen, and Gani Aldashev
Volume 2012, Article ID 868413, 3 pages

Extracting Fetal Electrocardiogram from Being Pregnancy Based on Nonlinear Projection,
Truong Quang Dang Khoa, Ho Huu Minh Tam, and Vo Van Toi
Volume 2012, Article ID 439084, 13 pages

Modeling Quantum Well Lasers, Dan Alexandru Anghel, Andreea Rodica Sterian,
and Paul E. Sterian
Volume 2012, Article ID 736529, 11 pages

Fisheye-Lens-Based Visual Sun Compass for Perception of Spatial Orientation, Liang Wang,
Fuqing Duan, and Ke Lv
Volume 2012, Article ID 460430, 15 pages

Multifractals Properties on the Near Infrared Spectroscopy of Human Brain Hemodynamic,
Truong Quang Dang Khoa and Vo Van Toi
Volume 2012, Article ID 670761, 12 pages

Adaptive Control and Synchronization of the Shallow Water Model, P. Sangapate
Volume 2012, Article ID 529251, 9 pages

Photovoltaic Greenhouses: Comparison of Optical and Thermal Behaviour for Energy Savings,
Maurizio Carlini, Tommaso Honorati, and Sonia Castellucci
Volume 2012, Article ID 743764, 10 pages

Distinguishing Stationary/Nonstationary Scaling Processes Using Wavelet Tsallis q -Entropies,
Julio Ramirez Pacheco, Deni Torres Román, and Homero Toral Cruz
Volume 2012, Article ID 867042, 18 pages

Grouping Optimization Based on Social Relationships, Rong-Chang Chen
Volume 2012, Article ID 170563, 19 pages

**Synchronization of Complex Networks with Time-Varying Delayed Dynamical Nodes via
Pinning Control**, Shuguo Wang, Hongxing Yao, and Qiuxiang Bian
Volume 2012, Article ID 654742, 13 pages

Parameterization Method on B-Spline Curve, H. Haron, A. Rehman, D. I. S. Adi, S. P. Lim,
and T. Saba
Volume 2012, Article ID 640472, 22 pages

Entropy and Multifractality for the Myeloma Multiple TET 2 Gene, Carlo Cattani,
Gaetano Pierro, and Giuseppe Altieri
Volume 2012, Article ID 193761, 14 pages

A Simple Framework for Face Photo-Sketch Synthesis, Xuwei Li and Xiaochun Cao
Volume 2012, Article ID 910719, 19 pages

Key Issues in Modeling of Complex 3D Structures from Video Sequences, Shengyong Chen, Yuehui Wang, and Carlo Cattani
Volume 2012, Article ID 856523, 17 pages

Advanced Signal Processing and Command Synthesis for Memory-Limited Complex Systems, Cristian Toma
Volume 2012, Article ID 927821, 13 pages

A Method for Designing Assembly Tolerance Networks of Mechanical Assemblies, Yi Zhang, Zongbin Li, Jianmin Gao, Jun Hong, Francesco Villecco, and Yunlong Li
Volume 2012, Article ID 513958, 26 pages

The Inverse Problem for a General Class of Multidimensional Hyperbolic Equations, Gani Aldashev and Serik A. Aldashev
Volume 2012, Article ID 891078, 11 pages

Noise Estimation for Single-Slice Sinogram of Low-Dose X-Ray Computed Tomography Using Homogenous Patch, Zhiwu Liao, Shaoxiang Hu, Ming Li, and Wufan Chen
Volume 2012, Article ID 696212, 16 pages

Emergent Behaviors in Social Networks of Adaptive Agents, Florin Leon
Volume 2012, Article ID 857512, 19 pages

Traffic Dynamics on Complex Networks: A Survey, Shengyong Chen, Wei Huang, Carlo Cattani, and Giuseppe Altieri
Volume 2012, Article ID 732698, 23 pages

A Model to Partly but Reliably Distinguish DDOS Flood Traffic from Aggregated One, Ming Li and Wei Zhao
Volume 2012, Article ID 860569, 12 pages

Haar-Wavelet-Based Just Noticeable Distortion Model for Transparent Watermark, Lu-Ting Ko, Jwu-E Chen, Hsi-Chin Hsin, Yaw-Shih Shieh, and Tze-Yun Sung
Volume 2012, Article ID 635738, 14 pages

A Unified Algorithm for Subband-Based Discrete Cosine Transform, Lu-Ting Ko, Jwu-E Chen, Hsi-Chin Hsin, Yaw-Shih Shieh, and Tze-Yun Sung
Volume 2012, Article ID 912194, 31 pages

A Novel SFP Extracting/Tracking Framework for Rigid Microstructure Measurement, Sheng Liu, Ting Fang, and Zichen Chen
Volume 2012, Article ID 356454, 11 pages

Markov Models for Image Labeling, S. Y. Chen, Hanyang Tong, and Carlo Cattani
Volume 2012, Article ID 814356, 18 pages

Numerov's Method for a Class of Nonlinear Multipoint Boundary Value Problems, Yuan-Ming Wang, Wen-Jia Wu, and Massimo Scalia
Volume 2012, Article ID 316852, 29 pages

An Efficient Approach to Solve the Large-Scale Semidefinite Programming Problems,
Yongbin Zheng, Yuzhuang Yan, Sheng Liu, Xinsheng Huang, and Wanying Xu
Volume 2012, Article ID 764760, 12 pages

Visiting Power Laws in Cyber-Physical Networking Systems, Ming Li and Wei Zhao
Volume 2012, Article ID 302786, 13 pages

Mixed Signature: An Invariant Descriptor for 3D Motion Trajectory Perception and Recognition, Jianyu Yang, Y. F. Li, Keyi Wang, Yuan Wu, Giuseppe Altieri, and Massimo Scalia
Volume 2012, Article ID 613939, 29 pages

Radial Basis Functional Link Network and Hamilton Jacobi Issacs for Force/Position Control in Robotic Manipulation, Shuhuan Wen, Junying Yuan, and Jinghai Zhu
Volume 2012, Article ID 568796, 10 pages

Incorporating a Local Binary Fitting Model into a Maximum Regional Difference Model for Extracting Microscopic Information under Complex Conditions, Chunyan Yao, Jianwei Zhang, Min Chen, Qiu Guan, and Massimo Scalia
Volume 2012, Article ID 474938, 10 pages

Editorial

Information and Modeling in Complexity

Carlo Cattani,¹ Shengyong Chen,² and Gani Aldashev³

¹ *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

² *Department of Computer Science, Zhejiang University of Technology, 310023 Hangzhou, 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 13 February 2012; Accepted 13 February 2012

Copyright © 2012 Carlo Cattani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Overview of the Issue

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

The topics that deal with these issues can be divided into two large 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.

The complexity theory is the analysis of complex systems, which can be approached from many different points of view. Mathematically it refers to nonlinear dynamical systems which are very sensitive to initial conditions, in a such way that a small perturbation of these could have unpredictable consequences on the evolution equations. Moreover the future behaviour or long-term prediction of the system cannot be fully determined, so that the uniqueness of the solution fails. Complex systems were also related to chaotic systems and to chaos dynamics. Common sense reasoning might lead one to believe that complexity and chaos should be considered as the same characteristic of the state of disorder. However, these two concepts should not be confused, since they refer to different questions. A chaotic system

is a disordered state which follows from the strong dependence on initial conditions, from the existence of some attractors and dense periodic orbits. The strong dependence on initial-boundary conditions means that a small variation thereof, from one initial state to another very close, could lead to some different future orbits. In other words, the time evolution for two very close states is divergent in future. Sometime a state of disorder can be also considered as a random state. In other words, one can hardly understand from observed experimental data if they come from the evolution of a chaotic system or if they are simply a collection of random values. Frequency analysis, wavelet analysis, power spectrum method, information theory, entropy, and fractal dimension are expedient tools which enable us to analyze data trying to single out some deterministic time series by removing what can be considered a pure random process.

A key issue in all fields of complexity is the signal extraction and processing. In fact, signals extract such information from the complex phenomena being measured; therefore, the signal analysis can be used to study the complex phenomena. However, a main difficulty comes from the fact that signals are typically some time series having both a regular and a random component. Separating these two components is the essence of signal extraction problem.

Randomness, multifractality and chaos are also epitomizing features of complexity. A complex system can be also considered as a chaotic system, in the sense that it can be sensitive to initial conditions, that its evolution could be also disturbed by some random noise and that it shows some noninteger dimensions typical of fractals. However, a complex system is characterized by some more interesting features, such as a multiparametric dependence or self-organizing multiscale activity. Indeed, being a multidisciplinary topic, complexity does not have a unique definition. We can roughly define complexity as one of the following three:

- (1) scale transition from a micro- to macroscale, such that a disordered set of individuals behaves as sole entity;
- (2) the analysis of a unique entity resulting from the interaction of many parts;
- (3) transition from disorder to organization.

It is clear that a special feature of complex system is the organization or self-organization arising from a chaotic-random state. However, there are still some open debates on a clear definition of organization. We can assume for organization the meaning of predictable, in the sense that there exists a dynamical system able to describe the evolution of the complex system when it is organized. The transition from disorder to organization is still unclear. Many models propose the emergence of order as the scale transition from the microscale, where each individual is characterized by its own activity, to a macroscale where more individuals agree on a unique strategy, which could be different from the usual activity at the microscale.

Self-organizing complex systems are often/essentially maintained by information flows, thus implying the quantification of information. Fundamental tools of signal processing have been recently developed from the elementary topics of compression data, storing and communicating data into more broad applications of information theory such as neurobiology, Internet networks, language processing, quantum computing, data mining, and many more. Any information flow is characterized by a measure, originally given by Shannon, related to uncertainty and disorder in the flow. Shannon entropy is the most popular tool to give a measure of the uncertainty in predicting the evolution of a stochastic variable. More outcomes have a variable, the higher is the entropy, so that this measure can

be also taken as the measure of randomness: the higher is the entropy, the more random is the variable evolution. Many alternative definitions of the Shannon entropy have been recently proposed starting from the Kolmogorov-Sinai entropy, Rényi entropy, topological entropy, and so forth.

The theory of complexity applied to advanced engineering systems requires subtle and efficient mathematical models obtained by developing classical tools for complex systems. The increasing demand for engineering systems able to handle complex phenomena is nowadays pushing scholar's investigations in all directions of knowledge, such as medical biological and visual systems, environment, optimisation, structural complexity, advanced composite material, vibrations, social science, economy, Internet architecture, and processing in networks.

The analysis of general principles that lie behind these problems in so many different fields, transversally across different specific applications is the area of highly active research. 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.

From the current states and trends in the issue and related topics of information and modelling in complexity, although some solutions and models become available, most problems remains open and research is highly active in this field. In the near future, we expect many more contributions that will address all of the key aspects raised above, in particular, the development of general mathematical methods and models, analyses of practical signal-extraction problems, and applications in the study of self-organizing problems in social sciences.

Carlo Cattani
Shengyong Chen
Gani Aldashev

Research Article

Extracting Fetal Electrocardiogram from Being Pregnancy Based on Nonlinear Projection

Truong Quang Dang Khoa,¹ Ho Huu Minh Tam,² and Vo Van Toi¹

¹ Biomedical Engineering Department, Ho Chi Minh City International University and Vietnam National University-Ho Chi Minh City, Vietnam

² University of Technology-Ho Chi Minh City, Vietnam

Correspondence should be addressed to Truong Quang Dang Khoa, khoa@ieee.org

Received 13 September 2011; Accepted 19 December 2011

Academic Editor: Carlo Cattani

Copyright © 2012 Truong Quang Dang Khoa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fetal heart rate extraction from the abdominal ECG is of great importance due to the information that carries in assessing appropriately the fetus well-being during pregnancy. In this paper, we describe a method to suppress the maternal signal and noise contamination to discover the fetal signal in a single-lead fetal ECG recordings. We use a locally linear phase space projection technique which has been used for noise reduction in deterministically chaotic signals. Henceforth, this method is capable of extracting fetal signal even when noise and fetal component are of comparable amplitude. The result is much better if the noise is much smaller (*P* wave and *T* wave can be discovered).

1. Introduction

Since the early work of Cremer in 1906, various methods for fetal monitoring have been proposed to obtain information about the heart status. The cardiac electrical activity of a fetus can be recorded noninvasively from electrodes on the mother's body surface. Such recordings of fetal electrocardiograms (ECGs) are complicated by the existence of the mother's ECG and effectively random contaminations due to noncardiac sources. Furthermore, the fetal signal is rather small due to the size of the fetal heart and the intervening tissue. We have to separate the fetal ECG (FECCG) from the maternal trace and from the other contaminations. The ECGs is the tool for the clinical diagnostic because the nonlinear chemiological excitation of cardiac tissue and signal show both fluctuation and remarkable structure. Moreover, because the length of cardiac cycle which is measured by the distance of two successive QRS spike fluctuates with the predictable component, ECGs is deterministic chaotic. Besides, ECGs also comprise the excitation of the mother, the distortion of tissues on the transmission.

In general, linear filter is used to separate signals based on their difference in frequency domain which can be expressed in terms of Fourier spectrum. However, even the optimal linear filter, the Wiener filter, cannot be successful in this case because ECGs of the pregnant include both maternal and fetal signals which have the same spectral contents, and the noise coming from the electric equipments has a broad band and random [1].

Another solution is nonlinear filter which offers some superior features to linear projection in this case. However, because the method is based on theory of deterministic dynamical system, we must make sure that the observed data contents the typical properties of deterministic chaotic signals. We can find it true with ECGs by seeing that the maximal Lyapunov exponent is positive [2].

In nonlinear filter, though the fetal signal and maternal signal are similar in shape and spectral contents, we can separate the components by a very natural way: the magnitude and the heart beat. In fact, because the fetal heart is much smaller than the maternal heart, the fetal signal is much smaller than the maternal signal. Generally, the heart beat of the fetus is about one-third of the mother. In fact, our method is used for noise reduction, we just consider fetal signal as a contaminated noise.

Up to now, in order to deal with this problem, many works have been done and have given satisfactory results: "wavelet transform" was applied to extract wavelet-based features of fetal signal [3], "blind source separation (BSS)" was used to separate a set of source signal from numerous observed signals [4], "source extraction" is quite related to BSS except using additionally prior information about FECCG [5], and some other method such as matched filtering [6], dynamic neural network [7], adaptive neurofuzzy inference systems [8], fuzzy logic [9], frequency tracking [10], and polynomial networks [11].

The technique we apply below is actually based on phase space reconstruction. The posttransient trajectory of the system is frequently confined to a set of points in state space called an "attractor" [12]. By using the delay coordinates, attractor is then empirically found to be constrained to a low-dimension manifold [13]. Hence, by estimating the attractor, noise can be reduced by projecting onto it. Whenever a multidimensional reconstruction of a signal can be approximated by a low-dimensional surface (or attractor), a projection onto this surface can improve the signal-to-noise ratio. In the present application, the fetal component is first treated as a contamination of the maternal ECG, whence noise reduction techniques are suitable for signal separation.

On the other hand, the extracted FECCG recorded in the form of the nonstress test (NST) by using cardiotocography (CTG) was analyzed by wavelets to monitor fetal well-being [14].

2. Method

It has been proved that if a system is controlled by an attractor, we can find out the dynamics of the full system just by single variables (theorem of Takens). In this paper, we use the delay coordinates of m dimension. The geometry of a state space trajectory or a shape of attractor can be obtained by using delay coordinates to construct vectors valued time from a single-channel observation ($\vec{s}_n, n = m^*t, \dots, N$):

$$\vec{s}_n = (s_{n-(m-1)t}, s_{n-(m-2)t}, \dots, s_n). \quad (2.1)$$

Here time n is measured in the sampling intervals, t is called delay or lag, and m is the embedded dimension. All of the vectors are then inserted into a matrix, in this way we

can draw delay plots: a plot of s_n versus a delay itself is able to reveal the characteristic of the attractor. According to Taken's theorem, under general conditions, if the embedded dimension m is large enough, the local topology of the attractor is preserved. This process is called a delay reconstruction in m dimension (see Figure 1).

For noise reduction, according to Schreiber's algorithm [12], we have to move all points \vec{s}_n of the phase space to the attractor manifold. First, we have to find the nearest points to \vec{s}_n , called neighbors, within the radius ε and the set of these points called U_n . In addition, the number of neighbor $|U_n|$ should not be lower than a specified number which is usually 50 [15]. From here, we compute the mean:

$$\langle \vec{s} \rangle^{(n)} = \frac{1}{|U_n|} \sum_{k \in U_n} \vec{s}_k. \quad (2.2)$$

Then we compute the covariance matrix:

$$C_{ij}^{(n)} = \sum_{\vec{s}_n \in U_n} \left[R \left(\vec{s}_n - \langle \vec{s} \rangle^{(n)} \right) \right]_i \left[R \left(\vec{s}_n - \langle \vec{s} \rangle^{(n)} \right) \right]_j. \quad (2.3)$$

R is a weight matrix R which is chosen to be diagonal with R_{11} and R_{mm} large (about 1000) and all other diagonal entries $R_{ii} = 1$. This makes the two largest eigenvalue of $C^{(n)}$ lying in subspace spanned by the first and the last coordinates of embedded space and prevents the correction vector from having any component in these direction.

Particularly, when using MATLAB, instead of using the for loop to compute covariance matrix, the method we use is a little bit different. First, we create a "deviation" matrix $D^{(n)}$ whose each column is the "deviation" vector $\vec{\delta}_j^{(n)} = R * (\vec{s}_n - \vec{s}_j)$ ($\vec{s}_j \in U_n$). Then, the covariance matrix in equation is computed simply by $C^{(n)} = \text{transpose}(D^{(n)}) * D^{(n)}$. Empirically, this makes the result more accurate and much faster than using for loop.

In the next step, we determine the orthonormal eigenvector \vec{c}_q and the eigenvalue of $C^{(n)}$. The eigenvector of $C^{(n)}$ represent the semiaxes of the ellipsoid best approximating the cloud of neighbor points U_n . Ideally, the largest eigenvalues of the covariance matrix span the attractor manifold and the lower span the others. Projecting vector onto the subspace spanned by the largest eigenvalue will move it closer to the attractor manifold thereby creating a more accurate approximation of the true dynamics of the system, because the contaminating noise and the fetal signal span in another subspace.

$$\hat{\vec{s}} = \vec{s}_n - R^{-1} \sum_{q=1}^Q \vec{c}_q \left[\vec{c}_q \cdot R \left(\vec{s}_n - \langle \vec{s} \rangle^{(n)} \right) \right], \quad (2.4)$$

(where Q is the number of dimensions of the manifold that will be locally approximate by Q eigenvector corresponding to the largest eigenvalues).

When the projection is finished for all the points, we will get m corrected vector because each element in scalar time series exists in m vector. Therefore, we just average them all; this will not project the vector exactly to the manifold but will still move it closer to the manifold.

In this algorithm, there are three important parameters: the embedded window $(m-1)t$ which is used to select components by time scale, the radius of cloud of neighborhood

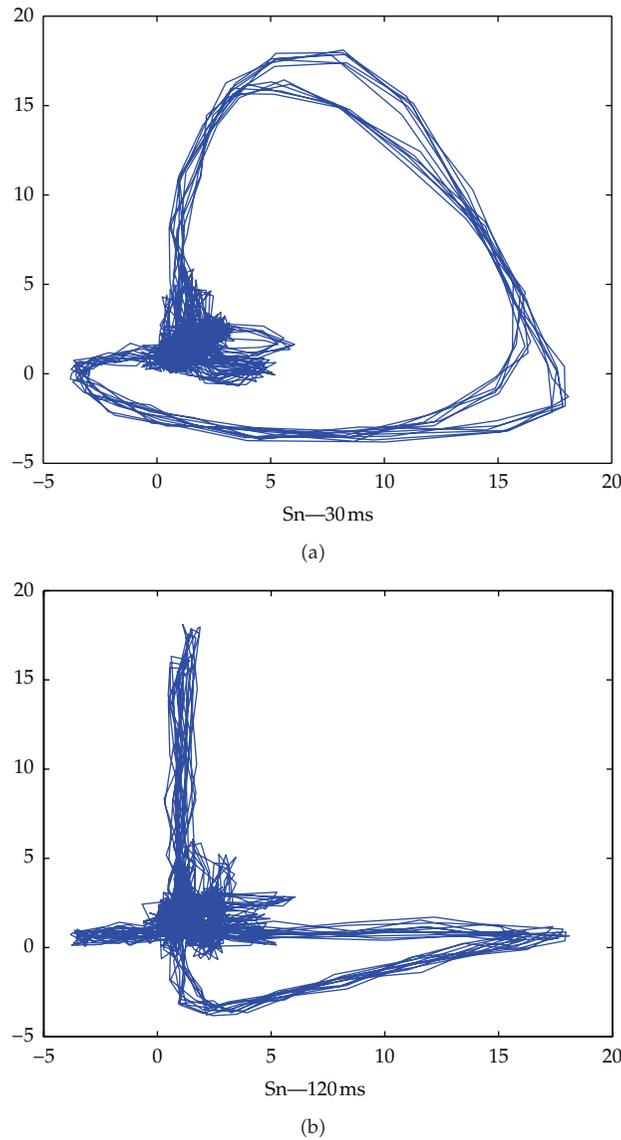


Figure 1: The delay plots of ECG. (a) The delay time is 30 ms; the large resolves the ventricular QRS complex while the smaller features in the center the atria *P* wave. (b) The bigger value of lag in comparison to the maternal time scale, we see the huge decrease in resolution of the QRS complex.

points which is used to select components by magnitude, and the number of dimension of the manifold Q .

In order to choose an optimal m , many studies have been done. According to [2], m should be large for two reasons. First, the larger m is, the more deterministic signs are presented in the dataset. In fact, due to the fluctuating of body condition such as the respiration, the biological signal such as ECG becomes nonstationary and that violates the condition to apply the locally projective noise reduction. A solution is to increase m . Instead of using $m > 2D$ in Taken's theory, $m > 2(D+P)$ is used where P is the number of nonstationary

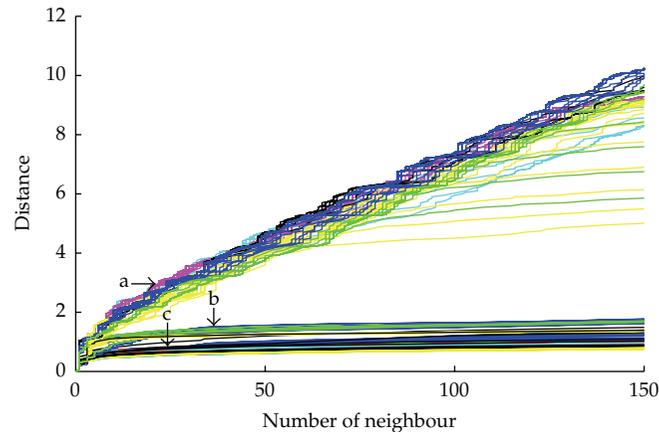


Figure 2: Number of neighbor versus distance.

parameters [16]. Although, mathematically, P must form a stationary chaotic system, this still works well if the nonstationary component varies slowly and rarely has any sudden change. The other reason is that with the large m , the selectivity or appropriate neighbor is increased because we are using max norm to define the distance between neighbors. For more precise value of m , the technique of Kennel and Abarbanel [17] is often used, which examines whether points that are near neighbors in this dimension are still near neighbor in the next dimension. Empirically, the larger m is, the fewer false neighbors and the more fully the image unfolded.

In alliance with m , choosing t is a very important factor. A suitable t should fulfill two criteria. First, t must be large enough compared to the time scale or the two successive \vec{s}_n or they will be strongly correlated, that is, the value of the time series at i must be significantly different from its value at $i + t$. Therefore, we can gather enough information to successfully reconstruct all whole phase space with the reasonable choice of m . Another criteria is that t must not be too large than the time scale of the system, in which the two successive \vec{s}_n will be independent and uncorrelated [18, 19], and the system will lose memory of its initial state. According to [18], it has been showed that the optimal value of t is typically around $1/10-1/2$ of the mean orbital period of the attractor. Often value of t is around the correlation time.

Finally, ε should be larger than the noise amplitude and the fetal amplitude (when the fetal signal is considered as a contaminated noise), but still small enough not to average out the curvature radius of the time series to reserve the manifold shape. However, if we find the neighbors for all points in phase space by using fixed ε will not give a good result. In this paper, we use a graph between the number of neighbor versus distance to determine ε for each point in phase space (see Figure 2 for relation between ε and distance).

In the graph, each curve represent one point in phase space. Clearly, we can see that it is separated into 3 parts. Part (a) is those points which lay on the peaks of ECG, part (b) is those points which lay on the peaks of FECG, part (c) is other points. Thus, basing on the slope and the height of the curve, we can apply this simple rule for better filter: because our aim is to calculate the fetal heart beat, ε should be sufficiently large for the (b) part, and for avoiding distortion, ε should be fair small for the (a) and (c) parts. However, at the (b) part, there is, in fact, some points that does not lay on the FECG's peak. The number of those points is few but the reason for this phenomenon is unknown. Fortunately, in some dataset, we find

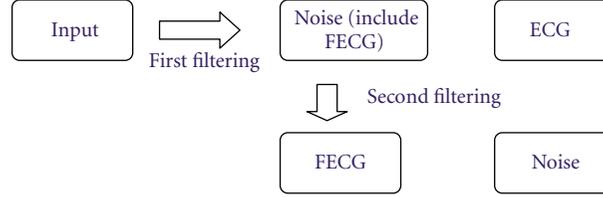


Figure 3: The processing diagram.

out that its slope is larger than the points lying on the FECG's peak when the number of neighbor grow sufficiently large (about > 150). Note that this method is primarily used in the first filtering for we need to avoid the large distortion at this point (large error will severely affect the second filtering), while in the second filtering, we will come back with the fix ε to suppress the noise better. In further research, a better rule or algorithm for the neighbor finding procedure may solve this problem.

Besides, we approximate the attractor as a collection of locally linear manifolds. For example, the loop can be approximated by a collection of short line segments; in this case the approximating manifold is one dimensional. When the embedding dimension m is larger than two, it can be appropriate to select locally linear manifold with dimension Q where $1 \leq Q < m$. For instance $Q = 2$, the manifold is locally planar.

In order to acquire FECG, in this paper, we will follow this strategy. First step, using Schreiber's algorithm for the input data to extract to clear ECG, without noise and fetal signal. Second, subtracting the input data with the clear ECG to acquire the secondary input (including noise and fetal signal). The last step, using Schreiber's algorithm again for the secondary input data to extract the clear FECG. The first and the last step should be repeated 2-3 times to get a better result. In sum, there are a lot of tradeoffs in choosing parameters and times of iteration, so that we have do a careful visual inspection of time series and few test runs to find the optimal results (see Figure 3).

3. Result

3.1. Artificial Signal

At first, we generate an artificial ECG by repeating one clear heart cycle of mother (the sample rate is 4 ms), plus the fake fetal signal created by scale artificial mother signal and then plus random noise following the Gaussian distribution (see Figure 4).

(a) Clean FECG RMS: Noise RMS = 1 : 1 (noise ratio = 1 : 1).

Following the following strategy, at first, to extract the clear ECG, we form an embedded window of 200 ms (equal 1/3 heart cycle of the mother) with $m = 51$ and $t = 1$ and using dynamic ε . Q around 2 is enough to reconstruct the phase space. The noise reduction ratio $R = 1.426$ (see Figure 5).

After that, the secondary input (using dynamic ε) is processed again with the embedded window ($m = 61, t = 1$). With Q of about 3, we will get the noise reduction ratio $R = 1.265$, and see that P wave and T wave are distorted. However, we just need to calculate the fetal heart beat so that the result is good enough (see Figure 6).

(b) Clean FECG RMS: Noise RMS = 1 : 5 (noise ratio = 1 : 5).

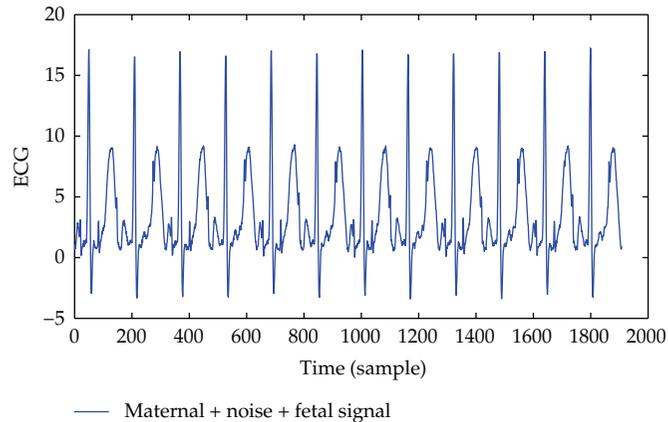


Figure 4: The artificial ECG.

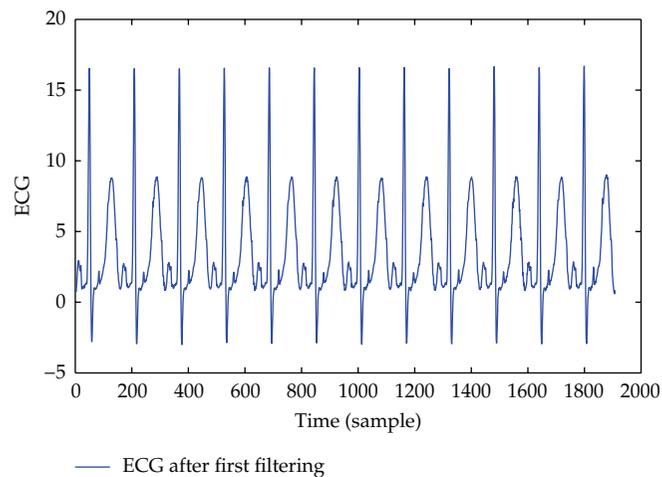


Figure 5: The artificial ECG after the first filtering.

The first filtering (using dynamic ε) is done with $m = 71$ and $t = 1$, we will get the noise reduction equal to 2.58. However, the noise ratio is too large so that none of the characteristic of the FECCG is reserved, thus the FECCG extraction failed (see Figure 7).

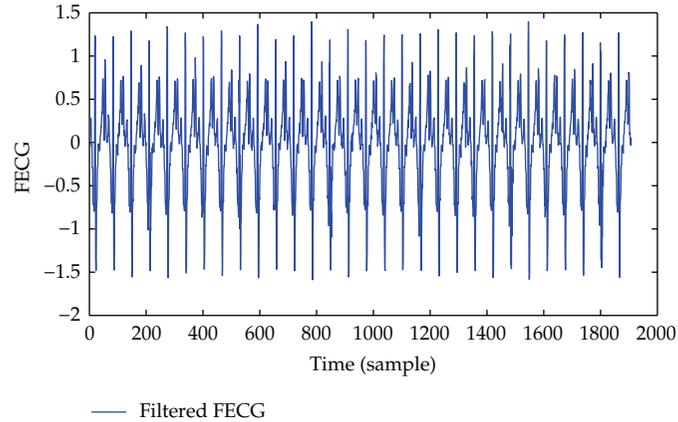
3.2. In the Real World

The entire following sample is taken from MIT website [20].

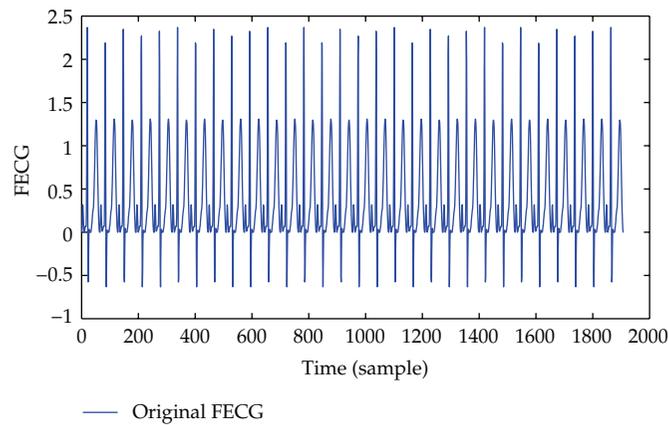
(1) ECG 23rd Weeks

Because the fetus has grown a lot, the magnitude of the fetal signal is quite big so that its heart beat can be seen virtually.

The first filtering is quite good despite the distortion at the T wave. For this result, we use $m = 31$, $t = 1$, dynamic ε . The manifold dimension Q is 2 (see Figure 8).



(a)

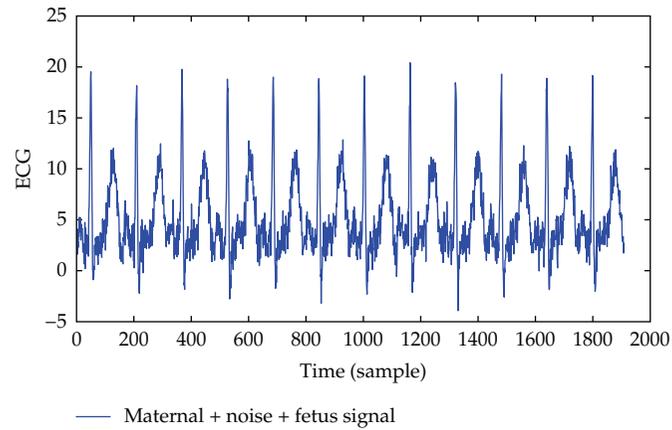


(b)

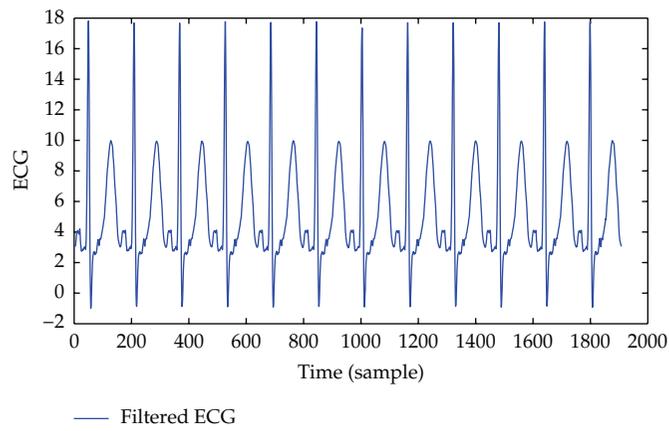
Figure 6: (a) The artificial FECG after the second filtering with the secondary input is the result of subtracting the artificial ECG to the filtered itself. (b) The true FECG.

For extracting the fetal signal, the next filtering is done with $m = 81, t = 1$, using fix $\varepsilon = 1.5$. The manifold dimension Q is 15. The extracted FECG lacks many details of P wave and T wave, its shape is distorted a lot. Due to the nonstationary characteristic of the real ECGs, the neighbor versus distance graph becomes somehow unstable (some points do not lay on the FECG's peak also have a slope as steep as the ones on the FECG's peak), this makes it quite difficult to extract the FECG at the "nonstationary" section (using dynamic ε) (see Figure 9).

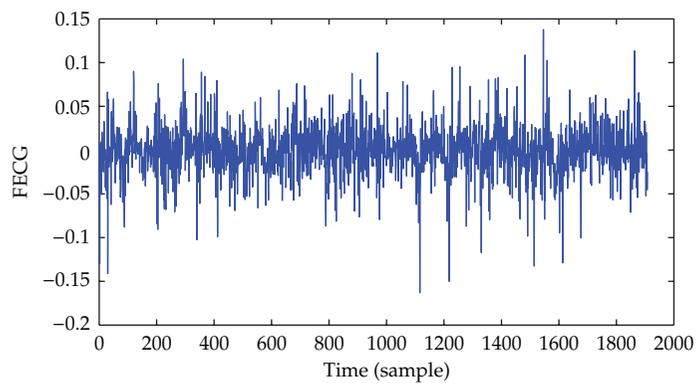
For comparison, we add here a result of Martín-Clemente et al. [21] which uses fast ICA. Clearly, the applied sample is more stationary than ours and its amplitude is a little bit larger than either of them may be in the same weeks. Hence, the result is very clear while ours is still interfered by some unwanted harmonics as Figure 7. Though the main purpose is to measure only the fetal heart beat, our algorithm still meets the requirement with just one channel.



(a)



(b)



(c)

Figure 7: (a) The artificial ECG. (b) ECG after the first filtering. (c) The extracted FECCG, nothing is revealed here cause the FECCG extraction failed.

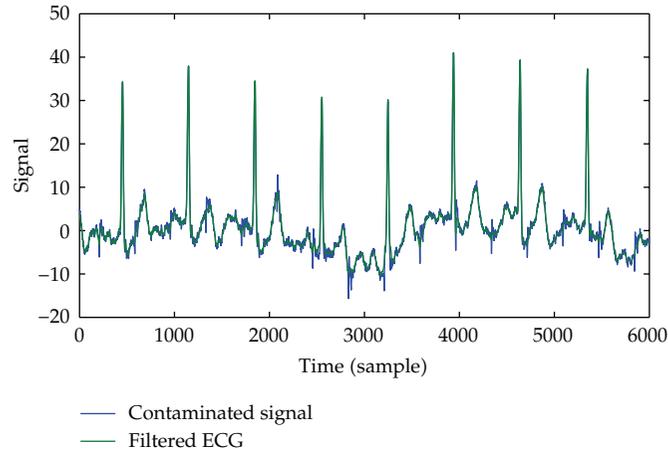


Figure 8: Blue line is the measured data. Green line is the clear ECG.

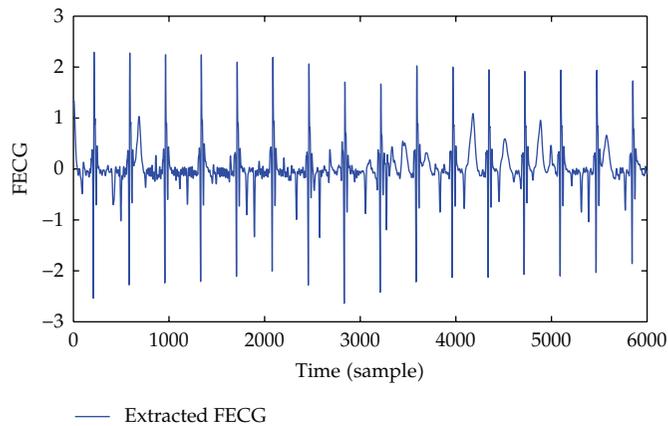


Figure 9: The extracted FECC, although it cannot be used for diagnosis, it is still good to calculate the heart rate.

(2) ECG 25th Weeks

In this case, the fetal signal is very small; the noise ratio is large so that any effort to recover FECC fails because our algorithm will average both noise and FECC. This phenomenon happens because at this time, the fetus will create a membrane to cover itself, thus the fetal heart beat signal received at the probe will be greatly reduced (see Figure 10).

(3) ECG 38th Weeks

At this time, the fetus has grown a lot, its heart beat is bigger as well. Thus the noise ratio is lower, then we can calculate its heart beat once again.

The first filtering is done with $m = 41, t = 1, Q = 2$, using dynamic ε . Then the second filtering is done with $m = 111, t = 1, Q = 2$, using fix $\varepsilon = 0.75$ (see Figure 11).

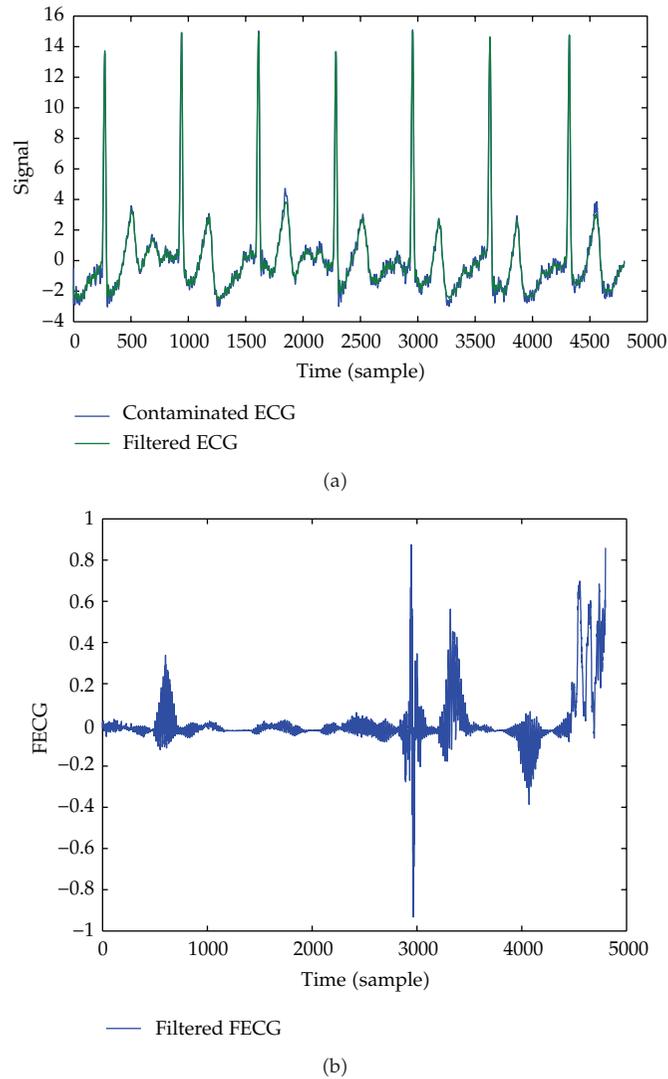


Figure 10: (a) Almost blue line is the measured data. Red line is the clear ECG. (b) All peaks of the FECC have been disappeared. This result is totally useless.

4. Conclusion

According to all results we got, this algorithm can be mainly used for counting the heart beat because many details such as P wave and T wave are distorted, few peaks of FECC disappear sometimes. on the other hand, if we want a better result, the diagnosis will be done after the 38th week or before the 25th week. This is a physiological characteristic that was reconfirmed by our results.

However, these results still show the dominance over the linear method which is based on frequency domain to separate signal. In our method, we define the difference between ECG and FECC in a very natural way: amplitude and time scale. The natural language to implement such filtering procedures is in terms of the geometry in a reconstructed state space.

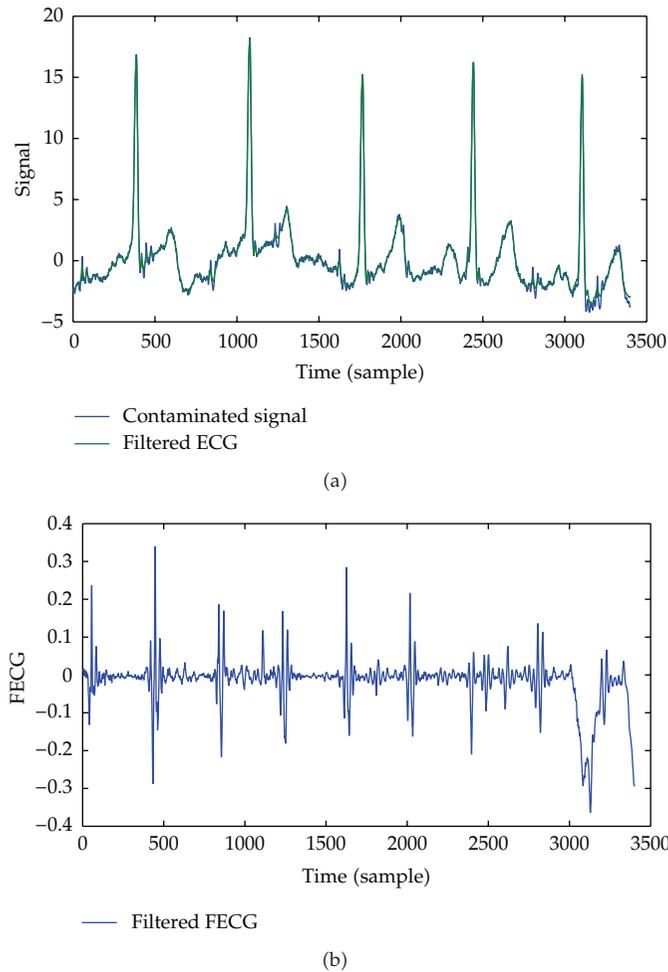


Figure 11: (a) Green line is the measured data. (b) The extracted FECG, due to the larger noise ratio (compare to ECG 23rd weeks), the result is not as good but it is good for heart beat counting.

Here, we have estimated the geometry of the maternal ECG using projections onto locally linear surfaces, which proves effective at filtering FECG from the measured signal.

In further works, some automatic information extraction method may be applied to test the efficiency of this method in reality as well as the development of a better algorithm or the research for the optimal parameters will be carried out to reserve as much information as possible for further diagnosis, manually or automatically.

Acknowledgments

This paper was partly supported research grants from Vietnam National University-Ho Chi Minh City, -Ho Chi Minh City International University, and Vietnam National Foundation for Science and Technology Development—NAFOSTED research Grant no. 106.99-2010.11.

References

- [1] T. Schreiber and D. T. Kaplan, "Signal separation by nonlinear projections: the fetal electrocardiogram," *Physical Review E*, vol. 53, no. 5, pp. R4326–R4329, 1996.
- [2] M. Perc, "Nonlinear time series analysis of the human electrocardiogram," *European Journal of Physics*, vol. 26, no. 5, pp. 757–768, 2005.
- [3] E. C. Karvounis, C. Papaloukas, D. I. Fotiadis, and L. K. Michails, "Fetal heart rate extraction from composite maternal ECG using complex continuous wavelet transform," in *Proceedings of Computers in Cardiology*, pp. 737–740, September 2004.
- [4] D. Graupe, Y. Zhong, and M. H. Graupe, "Extraction of fetal from maternal ECG early in pregnancy," *International Journal of Bioelectromagnetism*, vol. 7, no. 1, 3 pages, 2005.
- [5] Z. L. Zhang and Z. Yi, "Extraction of temporally correlated sources with its application to non-invasive fetal electrocardiogram extraction," *Neurocomputing*, vol. 69, no. 7–9, pp. 894–899, 2006.
- [6] J. F. Piéri, J. A. Crowe, B. R. Hayes-Gill, C. J. Spencer, K. Bhogal, and D. K. James, "Compact long-term recorder for the transabdominal foetal and maternal electrocardiogram," *Medical and Biological Engineering and Computing*, vol. 39, no. 1, pp. 118–125, 2001.
- [7] G. Camps-Valls, M. Martínez-Sober, E. Soria-Olivas, R. Magdalena-Benedito, J. Calpe-Maravilla, and J. Guerrero-Martínez, "Foetal ECG recovery using dynamic neural networks," *Artificial Intelligence in Medicine*, vol. 31, no. 3, pp. 197–209, 2004.
- [8] K. Assaleh and H. Al-Nashash, "A novel technique for the extraction of fetal ECG using polynomial networks," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 6, pp. 1148–1152, 2005.
- [9] K. A. K. Azad, "Fetal QRS complex detection from abdominal ECG: a fuzzy approach," in *Proceedings of the IEEE Nordic Signal Processing Symposium (NORSIG '00)*, p. 275, 2000.
- [10] A. K. Barros and A. Cichocki, "Extraction of specific signals with temporal structure," *Neural Computation*, vol. 13, no. 9, pp. 1995–2003, 2001.
- [11] K. Assaleh, "Extraction of fetal electrocardiogram using adaptive neuro-fuzzy inference systems," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 1, pp. 59–68, 2007.
- [12] T. Schreiber, M. Richter, and D. T. Kaplan, "Fetal ECG extraction with nonlinear state-space projections," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 1, pp. 133–137, 1998.
- [13] H. Kantz and T. Schreiber, "Nonlinear projective filtering I: background in chaos theory," in *Proceedings of the International Symposium on Nonlinear Theory and its Applications (NOLTA '98)*, Presses Polytechniques et Universitaires Romandes, Lausanne, Switzerland, 1998, chaos-dyn/9805024.
- [14] C. Cattani, O. Doubrovina, S. Rogosin, S. L. Voskresensky, and E. Zelianko, "On the creation of a new diagnostic model for fetal well-being on the base of wavelet analysis of cardiotocograms," *Journal of Medical Systems*, vol. 30, no. 6, pp. 489–494, 2006.
- [15] T. Schreiber and D. T. Kaplan, "Nonlinear noise reduction for electrocardiograms," *Chaos*, vol. 6, no. 1, pp. 87–92, 1996.
- [16] R. Hegger, H. Kantz, L. Matassini, and T. Schreiber, "Copy with nonstationary by over-embedding," *Physical Review Letters*, vol. 84, p. 4092, 2000.
- [17] M. B. Kennel and H. D. I. Abarbanel, "False neighbors and false strands: a reliable minimum embedding dimension algorithm," *Physical Review E*, vol. 66, no. 2, Article ID 026209, 2002.
- [18] S. H. Strogatz, *Nonlinear Dynamic and Chaos: With Application to Physics, Biology, Chemistry and Engineering*, Westview Press, 1994.
- [19] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, 2nd edition, 2004.
- [20] MIT-BIH Polysomnographic Database, Biomedical Engineering Centre, MIT, UK, <http://www.physionet.org/physiobank/database/>.
- [21] R. Martín-Clemente, J. L. Camargo-Olivares, S. Hornillo-Mellado, M. Elena, and I. Román, "Fast technique for noninvasive fetal ECG extraction," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 227–230, 2011.

Research Article

Modeling Quantum Well Lasers

**Dan Alexandru Anghel, Andreea Rodica Sterian,
and Paul E. Sterian**

*Academic Center for Optical Engineering and Photonics, Faculty of Applied Sciences,
University "Politehnica" of Bucharest, 060042 Bucharest, Romania*

Correspondence should be addressed to Andreea Rodica Sterian, andreea.sterian@physics.pub.ro

Received 29 September 2011; Accepted 28 November 2011

Academic Editor: Carlo Cattani

Copyright © 2012 Dan Alexandru Anghel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In semiconductor laser modeling, a good mathematical model gives near-reality results. Three methods of modeling solutions from the rate equations are presented and analyzed. A method based on the rate equations modeled in Simulink to describe quantum well lasers was presented. For different signal types like step function, saw tooth and sinus used as input, a good response of the used equations is obtained. Circuit model resulting from one of the rate equations models is presented and simulated in SPICE. Results show a good modeling behavior. Numerical simulation in MathCad gives satisfactory results for the study of the transitory and dynamic operation at small level of the injection current. The obtained numerical results show the specific limits of each model, according to theoretical analysis. Based on these results, software can be built that integrates circuit simulation and other modeling methods for quantum well lasers to have a tool that model and analysis these devices from all points of view.

1. Introduction

Laser semiconductor diodes are key components in modern optical communications, storage, printing, medicine, and information processing. These types of devices were developed and evolved constantly in the direction of size reduction and integration. A quantum well (QW) laser improves the functioning characteristics of laser diodes in the direction of low threshold current and narrow emission band as well as emitted wavelength dependence on nanostructure dimension (quantum size effect) [1–8].

A quantum well laser is a structure in which the active region of the device is so narrow that quantum confinement occurs, according to quantum mechanics. The wavelength of the light emitted by a quantum well laser is determined by the width of the active region rather than just the band gap of the material from which the device is realized. Consequently, much shorter wavelengths can be obtained from quantum well lasers than from conventional laser

diodes using a particular semiconductor material. The efficiency of a quantum well laser is also greater than a conventional laser diode due to the stepwise form of its density of states function. In the optoelectronic integrated circuits (OEICs), the quantum well lasers will interact optically and electrically with other devices as integrated modulators, optical fibers and coherent amplifiers, detectors, and optical integrated guides [9–19]. A fully analytical method to describe these kind of complex systems cannot be realistic without questionable simplifying assumptions so that the implementations of refined mathematical methods and algorithms as well as circuit simulators become essentially in the applications development [20–24].

2. General Modeling Methods and Techniques

This paper aims at presenting and analyzing modeling techniques for quantum well lasers starting from the rate equations. The issue of modeling of quantum well semiconductor lasers has been addressed by several authors [1–3, 7, 13]. In operation of a laser, effects are involved that can impact a variety of performance parameters of the device. In some cases is crucial to obtain a clear picture of how a laser works in various conditions or parameters. A good laser model which allows calculation and testing of certain parameters of the device is an important result. The main models of laser with quantum wells are based on the description of the semiconductor lasers using the rate equations formalism.

While the first models were based on one pair of equations to describe the density of photons and carriers in the active region, recent approaches include additional rate equations, to take into account, and carriers transport between active region and adjacent layers of the structure.

Note that in most cases the rate equations lead to multiple solutions, although only one solution is correct. Javro and Kang [25] showed that incorrect solutions or without physical sense can be eliminated or avoided through a change of variables in the rate equations.

However, the transformations used are available under certain conditions and for some cases give unrealistic solutions. These shortcomings are caused mainly due to the linear character of the gain-saturation coefficient. More general expression of the gain-saturation coefficient, proposed by Channin can be used to obtain models for operation regimes having a unique solution. Agrawal suggests another expression for this coefficient, which is also suitable. As it is shown, any of these two forms of the gain-saturation coefficient can be used to obtain models with a solution unique of the operation regime.

In this paper, based on known rate equations models we try, firstly, the circuit simulation for two different modeling systems: Simulink and SPICE. The Simulink modeling technique is very simple. It is based on standard rate equations with the sets of parameters given directly in Simulink. The second model is based on the standard rate equations that use a nonlinear gain-saturation term proposed by Channin. In this second model we also describe the circuit level in SPICE with simulation results.

A numerical experiment in MathCad which gives satisfactory results for the study of the transitory and dynamic operation at small level of the injection signal is shortly presented, also.

3. The Model with Linear Gain Saturation, in Simulink

One of the prevailing laser diode models is based on a set of rate equations. The rigorous derivation of these equations originates from Maxwell equations with a quantum mechanical

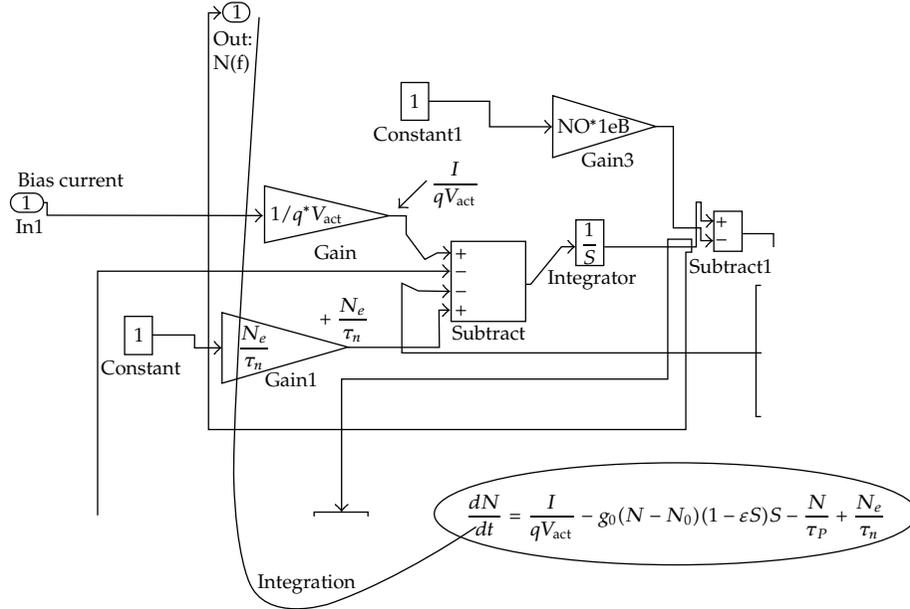


Figure 1: Construction of rate equation block.

approach for the induced polarization. However, the rate equation could also be derived by considering physical phenomena [1, 3, 5]:

$$\frac{dN}{dt} = \frac{I}{qV_{\text{act}}} - g_0(N - N_0)(1 - \epsilon S) - \frac{N}{\tau_p} + \frac{N_e}{\tau_n}, \quad (3.1)$$

$$\frac{dS}{dt} = \Gamma g_0(N - N_0)(1 - \epsilon S)S + \frac{\Gamma \beta N}{\tau_n} - \frac{S}{\tau_p}, \quad (3.2)$$

$$\frac{S}{P_f} = \frac{\Gamma \tau_p \lambda_0}{V_{\text{act}} \eta h c} = \vartheta. \quad (3.3)$$

Equation (3.1) relates the rate of change in carrier concentration N to the injection current I , the carrier recombination rate and the stimulated emission rate. Equation (3.2) relates the rate of change in photon density S to the photons loss, the rate of coupled recombination into the lasing mode, and the stimulated emission rate. The ratio between photon density S to the output power P_f is described in (3.3). The other parameters used have well-known significance [2]. This simple model can be directly implemented with Simulink without any problems.

A direct implementation looks like that in Figure 1.

Figures 1 and 2 show how the rate equations model is constructed in Simulink. with I as input parameter from a signal generator and S , N , and P_f as output parameters. All the parameters in the rate equations can be modified before the simulation starts.

For the simulation we used the following parameters found in specialized papers [1–5]: $\lambda_0 = 1.502 \times 10^{-4}$ cm (laser wavelength), $V_{\text{act}} = 9 \times 10^{-11}$ cm³ (active region volume),

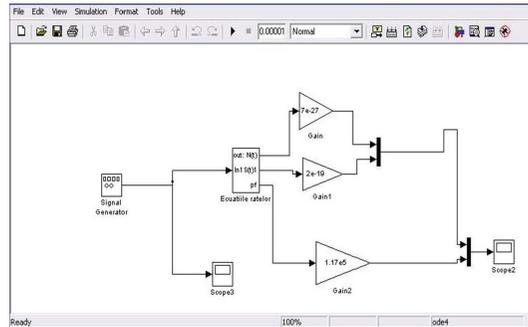


Figure 2: Rate equation in Simulink.



Figure 3: Output for step input signal of 10 mA amplitude and frequency of 5 MHz.

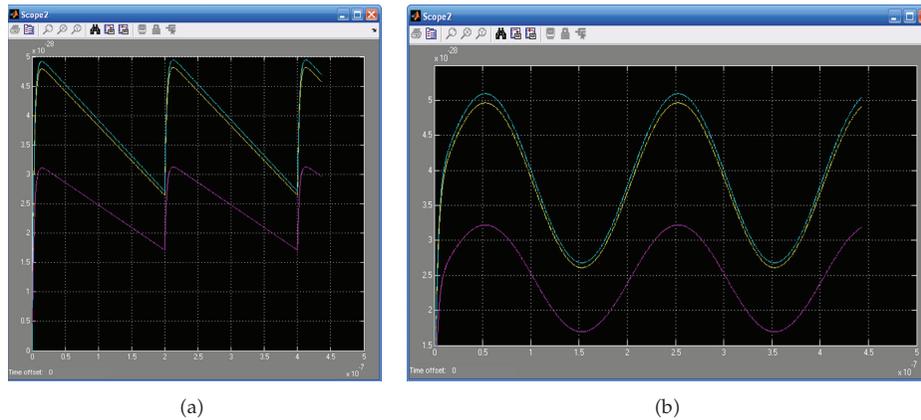


Figure 4: Output for saw tooth input signal of 10 mA amplitude and frequency of 5 MHz (a) and for sinusoidal input signal (b).

$\Gamma = 0.44$ (optical confinement factor), $B = 4 \times 10^{-4}$ (spontaneous emission factor), $g_0 = 3 \times 10^{-6} \text{ cm}^3/\text{s}$ (gain coefficient), $N_0 = 1.2 \times 10^{18} \text{ cm}^{-3}$ (optical transparency density), $\tau_n = 3 \text{ ns}$ (carrier lifetime), $\tau_p = 1 \text{ ps}$ (photon lifetime), $\eta = 0.1$ (quantum efficiency), $N_e = 5.41 \times 10^{10} \text{ cm}^{-3}$ (equilibrium carrier density), and $\varepsilon = 3.4 \times 10^{-17} \text{ cm}^3$ (gain saturation factor).

With the above model different signal forms can be used as input for the quantum well laser. They show that theoretical response of the equations is good in comparison with real results that are obtained in applications. Signals like step function, saw tooth, and sine types are used as input. The results are shown in Figures 3 and 4, which illustrate a very fast response at low current level. Low threshold current is the main feature of quantum well lasers, and it is directly shown for the basic form of rate equations. The simulation is not perfect, and this is because of negative solutions for N and S and high power solution of the equations.

4. The Model with Nonlinear Gain Saturation, in SPICE

To obtain nonnegative solutions for the photon density and the density of carriers in terms of a nonnegative current injection, a different version of the standard rate equations must be formulated. In the corresponding generalized equations (3.1) and (3.2), the linear gain-saturation term is replaced by a nonlinear term proposed by Channin or Agrawal as follows [1-3]:

$$\frac{dN}{dt} = \frac{\eta_i I}{qN_w V_{act}} - R_w(N) - \Gamma_c \nu_{gr} \frac{\alpha(N)}{\Phi(S)} S, \quad (4.1)$$

$$\frac{dS}{dt} = -\frac{S}{\tau_p} + N_w R_{w\beta}(N) + N_w \Gamma_c \nu_{gr} \frac{\alpha(N)}{\Phi(S)} S, \quad (4.2)$$

$$\frac{S}{P_f} = \frac{\lambda \tau_p}{\eta_c V_{act} h c} = \vartheta. \quad (4.3)$$

Equation (4.1) describes the carrier concentration rate dependence on the injection current, the carrier recombination rate $R_w(N)$ and the stimulated emission rate. To take into account all mechanisms of recombination, we consider: $R_w(N) = AN + BN^2 + CN^3$, where A , B , and C are, respectively, the unimolecular, radiative, and Auger recombination coefficients.

Equation (4.2) shows the photon density rate dependence on photon loss, the rate of coupled recombination into the lasing mode, and the stimulated emission rate. Equation (4.3) relates the output power P_f to the photon density S . The other parameters used have well-known significance [2]. In (4.1) and (4.2), the correlation between the material gain and the carrier density is given by the logarithmic carrier-dependent gain term:

$$\alpha(N) = G_0 \ln \left(\frac{R_w(N)}{R_w(N_0)} \right), \quad (4.4)$$

where G_0 is the one quantum well gain coefficient and N_0 is the optical transparency density while the gain-saturation function can take on one of the following two forms:

$$\phi^{-1}(S) = \frac{1}{1 + \varepsilon \Gamma_c S} \text{ (Channin)} \quad \text{or} \quad \phi^{-1}(S) = \frac{1}{\sqrt{1 + \varepsilon \Gamma_c S}} \text{ (Agrawal)}. \quad (4.5)$$

An equivalent circuit model based on (4.1)–(4.3) can be implemented in SPICE as in Figure 5. Unlike models based on the rate equations that use a linear gain-saturation term, this circuit model is applicable for all nonnegative values of injection current.

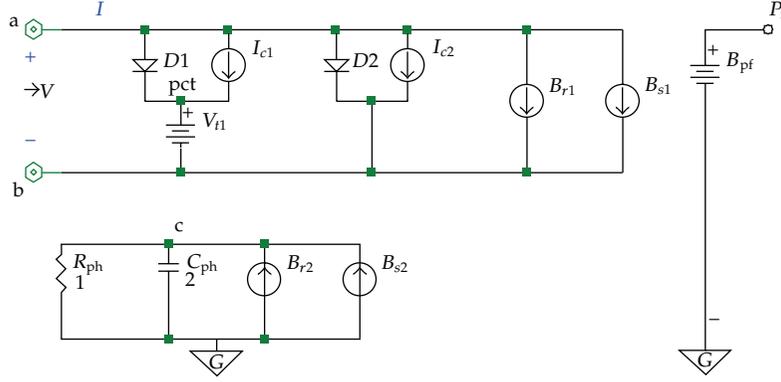


Figure 5: The circuit implementation of the model.

Figure 6 shows the circuit implementation of the model, using suitable variables transformations [25]: $N = N_e \exp(qV/nkT)$ and $P_f = (m + \delta)^2$, which impose nonnegative solutions for N and P_f . The linear recombination and charge storage in the device are described by diodes $D1$ and $D2$ and current sources I_{c1} and I_{c2} . The effects of additional recombination mechanisms and stimulated emission, respectively, on the carrier density are modeled by B_{r1} and B_{s1} , according to (4.11) and (4.13).

The effects of spontaneous and stimulated emission, respectively, are accounted in the circuit by B_{r2} and B_{s2} while R_{ph} and C_{ph} completes the model for the time variation of the photon density. The voltage B_{pf} describe the optical output power of the laser.

The equations for the elements of the circuit in Figure 8 according to [2] are as follows: $I = I_{T1} + I_{D1} + I_{C1} + B_{r1} + B_{s1}$, where

$$I_{T1} = I_{D1} + I_{C1}, \quad (4.6)$$

$$2\tau_p \frac{dm}{dt} + m = B_{r2} + B_{s2}, \quad B_{pf} = (m + \delta)^2, \quad (4.7)$$

$$I_{D1} = \frac{qN_w V_{act} N_e}{2\eta_i \tau_n} \left[\exp\left(\frac{qV}{nkT}\right) - 1 \right], \quad (4.8)$$

$$I_{D2} = \frac{qN_w V_{act} N_e}{2\eta_i \tau_n} \left[\exp\left(\frac{qV}{nkT}\right) - 1 + \frac{2q\tau_p}{nkT} \exp\left(\frac{qV}{nkT}\right) \frac{dV}{dt} \right], \quad (4.9)$$

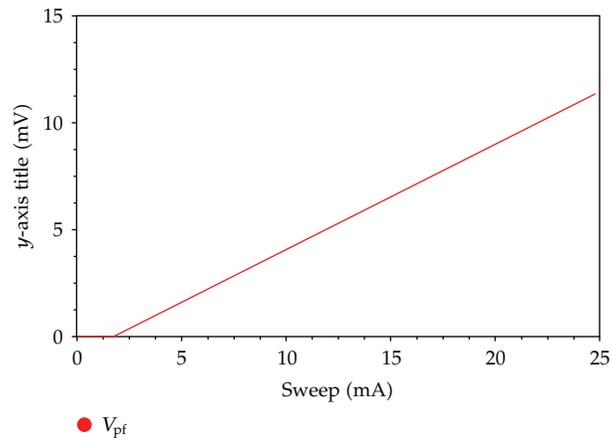
$$I_{c1} = I_{c2} = \frac{qN_w V_{act} N_e}{2\eta_i \tau_n}, \quad (4.10)$$

$$B_{r1} = \frac{qN_w V_{act}}{\eta_i} R_{\omega 2}(\Theta I_{T1}), \quad (4.11)$$

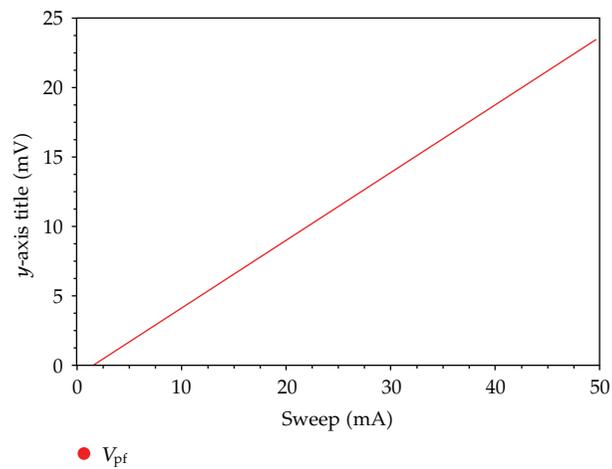
$$B_{s1} = \frac{\lambda \tau_p N_w \Gamma_c v_{gr}}{\eta_i \eta_c hc} \frac{\alpha(\Theta I_{T1})}{\phi(\vartheta(m + \delta)^2)} \vartheta(m + \delta)^2, \quad (4.12)$$

$$B_{r2} = \frac{N_w \eta_c V_{act} hc}{\lambda \vartheta(m + \delta)} R_{\omega 2}(\Theta I_{T1}), \quad (4.13)$$

$$B_{s2} = \tau_p N_w \Gamma_c v_{gr} \frac{\alpha(\Theta I_{T1})}{\phi(\vartheta(m + \delta)^2)} (m + \delta) - \delta. \quad (4.14)$$



(a)



(b)

Figure 6: Answer of optical power with variation of the input bias current between 0 and 25 mA (a) and between 0 and 50 mA (b).

The SPICE netlist was implemented in AIM-SPICE. Calculating the parameters in the netlist is time consuming, and if a calculation is wrong, the netlist will fail in SPICE or the result will not be good. Figures 6, 7, and 8 resulting from simulation with simple DC sweep and transitory sweep illustrate the following:

- (i) answer of optical power with variation of the input bias current between 0 and 25 mA (a) and between 0 and 50 mA (b) (Figure 6);
- (ii) transient output power in response between 0 and 100 ps (Figure 7);
- (iii) transient output power in response between 0 and 5 ns (Figure 8).

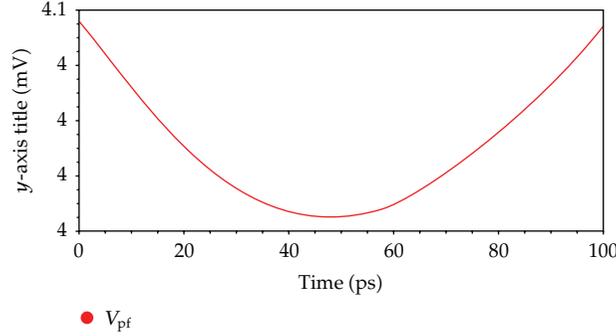


Figure 7: Transient output power in response between 0 and 100 ps.

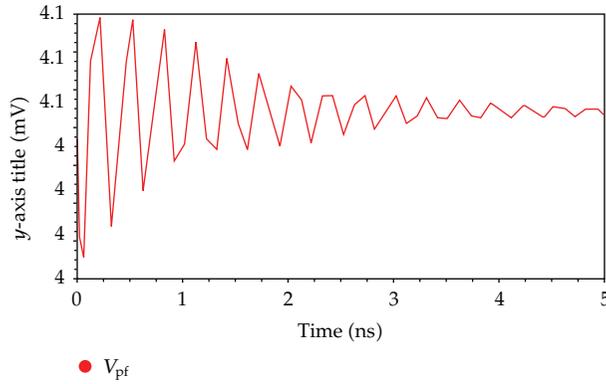


Figure 8: Transient output power in response between 0 and 5 ns. Note: Output power is in V because SPICE cannot output variables in W.

5. Numerical Experiment in MathCad

Study of functional characteristics of QWL is possible based on the models described previously, by implementing MathCad programs to integrate these equations using appropriate algorithms corresponding to different types of pumping signals.

The model given by the equations with linear gain-saturation term (3.1)–(3.3) was integrated for a constant current injection $I_i = 0.05$ A. The waveforms corresponding to I_i and to output optical power P are obtained [13].

The model given by the equations with nonlinear gain saturation (4.1)–(4.3) was integrated for a constant current injection $I_i(t) = 0.025$ A (Figure 9), sinus current injection (Figure 10):

$$I_i(t) = 1 + \sin(5 \cdot 10^8 t), \quad (5.1)$$

and rectangular current injection (Figure 11), which for several periods can be written as:

$$I_i(t) = 10^{-8} + 0.01 + 0.25I_i(t - 2 \cdot 10^{-9}) \\ + 0.25I_i(t - 7 \cdot 10^{-9}) + 0.25I_i(t - 12 \cdot 10^{-9}). \quad (5.2)$$

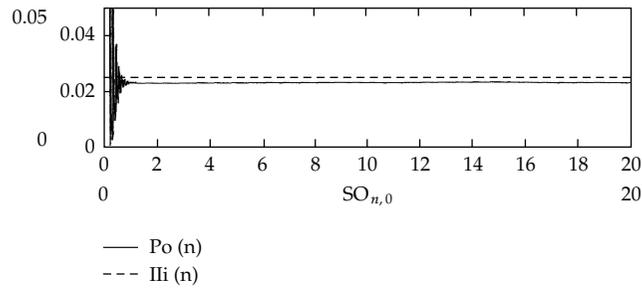


Figure 9: The waveforms corresponding to I_i and to output optical power P .

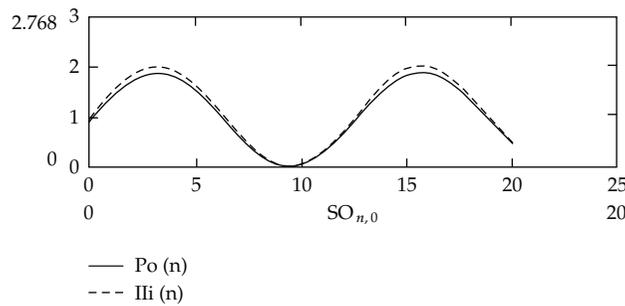


Figure 10: The waveforms corresponding to I_i and to output optical power P .

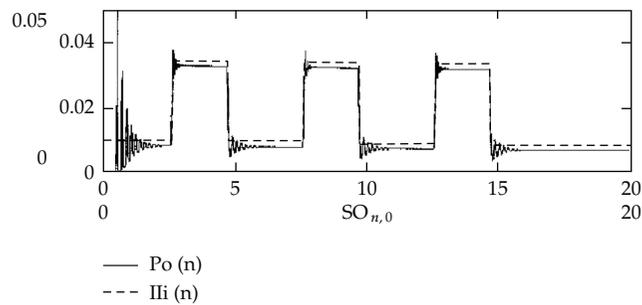


Figure 11: The waveforms corresponding to I_i and to output optical power P .

Simulations results for these large-signal models have shown excellent agreement with experimental data. A corresponding small-signal model has been derived for the quantum well lasers, to study the modulation properties in the frequency domain [13].

6. Conclusions

- (1) In the paper, some modeling methods and techniques for the quantum well lasers have been implemented and validated. A suitable modeling technique is very important for technology designing and analyzing of the integrated optoelectronic circuits.

- (2) A method based on the rate equations for using Simulink to describe quantum well lasers was presented. For different signal types like step function, saw tooth, and sinus used as input, the results show a good theoretical response of the equations in comparison with the real results obtained in applications. This method is useful to determine the various regions of operation of the laser diode also.
- (3) The SPICE simulation of the circuits shows that modifying one parameter will result in new calculations and new SPICE netlist, which is a rather complicated procedure. Simulation is not limited to SPICE; any all-purpose circuit simulator can be used to get similar results.
- (4) Generally, all studied models give satisfactory results for the study of the transitory and dynamic operation at small level of the injection signals. At large injection signal levels the obtained numerical results show the specific limits of each model, according to theoretical analysis.
- (5) As a future development, software can be built that integrates circuit simulations and other modeling methods for quantum well lasers to have a tool that models these devices from all points of view.

Acknowledgments

The author A. R. Sterian acknowledges the financial support from CNCSIS-UEFISCSU, project number PNII-IDEI ID.954/2007 and ID.123/2008.

References

- [1] P. S. Zory, *Quantum Well Lasers*, Academic Press, 1993.
- [2] P. V. Mena, S.-M. Kang, and T. A. DeTemple, "Rate-equation-based laser models with a single solution regime," *Journal of Lightwave Technology*, vol. 15, no. 4, pp. 717–729, 1997.
- [3] D. S. Gao, S. M. Kang, R. P. Bryan, and J. J. Coleman, "Modeling of quantum-well lasers for computer-aided analysis of optoelectronic integrated circuits," *IEEE Journal of Quantum Electronics*, vol. 26, no. 7, pp. 1206–1216, 1990.
- [4] S. Ghoniemy, L. MacEachern, and S. Mahmoud, "Extended robust semiconductor laser modeling for analog optical link simulations," *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 9, no. 3, pp. 872–878, 2003.
- [5] M. Dehghan and P. Derakhshan-Barjoei, "A novel dynamic analysis and simulation for quantum-well distributed feedback laser," *Progress In Electromagnetics Research B*, vol. 3, pp. 105–114, 2008.
- [6] G. P. Agrawal, *Fiber-Optic Communication Systems*, John Wiley and Sons, New York, NY, USA, 1997.
- [7] B. P. C. Tsou and D. L. Pulfrey, "A versatile SPICE model for quantum-well lasers," *IEEE Journal of Quantum Electronics*, vol. 33, no. 2, pp. 246–254, 1997.
- [8] N. Bewtra, D. A. Suda, G. L. Tan, F. Chatenoud, and J. M. Xu, "Modeling of quantum-well lasers with electro-opto-thermal interaction," *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 1, no. 2, pp. 331–340, 1995.
- [9] A. R. Sterian, "Coherent radiation generation and amplification in erbium doped systems," in *Advances in Optical Amplifiers*, P. Urquhart, Ed., InTech, Vienna, Austria, 2011.
- [10] A. R. Sterian, *Amplificatoare Optice*, Editura Printech, București, Romania, 2006.
- [11] G. Mattioli, M. Scalia, and C. Cattani, "Analysis of large-amplitude pulses in short time intervals: application to neuron interactions," *Mathematical Problems in Engineering*, vol. 2010, Article ID 895785, 15 pages, 2010.
- [12] A. R. Sterian and F. C. Maciuc, "Numerical model of an EDFA based on rate equations," in *the 12th International School on Quantum Electronics Laser Physics and Applications*, vol. 5226 of *Proceedings of SPIE*, pp. 74–78, 2003.
- [13] A. R. Sterian, *Laserii in ingineria electrica*, Editura Printech, București, Romania, 2003.

- [14] E. Ștefănescu, A. R. Sterian, and P. Sterian, "Study on the fermion systems coupled by electric dipol interaction with the free electromagnetic field," in *Advanced Laser Technologies*, vol. 5850 of *Proceedings of SPIE*, pp. 160–165, 2005.
- [15] E. Ștefănescu, P. Sterian, and A. R. Sterian, "The Lindblad dynamics of a Fermi system in a particle dissipative environment," in *ALT '02 International Conference on Advanced Laser Technologies*, Proceedings of SPIE, pp. 160–168, September 2002.
- [16] A. Sterian and V. Ninulescu, "Nonlinear phenomena in erbium-doped lasers," in *the International Conference on Computational Science and Its Applications (ICCSA '05)*, O. Gervasi et al., Ed., vol. 3482 of *Lecture Notes in Computer Science*, pp. 643–650, 2005.
- [17] A. D. Petrescu, A. R. Sterian, and P. E. Sterian, "Solitons propagation in optical fibers computer experiments for students training," in *the International Conference on Computational Science and its Applications (ICCSA '07)*, O. Gervasi and M. L. Gavrilova, Eds., vol. 4705 of *Lecture Notes in Computer Science*, pp. 450–461, 2007.
- [18] A. R. Sterian, *Mecanica cuantică*, Omnia Univ. S.A.S.T., Brașov, Romania, 2008.
- [19] A. R. Sterian, "Computer modeling of the coherent optical amplifier and laser systems," in *the International Conference on Computational Science and its Applications (ICCSA '07)*, vol. 4705 of *Lecture Notes in Computer Science*, pp. 436–449, 2007.
- [20] C. H. Lee, T. H. Yoon, and S. Y. Shin, "Period doubling and chaos in a directly modulated laser diode," *Applied Physics Letters*, vol. 46, no. 1, pp. 95–97, 1985.
- [21] R. Nagarajan, M. Ishikawa, T. Fukushima, R. S. Geels, and J. E. Bowers, "High speed quantum-well lasers and carrier transport effects," *IEEE Journal of Quantum Electronics*, vol. 28, no. 10, pp. 1990–2008, 1992.
- [22] C. Cattani and E. Nosova, "Transversal waves in nonlinear Signorini model," in *the International Conference on Computational Science and Its Applications (ICCSA '08)*, vol. 5072 of *Lecture Notes in Computer Science*, pp. 1181–1190, 2008.
- [23] C. Cattani, "Harmonic wavelets towards the solution of nonlinear PDE," *Computers & Mathematics with Applications*, vol. 50, no. 8-9, pp. 1191–1210, 2005.
- [24] C. Cattani and I. Bochicchio, "Wavelet analysis of bifurcation in a competition model," in *the 7th International Conference on Computational Science (ICCS '07)*, vol. 4488 of *Lecture Notes in Computer Science*, pp. 990–996, 2007.
- [25] S. A. Javro and S. M. Kang, "Transforming Tucker's linearized laser rate equations to a form that has a single solution regime," *Journal of Lightwave Technology*, vol. 13, no. 9, pp. 1899–1904, 1995.

Research Article

Fisheye-Lens-Based Visual Sun Compass for Perception of Spatial Orientation

Liang Wang,¹ Fuqing Duan,² and Ke Lv³

¹ College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China

² College of Information Science and Technology, Beijing Normal University, Beijing 100875, China

³ College of Computing and Communication Engineering, Graduate University of the Chinese Academy of Sciences, Beijing 100049, China

Correspondence should be addressed to Liang Wang, wangliang@bjut.edu.cn

Received 11 October 2011; Accepted 9 December 2011

Academic Editor: Sheng-yong Chen

Copyright © 2012 Liang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In complex aeronautical and space engineering systems, conventional sensors used for environment perception fail to determine the orientation because of the influences of the special environments, wherein geomagnetic fields are exceptional and the Global Positioning System is unavailable. This paper presents a fisheye-lens-based visual sun compass that is efficient in determining orientation in such applications. The mathematical model is described, and the absolute orientation is identified by image processing techniques. For robust detection of the sun in the image of the visual sun compass, a modified maximally stable extremal region algorithm and a method named constrained least squares with pruning are proposed. In comparison with conventional sensors, the proposed visual sun compass can provide absolute orientation with a tiny size and light weight in especial environments. Experiments are carried out with a prototype validating the efficiency of the proposed visual sun compass.

1. Introduction

Mobile robots possessing various sensors are widely employed for taking measurements and performing tasks in outdoor applications in many fields, including the complex aeronautical and space engineering systems. These applications present many great challenges, one of which is that conventional sensors do not work in emerging special environments. For example, in some planets, magnetic heading detection devices are nonfunctional because of the negligible magnetic fields, and the Global Positioning System (GPS) receiver does not work because there is no GPS available. Therefore, a novel sensor is required to perceive the environment and the robot's own state.

The special aeronautical and space environments have little influence on visual sensors. In addition, being a type of passive sensor, visual sensors do not need a precise and complex mechanical structure, indicating that vibration has little effect on them and there is no mechanical fault. More importantly, visual sensors can provide abundant information, including details of depth, texture, illumination, and so on, with only one measurement (image) [1, 2]. All of these characteristics ensure that visual sensors have the potential to overcome the aforementioned challenges. Therefore, the visual sun compass is proposed to determine orientation for applications in aeronautical and space engineering systems.

The concept of the sun compass is derived from studies of the migration of birds [3]. It has been found that birds navigate and orient themselves with the sun as the reference during their migration. Some ornithologists analyzed this sun compass both in theory and by experiment [3]. Then, sun compasses were designed and used for navigation. These compasses were found to perform well in regions with exceptional geomagnetic fields where the magnetic compass fails to work. However, the existing sun compasses (denoted as the classical sun compass) were designed on the basis of the sundial. To obtain precise orientation measurements, the sundial should be sufficiently large, which makes its volume and weight very huge and limits its applications. As a mechanical sensor, it is difficult to embed such a compass in a measuring system and coordinate it with electronic sensors.

The visual sun compass proposed in this paper is composed of a fisheye lens and a camera. With improvements and advances in manufacturing techniques, fisheye lens and cameras with high precision and compact structure have been developed, which make the visual sun compass satisfy the requirements of many applications and be easily embedded in a measuring system. In addition, such cameras can cooperate with other electronic sensors because of its electronic output. This type of lens has a large field of view of about 180 degrees, and there is no occlusion to the view and no requirement for precise mechanical assembly.

There exist some related works in the literature. The star tracker [4], which comprises a conventional lens and an Active Pixel Sensor camera, is a very precise attitude-measurement sensor. This sensor uses stars as the frame of reference and performs estimation of orientation by identifying the observed stars and measuring their relative positions. However, this sensor is susceptible to various errors, particularly the bright sunlight, because it has high sensibility and is significantly dependent on the star identification algorithm. In addition, the high costs limit its use. Sun sensors are another type of instrument that is similar to the proposed visual sun compass. The sun sensor proposed by Cozman and Krotkov [5] is made up of a telephoto lens and a camera. Because a telephoto lens has a very small field of view, a precise mechanical tracking system is needed to capture the sun, and this makes the sensor susceptible to vibration and mechanical fault. Sun sensors proposed by Deans et al. [6] and Trebi-Ollennu et al. [7], which comprise a fisheye lens and a camera, can determine the relative orientation. The imaging model of a pinhole camera is used to describe the imaging mechanism of the fisheye lenses. In addition, extensions with some distortion terms have been proposed; these do not take the characteristics of fisheye lenses into account, but rather introduce some nuisance parameters and increase the computational burden. The centroid of pixels whose intensity values are above a certain threshold is taken as the image of the sun's center while taking image measurements of the sun [6–9]. In fact, the centroid of these pixels may not be the image of the sun's center because of image noise and outliers.

In this paper, a mathematical model of the visual sun compass, which takes into account characteristics of fisheye lenses, is presented. An effective method for detecting and delineating the sun's image is provided to robustly calculate the azimuth and elevation angle of the sun. Finally, with the knowledge of the celestial navigation, an estimation of

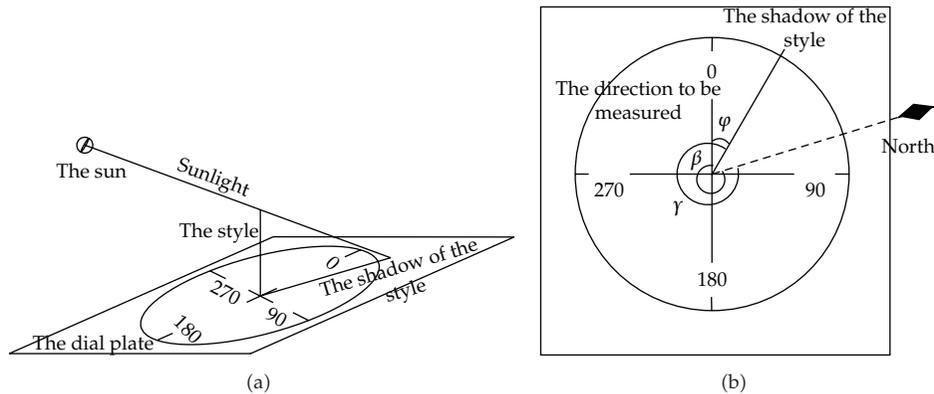


Figure 1: Illustration of the classical sun compass. (a) A sketch of the principle of the classical sun compass. (b) Angles on the dial plate (β : the direction to be measured, γ : the azimuth angle of the sun, φ : the angle of the style's shadow).

absolute orientation can be obtained. The remainder of this paper is organized as follows. In Section 2, the classical sun compass has been introduced. The model of the proposed visual sun compass has been described in Section 3. Section 4 provides a method for detecting and delineating the sun's image in the visual sun compass. Section 5 presents some experimental results and, finally, in Section 6 the conclusions of this study are presented.

2. The Classical Sun Compass

The design of the classical sun compass is based on the sundial, which is originally used to measure time on the basis of the position of the sun's shadow. A sundial is composed of a style (usually a thin rod or a straight edge) and a dial plate on which hour (or angle) lines are marked. The sun casts a shadow from the style onto the dial plate (see Figure 1(a)). As the sun moves, the shadow aligns itself with different angle lines on the plate. If the absolute direction is known, the time at the measuring moment can be calculated on the basis of the shadow. The classical sun compass inverts the traditional role of a sundial. Instead of measuring time, it calculates the absolute direction with the shadow and the time information at the measuring moment.

To calculate the absolute direction, first, the dial plate of the classical sun compass should be adjusted to be horizontal. Then, the direction to be estimated should be aligned with one angle line; in general, the angle line with mark 0 is selected. The angle of the style's shadow, that is, the angle φ in Figure 1(b), and the time at the measuring moment should be measured. To determine the direction β ($\beta = \gamma - \alpha$), the azimuth angle of the sun γ must be calculated. With the knowledge of celestial navigation, the azimuth angle γ can be calculated. As shown in Figure 2, O is the observer's position, L is the latitude of the observer, N is the north pole of the planet, P is the projection point of the sun on the planet, D is the declination of the sun (the declination is the angle between the rays of the sun and the planet's equator), and H is the local hour angle (the local hour angle of a point is the angle between the half plane determined by this point and the planet's axis, and the half plane determined by the sun's projection point and the planet's axis, and it can be used to describe the position of a

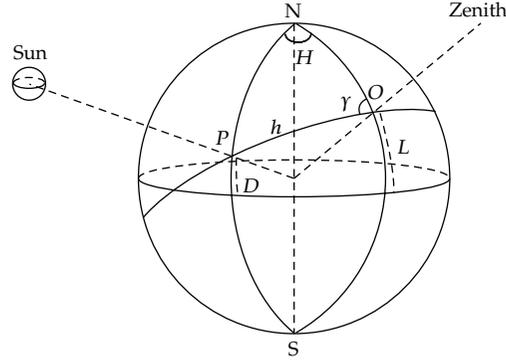


Figure 2: Illustration of various angles on the planet when measuring with the sun compass (see text for details of N , P , H , D , O , L , and γ).

point on the celestial sphere). In the spherical triangle ΔPNO , by the spherical law of cosines (also called the cosine rule for sides), we have

$$\cos(h) = \sin(L) \sin(D) + \cos(L) \cos(D) \cos(H). \quad (2.1)$$

In (2.1), the observer's latitude L should be known, and from this the declination of the sun D can be calculated on the basis of the time at measuring moment, and the local hour angle H can be calculated with longitudes of the observer and the projection point of the sun. After the side of spherical triangle ΔPNO , h , being obtained by the spherical law of cosines, the azimuth angle of the sun, γ , can be expressed as

$$\sin(\gamma) = -\frac{\cos(D) \sin(H)}{\sin(h)}. \quad (2.2)$$

In addition, by the spherical cosine law for sides,

$$\cos(\gamma) = \frac{(\sin(D) - \cos(L) \cos(h))}{\cos(L) \sin(h)}. \quad (2.3)$$

Thus, the azimuth angle γ can be determined uniquely, and the direction $\beta (= \gamma - \alpha)$ can be obtained. In general, it is difficult to obtain precise values of the observer's longitude and latitude in advance. However, the exact position can be obtained with assumed values of two or more positions and observed azimuth angles by the intercept method [10].

3. The Visual Sun Compass

The proposed visual sun compass, which is composed of a fisheye lens and a camera, is essentially a fisheye camera. Figure 3 shows its imaging process. For a space point $\mathbf{P} = [X \ Y \ Z]^T$ (where $[\bullet]^T$ denotes the transpose of a vector or matrix), its fisheye image is $\mathbf{p} = [x \ y]^T$, whereas its virtual perspective image would be $\mathbf{p}' = [x' \ y']^T$ as obtained by a pinhole camera. In comparison with a pinhole camera, a fisheye camera has a large

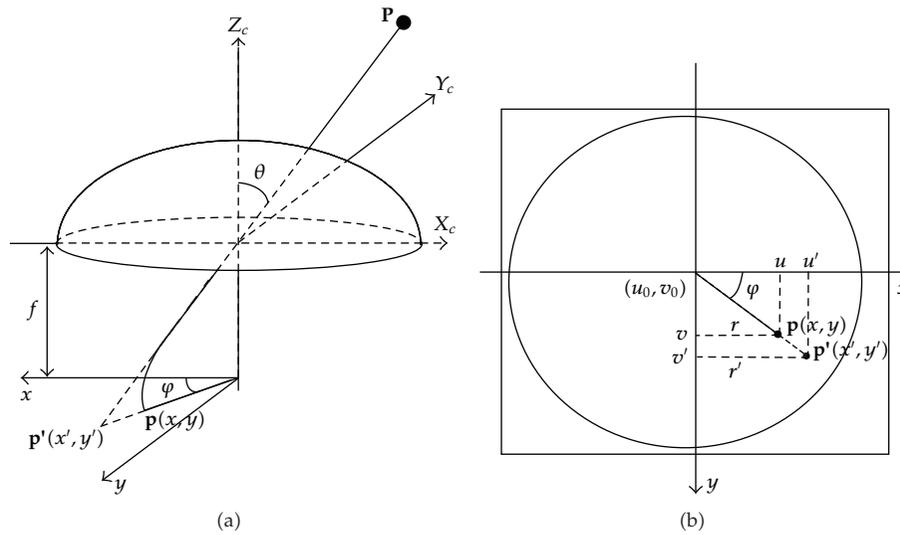


Figure 3: Illustration of a visual sun compass. (a) Light path. (b) A captured image.

field of view of about 180 degrees, which ensures that the sun can be captured with only one image when its optical axis points at the sky. Further, it can be regarded as a sundial, where the optical axis is the style and the image plane is the dial plate. Thus, the sun needs to be identified in the captured fisheye image. Then the azimuth and elevation angle of the sun can be calculated with its imaging model. If the optical axis upright points at the sky when capturing images, the absolute direction (and by the intercept method [10] even the exact position) can be calculated with the time at the image-capture moment and the longitude and latitude of the observer by the method presented in Section 2. Otherwise, the measured azimuth and elevation angles should, first, be transformed into the ground coordinate system with the assistance of an inclinometer or inertial sensor.

The image of a space point P is p whereas it would be p' by a pinhole camera; (θ, φ) is the incidence angle of p (or the direction of the incoming ray from P); r is the distance between the image point p and the principal point (u_0, v_0) ; and f is the focal length.

3.1. Imaging Model of Fisheye Camera

A pinhole camera model with some distortion terms is used to describe the fisheye lens and determine the relative direction in [6, 7], which do not consider the characteristics of fisheye lenses. In fact, fisheye lenses are usually designed to follow some types of projection geometry. Table 1 shows some ideal projection models of fisheye lenses [11]. Because the ideal model is not strictly satisfied and, in some cases, the type of a fisheye lens may be unknown in advance; a more general mathematical model is needed to describe the visual sun compass.

Various imaging models of fisheye cameras have been proposed in the literature. These models can be divided into two categories as follows.

Table 1: Projection geometry of fisheye lens.

Project type	Model
Stereographic	$r = 2f \tan(\theta/2)$
Orthogonal	$r = f \sin(\theta)$
Equidistant	$r = f\theta$
Equisolid angle	$r = 2f \sin(\theta/2)$

3.1.1. Mapping from an Incident Ray to a Fisheye Image Point

A fisheye camera can be described by the mapping from an incident ray to a fisheye image point: $[X/R \ Y/R \ Z/R]^T \rightarrow [x \ y]^T$ or $\theta \rightarrow r$, where $R = \sqrt{X^2 + Y^2 + Z^2}$, the incidence angle θ is the angle between the principle axis and the incident ray $\mathbf{p}'\mathbf{P}$ (see Figure 3), and $r = \sqrt{x^2 + y^2}$.

3.1.2. Mapping from a Virtual Perspective Image to a Fisheye Image

A fisheye camera can also be described with the mapping from a virtual perspective image to a fisheye image by $[x' \ y']^T \rightarrow [x \ y]^T$ or $r' \rightarrow r$, where $r' = \sqrt{x'^2 + y'^2}$. The rational function [12] belongs to this class

$$r = \frac{\varepsilon_1 r'}{(1 - \varepsilon_2 r'^2)}, \quad (3.1)$$

where ε_1 and ε_2 are two parameters for the lens type.

The former class is more straightforward and popular and represents the real imaging process of a fisheye camera [12]. In fact, the latter class can also be described in the form of the former. For example, (3.1) can be expressed as $r = \varepsilon_1 f \tan(\theta) / (1 - \varepsilon_2 f^2 \tan^2(\theta))$.

Different to fisheye cameras, central catadioptric cameras have an elegant unified imaging model [13]. Ying and Hu [14] discovered that this unified imaging model could be extended to some fisheye cameras.

As shown in Figure 4, the unified imaging model for central catadioptric cameras can be described as follows. A space point $[X \ Y \ Z]^T$ is first projected to a point $[X/R \ Y/R \ Z/R]^T$ on the unit sphere. Then, it is mapped on the image plane, $z = -m$, via a projection from the point $[0 \ 0 \ \xi]^T$. The coordinates of the point on the image plane are

$$[x \ y]^T = \begin{bmatrix} \frac{(\xi + m)X}{\xi R - Z} & \frac{(\xi + m)Y}{\xi R - Z} \end{bmatrix}^T. \quad (3.2)$$

Now, with $r = \sqrt{x^2 + y^2}$, we have a mapping: $\theta \rightarrow r$,

$$r = \frac{(\xi + m) \sin(\theta)}{\xi + \cos(\theta)}. \quad (3.3)$$

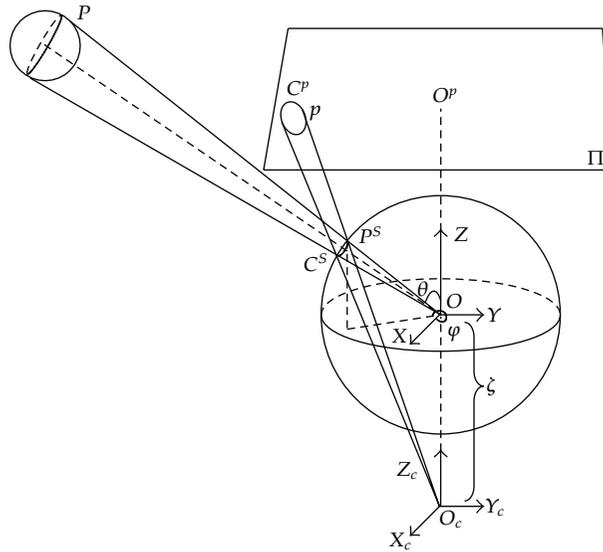


Figure 4: The unified imaging model for central catadioptric cameras.

Equation (3.3) can describe the imaging process of catadioptric projections with parabolic, hyperbolic, and elliptical mirrors with different ξ and m [13]. Let $\xi = \infty$, for all $m \neq \infty$; then (3.3) can describe the imaging process of a fisheye camera with the orthogonal projection. Let $\xi = 1$, $m = 2f - 1$, then (3.3) can describe the imaging process of a fisheye camera with the stereographic projection. All of these mean that the unified imaging model (3.3) can be used to describe some fisheye cameras, which are consistent with conclusions of Ying and Hu [14].

However, for fisheye cameras with equidistant and equisolid angle projections, which are more widely used in real applications, the unified imaging model does not fit. Assuming without loss of generality that $m = 1$ and $f = 1$, submit $r = f\theta$ and $r = 2f \sin(\theta/2)$ into (3.3), respectively. The obtained type parameters ξ are $(\sin(\theta) - \theta \cos(\theta)) / (\theta - \sin(\theta))$ and $\cos(\theta/2) / (1 - \cos(\theta/2))$. This shows that the parameters ξ of these two types of fisheye cameras vary with the incident angle θ . This indicates that these fisheye cameras do not have a unique central view point. Therefore, the unified model (3.3) does not exactly describe the imaging process of all fisheye cameras.

In fact, the Maclaurin Series of all fisheye lenses listed in Table 1 have the same form

$$r = k_1\theta + k_2\theta^3 + k_3\theta^5 + \dots \quad (3.4)$$

Without loss of generality, we take (3.4) as the model of fisheye lenses. Different from the model proposed by Trebi-Ollennu et al. [7], there are only odd-order terms. As shown in Figure 5, there are only slight differences between the ideal models and their 3rd- and 5th-order Maclaurin Series when $\theta \in [0, \pi/2]$, that is, within the field of view of fisheye lenses. In practice, the 3rd-order Maclaurin Series is taken as a compromise between precision and computation efficiency:

$$r = k_1\theta + k_2\theta^3. \quad (3.5)$$

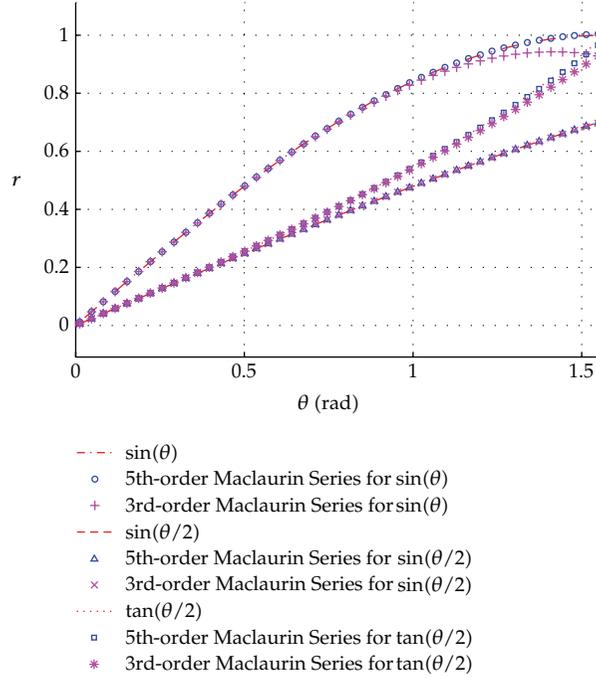


Figure 5: The ideal models of fisheye lenses and their 3rd- and 5th-order Maclaurin Series.

Then, the projection procedure of a fisheye camera can be described with the following four steps.

Step 1. A space point \mathbf{P} in the world coordinate system is transformed to incident angles (θ, φ) in the camera coordinate system with the relative rotation \mathbf{R} and translation \mathbf{t} of two coordinate systems:

$$[\sin(\theta) \cos(\varphi) \quad \sin(\theta) \sin(\varphi) \quad \cos(\theta)]^T = \frac{(\mathbf{R}\mathbf{X} + \mathbf{t})}{\|\mathbf{R}\mathbf{X} + \mathbf{t}\|}. \quad (3.6)$$

Step 2. The incoming direction (θ, φ) is transformed into the normalized image coordinates (x, y) . In this case, the general form (3.5) of fisheye cameras' model is considered. Then, the normalized image coordinates can be expressed as

$$[x \quad y]^T = (k_1\theta + k_2\theta^3) [\cos(\varphi) \quad \sin(\varphi)]^T. \quad (3.7)$$

Step 3. To compensate for deviations of the fisheye model from the real camera, distortion terms, called radial distortion and decentering distortion, can be introduced. The former represents the distortion of ideal image points along radial directions from the distortion

center. The latter represents the distortion of ideal image points in tangential directions. The distortions can be expressed as functions of radial distance $r = \sqrt{x^2 + y^2}$:

$$\begin{aligned}\delta_x &= x(m_1 r^2 + m_2 r^4) + [p_1(r^2 + 2x^2) + p_2 xy](1 + p_3 r^2), \\ \delta_y &= y(m_1 r^2 + m_2 r^4) + [p_1 xy + p_2(r^2 + 2y^2)](1 + p_3 r^2),\end{aligned}\quad (3.8)$$

where m_i is the coefficient of radial distortions and p_i the coefficient of decentering distortions.

Step 4. Apply affine transformation to the coordinates (x, y) . Assuming that the pixel coordinate system is orthogonal, we get the pixel coordinates

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_u & 0 \\ 0 & f_v \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix} \right) + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}, \quad (3.9)$$

where f_u, f_v are the scale factors along the horizontal and vertical directions, respectively, and (u_0, v_0) is the principal point.

Therefore, we have a more general imaging model of fisheye cameras:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_u & 0 \\ 0 & f_v \end{bmatrix} \left((k_1 \theta + k_2 \theta^3) \begin{bmatrix} \cos(\varphi) \\ \sin(\varphi) \end{bmatrix} + \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix} \right) + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}. \quad (3.10)$$

Parameters in this model can be obtained by camera calibration. It appears rational that an elaborate model (3.10) with distortions (3.8) may perfectly model the real imaging process. However, we find that, just like the case of a pinhole camera [15], a more elaborate modeling would not help (negligible when compared with sensor quantization), but cause numerical instability. In addition, it makes the calibration of the system more complicated and brings in a very high computation cost, which most applications cannot afford currently. Therefore, we omit distortions and take the following model in practice:

$$\begin{bmatrix} u \\ v \end{bmatrix} = (k_1 \theta + k_2 \theta^3) \begin{bmatrix} f_u & 0 \\ 0 & f_v \end{bmatrix} \begin{bmatrix} \cos(\varphi) \\ \sin(\varphi) \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}. \quad (3.11)$$

3.2. The Model of the Visual Sun Compass

With the imaging model of fisheye cameras (3.11), the absolute direction of a point \mathbf{p} in the captured image can be determined from its pixel coordinates (u, v) . Firstly the normalized image coordinates can be expressed as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \left(\begin{bmatrix} f_u & 0 \\ 0 & f_v \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} u \\ v \end{bmatrix} - \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \right). \quad (3.12)$$

Then, the incidence angles (θ, φ) of the incoming ray in the visual sun compass coordinate system can be calculated with

$$\begin{aligned}\varphi &= \arctan 2(y, x), \\ k_1\theta + k_2\theta^3 &= \text{sqrt}(x^2 + y^2).\end{aligned}\quad (3.13)$$

If the optical axis upright points at the sky when taking images, which means the visual sun compass coordinate system is parallel with the ground coordinate system and the rotation matrix between them is a 3×3 identity matrix, angles (θ, φ) calculated by (3.13) are also the incidence angles in the ground coordinate system. Otherwise, the calculated (θ, φ) should be first transformed to the ground coordinate system. With an inclinometer or inertial sensor, the rotation matrix from the visual sun compass coordinate system to the ground plane coordinate system, \mathbf{R}_{f2g} , can be obtained. The unit vector indicating the incidence angle of an image point \mathbf{p} is denoted as $\hat{\mathbf{P}}^s = [\hat{x}^s \ \hat{y}^s \ \hat{z}^s]^T$ in the ground plane coordinate system, which satisfies

$$\begin{aligned}\hat{\mathbf{P}}^s &= [\hat{x}^s \ \hat{y}^s \ \hat{z}^s]^T = \left[\cos(\hat{\varphi}) \sin(\hat{\theta}) \ \sin(\hat{\varphi}) \sin(\hat{\theta}) \ \cos(\hat{\theta}) \right]^T \\ &= \mathbf{R}_{f2g} [\cos(\varphi) \sin(\theta) \ \sin(\varphi) \sin(\theta) \ \cos(\theta)]^T.\end{aligned}\quad (3.14)$$

The corresponding azimuth angle and elevation angle $(\hat{\theta}, \hat{\varphi})$ in the ground plane coordinate system can be expressed as:

$$\begin{aligned}\hat{\theta} &= \arccos(\hat{z}^s), \\ \hat{\varphi} &= \arctan 2\left(\frac{\hat{x}^s}{\sin(\hat{\theta})}, \frac{\hat{y}^s}{\sin(\hat{\theta})}\right).\end{aligned}\quad (3.15)$$

With (3.12)–(3.15), the incidence angle in the ground plane coordinate system of an image point can be calculated, where parameters of the fisheye lens, $(f_u, f_v, s, u_0, v_0, k_1, k_2)$, can be calibrated off-line in advance. The obtained azimuth angle $\hat{\varphi}$ corresponds to the φ shown in Section 2. Then, with local time and estimated longitude, the angle γ can be calculated. Finally, the absolute direction β can be derived with the obtained $\hat{\varphi}$ and γ , as shown in Section 2.

4. Detecting the Sun in the Image

To calculate the absolute direction with the proposed visual sun compass, the sun's image in captured images should be detected and measured. In practice, image noises may make the apparent outline of the sun's image greatly deviate from the real configuration, and outliers may generate false alarms with regard to the sun's image. Therefore, the conventional method, which takes the centroid of pixels whose intensity values are above a certain threshold as the image of the sun's center, may fail to find the real image of the sun's center. Because the image of a sphere under the unified imaging model is an ellipse and fisheye

lenses model (3.11) approximates the unified imaging model (3.3), we take the image of the sun as a blob. Then, use a modified maximally stable extremal region (MSER) algorithm, which can remove redundant blobs and accelerate the processing speed of MSER, to detect the blob. Finally, fit an ellipse with the blob's contour and take the ellipse's center as the image of the sun's center. The details are as follows.

Firstly, narrow down the threshold's varying range of the MSER and detect the sun's image. The MSER [16] was originally proposed to establish tentative correspondences between a pair of images taken from different viewpoints. Its principle can be summarized as follows [16]. For a gray-level picture, take all possible thresholds. The pixels whose intensity values are equal to or greater than the threshold t are taken as white, and the others are denoted as black. By increasing the threshold t continually, a series of binary pictures E_t would be seen, where the subscript t indicates the frame corresponding to a threshold t . Initially, a white frame would be obtained. Thereafter, black spots corresponding to local intensity minima would emerge and grow with increase in the threshold. Neighboring regions corresponding to two local minima will merge to a larger one gradually, and the last frame would be black. An MSER is a connected region in E_t with little size change across a range of thresholds.

By the original MSER method, not only the blob corresponding to the image of the sun but also some blobs corresponding to other regions would be found. Consider that the intensity value of the sun's image is very close to 255 because it is sufficiently bright. Let the threshold vary only in a small range $[255 - \delta_1, 255]$ when performing MSER detection, where δ_1 is a variable of limiting the varying range of thresholds. The computation cost can be dramatically reduced and those regions with great difference from the sun's image in intensity can be removed.

Secondly, remove the false alarms of the sun's image with the aspect ratio constraint. The sun's image in the visual sun compass is close to a circle. The constraint that the aspect ratio of the sun's image approximates to 1 can be used to further refine detected MSERs. By these modifications, only one or some nested MSERs can be obtained.

Thirdly, remove redundant nested MSERs and obtain the blob corresponding to the sun's image. In this step, the kernel MSER, which refers to the most stable one among nested MSERs, is proposed to remove redundant nested MSERs. For a sequence of nested MSERs, $Q_1, \dots, Q_{k-1}, Q_k, \dots, Q_m$, where $Q_{k-1} \subset Q_k$, the kernel MSER \hat{Q} satisfies

$$\hat{Q} : \min_{k=1}^m (\text{std}(I(q_k))), \quad q_k \in Q_k, \quad (4.1)$$

where $\text{std}(I(q_k))$ is the standard deviation of the intensity of all pixels q_k in MSER Q_k . The kernel MSER takes not only the maximally stable region's cardinality but also the statistical property of intensity into account.

Fourthly, prune contour points are contaminated by image noise and fit ellipse with the remaining points. Because the blob detected by the modified MSER is a continuous region, only few contiguous outliers and noises contaminate the sun's image. For further robust processing, an ellipse is fitted with points of the blob contour to describe the sun's image. A direct method to fit ellipse is the constrained least squares (LS) method [17]. In images captured by the visual sun compass, although contiguous outliers and contaminated points are within the detected blob, they are far away from the real geometric configuration and greatly change the shape of the sun's image. Therefore, the distance from the boundary point to the blob's center is taken as the measure to purify the data. Points whose distance is the



Figure 6: The prototype of the proposed visual sun compass.

least under a certain percentage are used to fit the ellipse, and other points which are far away from the geometric configuration are taken as outliers and pruned. This method is called the constrained LS with pruning.

Finally, use coefficients of the fitted ellipse to calculate its center. Then, take the center as the image of the sun's center to estimate the sun's azimuth and elevation angle.

5. Experiments

A prototype of the visual sun compass (as shown in Figure 6) was built, which comprised a Feith intelligent CANCam and a FUJION FE185C057HA-1 fisheye lens. A two-dimensional (2D) calibration board [15] was used to calibrate the visual sun compass. The calibrated parameters in (3.11) are $k_1 = 1.7941$, $k_2 = -0.0043$, $f_u = 166.1549$, $f_v = 167.1770$, $u_0 = 654.5771$, and $v_0 = 532.6644$. Experiments with this prototype were conducted to validate the proposed visual sun compass.

5.1. Detecting the Sun

Some images were captured with the optical axis of the sun compass's prototype upright pointing at the sky at different hours in the daytime from May to July. The sun was detected in these images with the original MSER method, the conventional method, and the proposed modified MSER method. Figure 7(a) shows an image captured by the visual sun compass. Figure 7(b) shows the blobs detected by the original MSER method. It can be seen that, with the original MSER method, some redundant blobs were detected and the image of the sun's center could not be uniquely determined. For the conventional method, the estimated center is far away from the real configuration as shown in Figure 7(c), because intensity values of outliers are close to those of the sun's image. By applying the proposed modified MSER and the constrained LS with pruning to this image, the obtained result is shown in Figure 7(d). It is evident that only one region is obtained and the difference between the fitted ellipse and the image of the sun is slight. These results prove the validity of the proposed method.

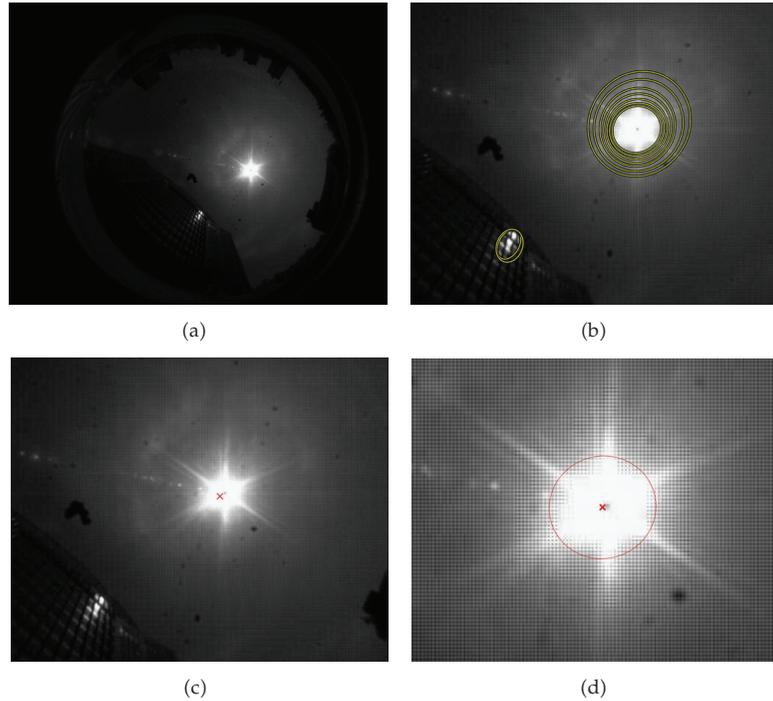


Figure 7: An image captured by the visual sun compass and experiment results. (a) Image captured by the visual sun compass. (b) Blobs detected by the original MSER. (c) The center detected by the conventional method. (d) The detected sun and its center by the proposed method.

Table 2: Orientation results with visual sun compass (in degrees).

Elevation θ		Angle φ	Azimuth γ		Orientation β	
G.T.	est.	est.	G.T.	est.	G.T.	est.
35.47	36.57	-87.64	-89.87	-89.87	0.0	-2.23
60.43	60.86	-125.10	62.92	62.92	185.0	188.02
58.10	59.92	-115.66	68.78	68.77	185.0	184.43

5.2. Orientation Estimation

By the proposed visual sun compass model, the elevation angle of the sun θ and “the angle of the style’s shadow” φ can be calculated with the calibrated parameters and the detected sun’s center. Table 2 shows some results, where “G.T.” denotes the ground truth and “est.” denotes the estimation. It is obvious that, for the estimated sun’s elevation angle θ , there is a little difference from the ground truth (values from the astronomic almanac corresponding to the time and the latitude and longitude information). Then, the azimuth angle γ of the sun and the absolute directions β in the ground plane coordinate system are calculated with the time, the latitude, and longitude information. Results are shown in Table 2. For the azimuth angle γ , the difference between the estimated values and the ground truth is very slight. The errors of the absolute direction with proposed method are less than 3.5° . They may arise from several sources. The primary source is the current rudimentary experimental setup: the visual sun compass is made horizontal using barely eye inspection with a spirit level, and the ground



Figure 8: The mobile robot mounted with the visual sun compass prototype and an IMU.

Table 3: Absolute orientation results with a mobile robot (in degrees).

G.T.	30.00	60.00	90.00	150.00	180.00
est.	28.08	59.63	92.75	152.31	180.77

truth of the estimated direction is obtained by eye inspection of a simple magnetic needle compass. Nonetheless, experiments show the validity of the proposed visual sun compass.

5.3. Application

We also applied the prototype of the visual sun compass on a mobile robot to carry out orientation estimation. Figure 8 shows the mobile robot, on which the visual sun compass prototype, a Crossbow VG700CB-200 IMU, and other sensors are mounted. The IMU can provide its own attitude relative to the ground coordinate system. The relative rotation between the IMU and the visual sun compass can be calibrated off-line in advance. Then, the rotation from the visual sun compass to the ground coordinate system, \mathbf{R}_{f2g} in (3.14), can be determined. Orientation experiments are conducted with this mobile robot to validate the proposed visual sun compass. Some results are shown in Table 3. With the help of the IMU, the deviations from the ground truth are less than those reported above. All of these results prove the validity of the proposed visual sun compass.

6. Conclusions

A visual sun compass is proposed in this paper. It is competent for orientation in environments with a tiny size and light weight, such as in aeronautical and space applications where the conventional sensors cannot function. A mathematical model and the absolute orientation method are presented. Further, a modified MSER and constrained LS with pruning are proposed to deal with severe distortion while detecting the sun in captured images. Real image experiments show the validity of the proposed visual sun compass. In comparison with conventional orientation sensors, the proposed visual sun compass can not only work in special environments but also provide the absolute orientation. The measurements of the proposed visual sun compass are not precise enough as yet. Future steps for improvement of the proposed visual sun compass include building a more precise

experiment platform, refining the measurements by finding a more precise calibration method, and analyzing the uncertainties of the projection of the sun's center.

Acknowledgments

This work was supported by the NSFC (Grant no.: 61101207, 60872127, and 61103130) and National Program on Key Basic Research Project (973 Program) (Grant no.: 2010CB731804-1, 2011CB706901-4).

References

- [1] M. Y. Kim and H. Cho, "An active trinocular vision system of sensing indoor navigation environment for mobile robots," *Sensors and Actuators, A*, vol. 125, no. 2, pp. 192–209, 2006.
- [2] S. Y. Chen, Y. F. Li, and J. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
- [3] K. M. Creer and M. Sanver, *Methods in Palaeo-Magnetism*, Elsevier, Amsterdam, The Netherlands, 1967.
- [4] W. Roellf and V. H. Bezooijen, "Star sensor for autonomous attitude control and navigation," in *Optical Technologies for Aerospace Sensing*, Proceedings of the SPIE, pp. 1–28, Boston, Mass, USA, November 1992.
- [5] F. Cozman and E. Krotkov, "Robot localization using a computer vision sextant," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 106–111, May 1995.
- [6] M. C. Deans, D. Wettergreen, and D. Villa, "A sun tracker for planetary analog rovers," in *Proceedings of the 8th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, pp. 577–583, September 2005.
- [7] A. Trebi-Ollennu, T. Huntsberger, Y. Cheng, E. T. Baumgartner, B. Kennedy, and P. Schenker, "Design and analysis of a sun sensor for planetary rover absolute heading detection," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 6, pp. 939–947, 2001.
- [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] 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.
- [10] J. Favill, *Primer of Celestial Navigation*, Mallock Press, 2007.
- [11] D. Schneider, E. Schwalbe, and H. G. Maas, "Validation of geometric models for fisheye lenses," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 3, pp. 259–266, 2009.
- [12] C. Bräuer-Burchardt and K. Voss, "A new algorithm to correct fish-eye- and strong wide-angle-lens-distortion from single images," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '01)*, pp. 225–228, October 2001.
- [13] C. Geyer and K. Daniilidis, "A unifying theory for central panoramic systems and practical implications," in *Proceedings of the European Conference on Computer Vision*, pp. 445–462, 2000.
- [14] X. Ying and Z. Hu, "Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model," in *Proceedings of the European Conference on Computer Vision*, pp. 442–455, 2004.
- [15] Z. Y. Zhang, "Camera calibration," in *Emerging Topics in Computer Vision*, G. Medioni and S. B. Kang, Eds., IMSC Press, 2004.
- [16] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proceedings of British Machine Vision Conference*, vol. 21, pp. 384–393, 2002.
- [17] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, 1999.

Research Article

Multifractals Properties on the Near Infrared Spectroscopy of Human Brain Hemodynamic

Truong Quang Dang Khoa and Vo Van Toi

Biomedical Engineering Department, International University of Vietnam National Universities, Block 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam

Correspondence should be addressed to Truong Quang Dang Khoa, khoa@ieee.org

Received 7 October 2011; Accepted 4 December 2011

Academic Editor: Carlo Cattani

Copyright © 2012 T. Quang Dang Khoa and V. Van Toi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nonlinear physics presents us with a perplexing variety of complicated fractal objects and strange sets. Naturally one wishes to characterize the objects and describe the events occurring on them. Moreover, most time series found in “real-life” applications appear quite noisy. Therefore, at almost every point in time, they cannot be approximated either by the Taylor series or by the Fourier series of just a few terms. Many experimental time series have fractal features and display singular behavior, the so-called singularities. The multifractal spectrum quantifies the degree of fractals in the processes generating the time series. A novel definition is proposed called full-width Hölder exponents that indicate maximum expansion of multifractal spectrum. The obtained results have demonstrated the multifractal structure of near-infrared spectroscopy time series and the evidence for brain imagery activities.

1. Introduction

Neurophysiological and neuroimaging technologies have contributed much to our understanding of normative brain function. Functional magnetic resonance imaging (fMRI) is currently considered the “gold standard” for measuring functional brain activation. The limitations of fMRI include the requirement that participants must lie within the confines of the magnet bore, which limits its use for many applications. The readout gradients in the imaging pulse sequences also produce a loud noise [1]. fMRI is also highly sensitive to movement artifact; subject movements on the order of a few millimeters can invalidate the data. And fMRI systems are quite expensive [2].

In recent years, functional near-infrared spectroscopy (NIRS) has been introduced as a new neuroimaging modality with which to conduct functional brain-imaging studies. NIRS technology uses specific wavelengths of light, introduced at the scalp, to enable

the noninvasive measurement of changes in the relative ratios of deoxygenated hemoglobin and oxygenated hemoglobin during brain activity. A wireless NIRS system consists of personal digital assistant software controlling the sensor circuitry, reading, saving, and sending the data via a wireless network. This technology allows the design of portable, safe, affordable, noninvasive, and minimally intrusive monitoring systems [3].

For such advanced features, NIRS signal processing really becomes an attractive field for computational science. Izzetoglu et al. investigated the canceling of motion artifact noise from NIRS signals by Wiener filter [4]. Izzetoglu et al. presented a statistical analysis of NIRS signals for the purpose of cognitive state assessment while the user performs a complex task [5]. The results indicated that the rate of change in the blood oxygenation of NIRS signals was significantly sensitive to task load changes and correlated fairly well with performance variables. Fantini et al. describe a specific frequency-domain instrument for near-infrared tissue spectroscopy. It has been proven that the hemodynamic changes monitored with NIR spectroscopy correlate with the activation state of the cortex in response to a stimulus [6, 7]. Sitaram et al. presented the results of signal analysis indicating that distinct patterns of hemodynamic responses exist that could be utilized in a pattern classifier [8].

Although there are many computing analyses on NIRS biomedical signals, there is not yet any work mentioning the aspects of NIRS physics. This paper continuously explores physical aspects of NIRS following a paper mentioned about nonlinear characteristics [9]. In this paper, we report an evidence for multifractality in biomedical NIRS signals and furthermore detect that the singularities indicate the state changes of brain activities. Since the conception of multifractal structure was first reported in 1986 by Halsey et al. [10], an approach based on wavelets transform was developed latter by Muzy et al. [11] and Mallat [12] called wavelet transform modulus maxima (WTMM). This theory has been greatly developed and applied in many study fields especially in biomedical researches. In 1999, Ivanov et al. [13] reported in *Nature* that multifractality is endogenous in healthy heartbeat dynamics both in awake and sleep states and thus does not depend on external factors such as levels of physical activities. In 2001, Amaral et al. [14] found the multifractal complexity of cardiac dynamics decreased or markedly lost when blocking the sympathetic or parasympathetic branch of the neuron autonomic system. Ohashi et al. [15] generalized WTMM in order to analyze positive and negative changes separately and show different singularity spectra depending on the direction of changes in human heartbeat interval data during sympathetic blockade, time series of daytime human physical activity of healthy individuals and daily stock price records. Shimizu et al. [16] investigated WTMM on functional magnetic resonance imaging (fMRI) time series to extract local singularity exponents to identify activated areas in human brain. In 2007, Yang et al. [17] distinguished among healthy people and heart diseased once by multifractal singularity spectrum area of synchronous 12-lead electrocardiogram (ECG) signals.

Although there are a lot of papers on the multifractality of the biological signals, there are few studies that clarify the reason of multifractal and the relation between the multifractality and biological functions.

During the last decades, a number of authors have claimed not only correlations between memory span and mental speed but also with electrophysiological and hemoglobin variables of brain waves. In [18], H. Weiss and V. Weiss determined the information entropy of working memory capacity. The congruence between multiples of memory span and multiples of a fundamental brain wave was the first important discovery. Relationships between different frequencies correspond to mechanisms designed to minimize interference, couple activity via stable phase interactions, and control the amplitude of one frequency

relative to the phase of another. These mechanisms are proposed to form a framework for spectral information processing [19]. In addition, we discovered the relationship between brain waves in motor imaging activities measured by NIRS and chaos properties in [20]. Furthermore, in this paper, we investigated WTMM to detect the singularities on NIRS time series. The obtained results indicate the task periods of brain activities. Furthermore, the parameters of WTMM models indicate physiological conditions in order to recognize left and right motor imagery tasks of human brain.

2. Methods

2.1. Wavelet Transform Modulus Maxima Method

2.1.1. Fractal Function

Self-affine functions are ones that are similar to themselves when transformed by anisotropic dilations. If $f(x)$ is a self-affine function, then $\forall x_0 \in \mathfrak{R}, \exists H \in \mathfrak{R}$ such that, for any $\lambda > 0$,

$$f(x_0 + \lambda x) - f(x_0) \approx \lambda^H (f(x_0 + x) - f(x_0)). \quad (2.1)$$

H is called the Hurst exponent. If $H < 1$, then f is not differentiable and the smaller H is the more singular f . Thus, H indicates the global irregularity or roughness of f . The fractal dimension of graph f is defined:

$$D_F = H - 2. \quad (2.2)$$

Fractal functions can possess multiaffine properties so that their roughness or the irregularity can fluctuate from point to point. Thus, the definition of the Hurst regularity becomes a local quantity of the velocity increment $\delta f(x_0 + l)$ around x_0 in the limit of inertial separation $l \rightarrow 0$:

$$f(x_0 + l) - f(x_0) \sim l^{h(x_0)}. \quad (2.3)$$

The local the Hurst exponent $h(x)$ is also called the Hölder exponent of f at point x . This is primarily related with the strength of the singularity of f at this point.

At any given point x_0 , the Hölder exponent is given by the largest exponent such that there exists a polynomial $P_n(x - x_0)$ of order $n < h(x_0)$ and a constant $C > 0$, so that, for any point x in the neighborhood of x_0 , the following relation holds.

$$|f(x) - P_n(x - x_0)| \leq C|x - x_0|^h. \quad (2.4)$$

$h(x_0)$ measures how irregular f is at x_0 . The higher the exponent $h(x_0)$, the more regular the function f .

In a signal with fractal features, an immediate question one faces is “how to quantify the fractal properties of such a signal?” The first problem is to find the set of locations of the singularities $\{x_i\}$ and to estimate the value of h for each x_i .

2.1.2. Using Wavelets Transform (WT) to Detect Singularities

WT is a space-scale analysis which consists in expanding signals in terms of wavelets which are constructed from a single function, the mother wavelet ψ , by means of translation and dilation. The WT of a real-valued function f is defined as

$$T_\psi[f](x_0, a) = \frac{1}{a} \int_{-\infty}^{+\infty} f(x) \psi\left(\frac{x - x_0}{a}\right) dx, \quad (2.5)$$

x_0 is the space parameter; a (>0) is the scale parameter.

The analyzing wavelet is usually well localized in both space and frequency. An interesting property of the wavelet transform is that the coefficients at these maxima are enough to encode the information contained in the signal. These maxima are defined, at each scale a , as the local maxima of $|T_\psi[f](x, a)|$. Moreover, as one follows a maxima line from the lowest scale to higher and higher scales, one is following the same singularity. This fact allows for the calculation of h_i by a power law fit to the coefficients of the wavelet transform along the maxima line.

The first possibility is that we find a single value $h_i = H$ for all singularities; the signal is then said to be monofractal. The second, more complex, possibility is that we find several distinct values for h ; the signal is then said to be multifractal.

2.1.3. Wavelet Transform Modulus Maxima Method (WTMM)

The term modulus maxima describes any point (x_0, a_0) such that $|T_\psi[f](x, a)|$ is locally maximum at $x = x_0$:

$$\frac{\partial T_\psi[f](x_0, a_0)}{\partial x} = 0. \quad (2.6)$$

This local maximum is a strict local maximum in either the right or the left neighborhood of x_0 . Maxima lines are called the connected curves of local maxima in the space-scale plane (x, a) along which all points are modulus maxima.

Let $\mathfrak{S}(a)$ be the set of all the maxima lines that exist at the scale a which contain maxima at any scale $a' \leq a$. A partition function is defined in terms of WT coefficients:

$$Z(q, a) = \sum_{l \in \mathfrak{S}(a)} \left(\sup_{\substack{(x, a') \in l \\ a' \leq a}} |T_\psi[f](x, a')| \right)^q. \quad (2.7)$$

The partition function Z measures the sum at a power q of all these wavelet modulus maxima. One can define the exponent $\tau(q)$ from the power-law behavior of partition function:

$$Z(q, a) \sim a^{\tau(q)}. \quad (2.8)$$

Thus, one can estimate $h(x_0)$ as the slope of log-log plot of Z versus scale a . The singularity spectrum can be determined from the Legendre transform of the partition function scaling exponent $\tau(q)$:

$$D(h) = \min_q (qh - \tau(q)), \quad (2.9)$$

where $h = \partial\tau/\partial q$.

A linear $\tau(q)$ curve indicates a homogenous fractal function. A nonlinear $\tau(q)$ curve indicates a nonhomogenous function exhibiting multifractal properties; that is, the Hölder exponent $h(x)$ depends on the spatial position x .

A novel definition is proposed in this paper called full-width the Hölder exponents that indicates maximum expansion of the Hölder exponents within spectrum $D(h)$. This parameter presents better separation of different multifractal time series:

$$f\omega H = \delta h = h_{\max} - h_{\min}. \quad (2.10)$$

3. Biomedical Time Series Acquisition

We used a multichannel NIRS instrument, OMM-3000, from Shimadzu Corporation, Japan, to acquire oxygenated hemoglobin and deoxygenated hemoglobin concentration changes. The system operated at three different wavelengths, 780 nm, 805 nm, and 830 nm, emitting an average power of $3 \text{ mW}\cdot\text{mm}^{-2}$. The illuminator and detector optodes were placed on the scalp. The detector optodes were fixed at a distance of 4 cm from the illuminator optodes. The optodes were arranged above the hemisphere on the subject's head.

Near-infrared rays leave each illuminator, pass through the skull and the brain tissue of the cortex, and are received by the detector optodes. The photomultiplier cycles through all the illuminator-detector pairings to acquire data at every sampling period. The data were digitized by the 16-bit analog-to-digital converter. Because oxygenated and deoxygenated hemoglobin types have characteristic optical properties in the visible and near-infrared light range, the change in concentration of these molecules during neurovascular coupling can be measured using optical methods. By measuring absorption changes at two (or more) wavelengths, one of which is more sensitive to Oxy-Hb and the other to Deox-Hb, changes in the relative concentrations of these chromophores can be calculated. Using these principles, researchers have demonstrated that it is possible to assess brain activity through the intact skull in adult humans.

The NIRS instrument was capable of storing the raw signals for each of the channels, one of which consists of the intensity values of 3 wavelengths, and also the derived values of oxygenated hemoglobin [Ox-Hb], deoxygenated hemoglobin [Deox-Hb], and total hemoglobin [Total-Hb] = [Ox-Hb] + [Deox-Hb] concentration changes for all time points in an output file in a prespecified format.

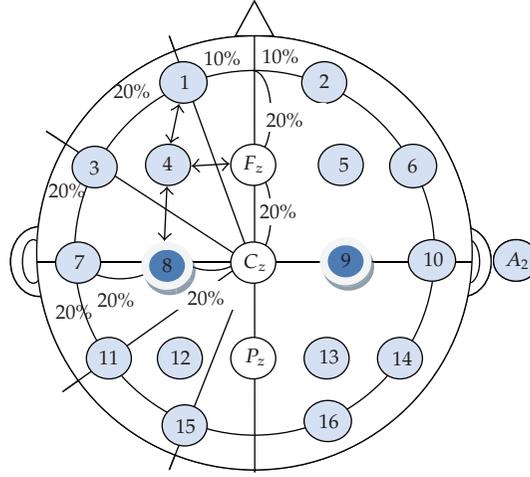


Figure 1: Measured positions based on international 10–20 system.

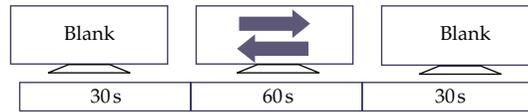


Figure 2: Experiment imagery moving tasks.

In this work, we investigate an experiment brain response on imagery moving tasks. The stimulus is a computer screen with arrows indicating left turn or right turn. The subject is a normal 30-year-old man measured during 2 mins, with the sampling time of 25 ms. In terms of optode placement, there is currently no standardized placement scheme for NIRS measurements. With such a standardized placement of electroencephalography (EEG), we have proposed 2 positions number 8 and number 9 in primary motor cortex of Brodmann's areas as shown in Figure 1 measuring left and right moving imagery tasks as shown in Figure 2.

4. Results and Discussion

This section included illustrated results in three tests, testing monofractal of fractional Brownian motion (fBM) signals, detecting singularities throughout artificial signals, and detecting singularities of real-life NIRS signals.

4.1. Testing Monofractal of Fractional Brownian Motion (fBM) Signals

Figure 3(a) displays one realization of a fractional Brownian with the Hurst exponent $H = 0.3$. The mother wavelet is chosen first derivative of Gaussian, and decomposition scale increases follow as exponent function, $a = 1.15^i$, $i = 0, \dots, N = 35$. Figure 3(b) gives the scaling exponent $\tau(q)$, which is nearly a straight line. Fractional Brownian motions are

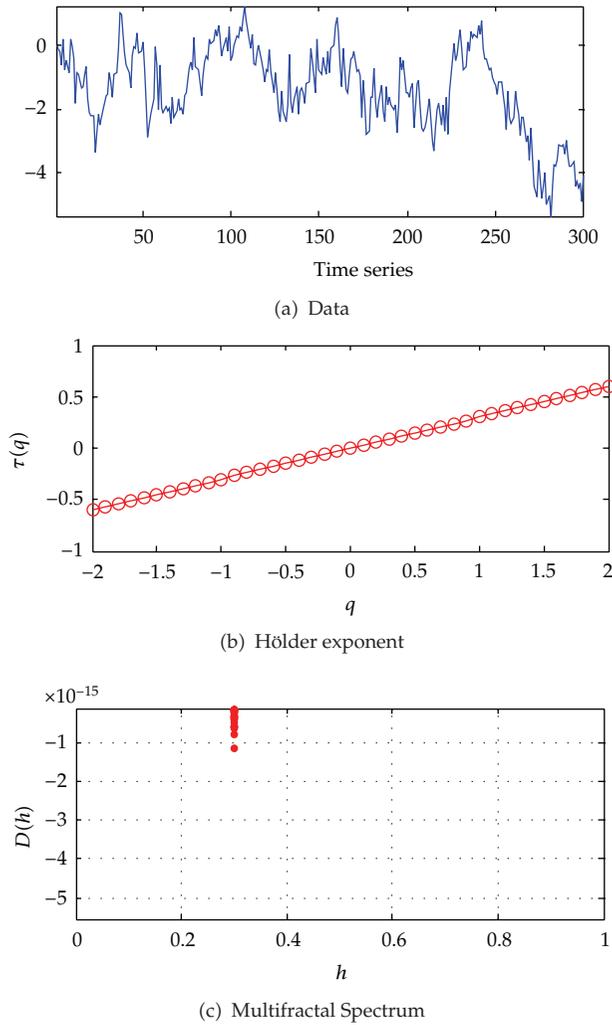


Figure 3: (a) A fractional Brownian with the Hurst exponent $H = 0.3$, (b) Scaling exponent $\tau(q)$ (c) Multifractal spectrum.

homogeneous fractals equal to H . The estimated spectrum in Figure 3(c) is calculated with a Legendre transform of $\tau(q)$. The theoretical spectrum $D(h)$ has therefore reduced to $\{0.3\}$.

4.2. Detecting Singularities of Multifractal Signal

Figure 4 clearly shows singularities detected by finding the abscissa where the wavelet modulus maxima is locally maximum. The mother wavelet is chosen first derivative of Gaussian, and decomposition scale increases follow as exponent function, $a = 1.15^i, i = 0, \dots, N = 35$. Figure 4(a) are original data taken from illustrated example of Matlab 1-dimension continuous wavelet analysis. Figures 4(b) and 4(d) are correspondent to the chains of local maxima and wavelet coefficients $|T_q[f](x, a)|$ at the maximum scale. It can be found

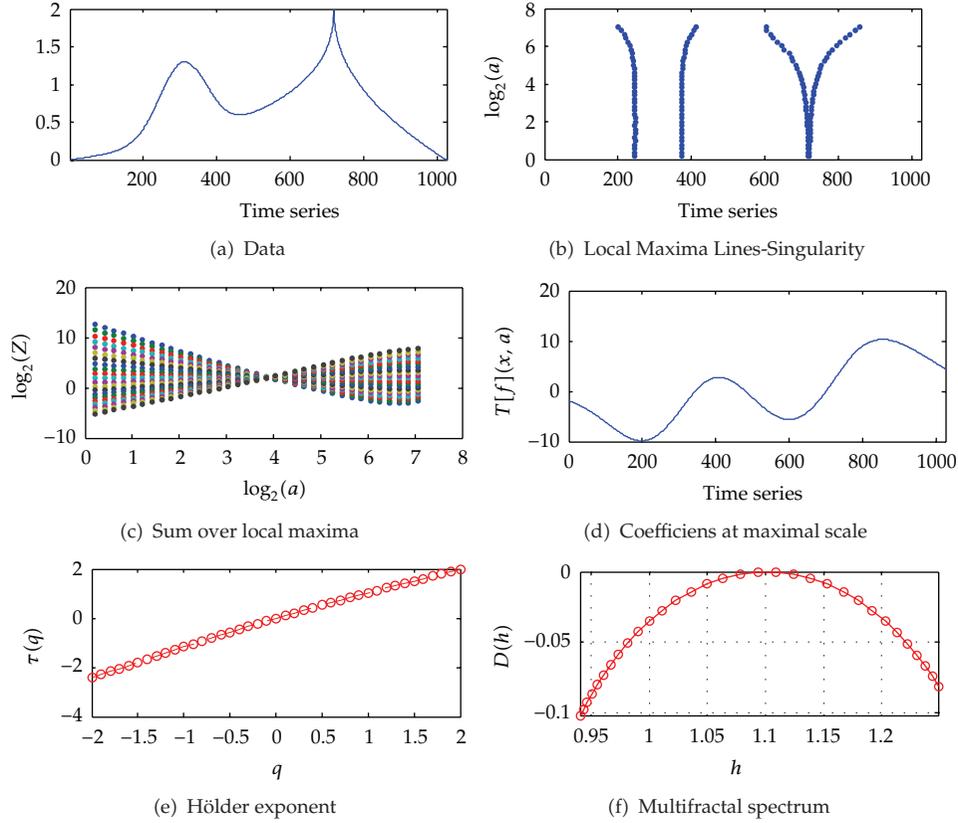


Figure 4: (a) Data testing singularities. (b) Local maxima line. (c) Partition functions. (d) Wavelet coefficients at the maximum scale. (e) Scaling exponents. (f) Multifractal spectrum.

that the beginning of chains is correspondent to local maxima of $|T_\psi[f](x, a)|$. The partition function Z measures the sum at a power $q = \{-2 : 2\}$ of all these wavelet modulus maxima shown in Figure 4(c). Figure 4(e) gives the scaling exponent $\tau(q)$, and spectrum $D(h)$ in Figure 4(f) indicates the signal is multifractal.

4.3. Relation between Multifractality and Biological Functions

The objective of this paper is detection of the singularities on NIRS time series and then finding the active periods of human brain. Figures 5 and 6 are correspondent to changes in concentrations of oxyhemoglobin (Oxy-Hb) and deoxy-hemoglobin (DeOxy-Hb) using the second derivative of Gaussian, and wavelet scale increases follow as exponent function, $a = 1.15^i, i = 0, \dots, N = 45$. Figure 5(b) shows all singularities, two of which reach to positive peaks of $|T_\psi[f](x, a)|$ shown in Figure 5(d) at which occur activities of brain. Concurrently Figure 6 displays two negative minima at the same positions. Only using characteristic points of interests, maxima and minima, as an extension of wavelet-based analysis of multifractal singularity, we can identify active periods of human brain. Furthermore, multifractal spectra shown in Figures 5(f) and 6(f) indicate NIRS is definitely multifractal time series. In near

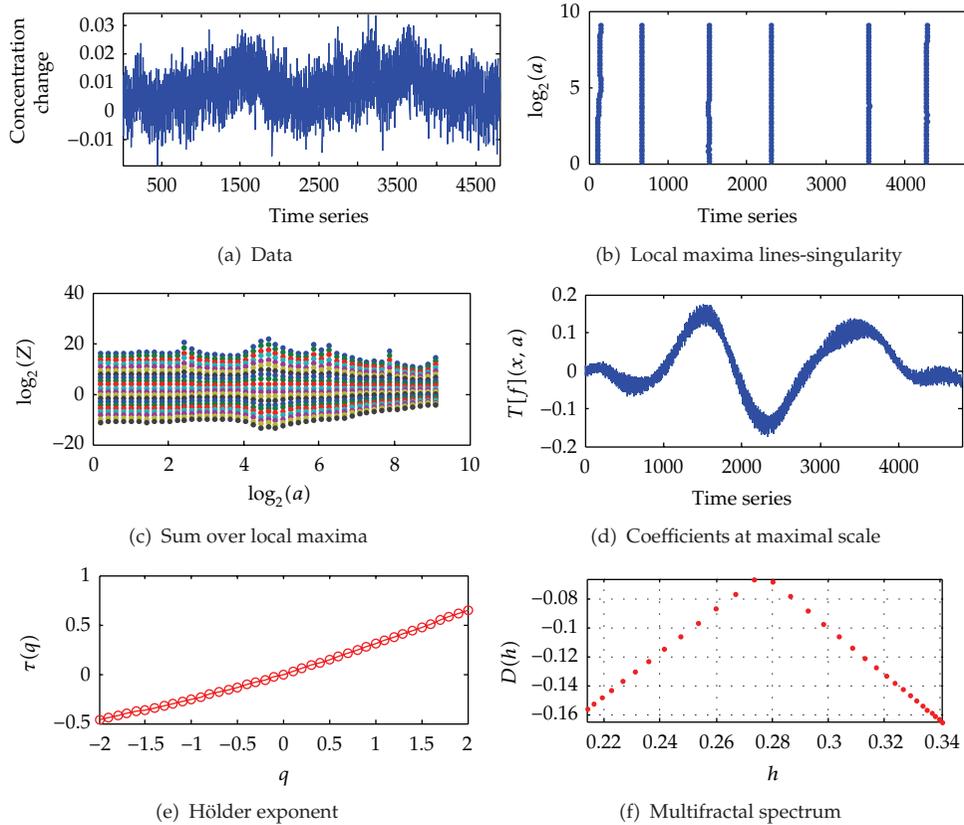


Figure 5: (a) Data of changes in concentrations of Oxy-Hb (b) Local maxima line (c) Partition functions (d) Wavelet coefficients at the maximum scale (e) Scaling exponents (f) Multifractal spectrum.

future, we believe that the results provide greater opportunities to identify the mechanisms responsible for complex biomedical systems.

Figure 7 shows evidences that the full-width Hölder exponents are clearly different corresponding to right-hand moving tasks. The fwH in these figures is average value of three trials for the same subject. In Figure 3, while brain implement to image the right-hand task, the measurements on the right side of head at the C3 position present the wide range of Hölder exponents. This indicates that multifractal behavior of right hand, channel 1, is stronger than that of left side of brain. The notice is completely right for all three acquisition data, (Ox-Hb), (Deox-Hb) and (To-Hb). The similar results of left-hand moving imagery are shown in Figure 8. The multifractal spectrum of the left-side measurements, channel 2, indicates a wide range of the Hölder exponents.

5. Conclusions

The advantages of NIRS are well demonstrated in many recent reports, although quantification of the changes of NIRS responses is still being developed. In the present paper, we have focused mainly on detection of multifractal characteristics of NIRS time series to identify the active-state period of human brain. Multifractal parameters are regarded as a

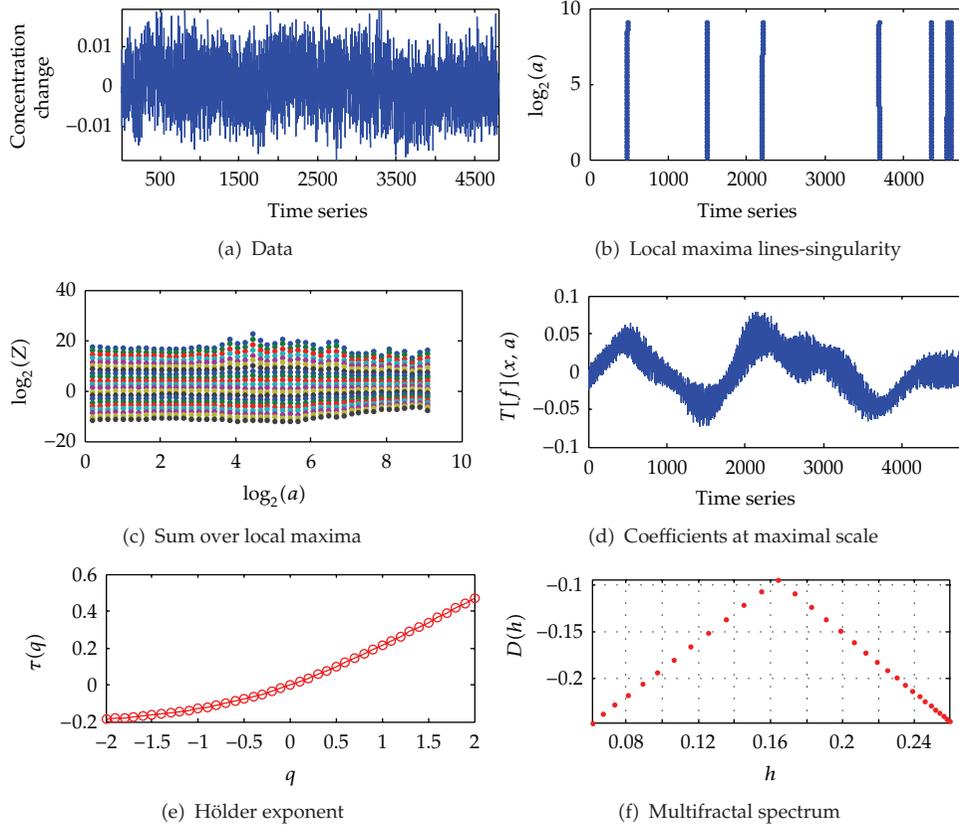


Figure 6: (a) Data of changes in concentrations of DeOxy-Hb. (b) Local maxima line. (c) Partition functions. (d) Wavelet coefficients at the maximum scale. (e) Scaling exponents. (f) Multifractal spectrum.

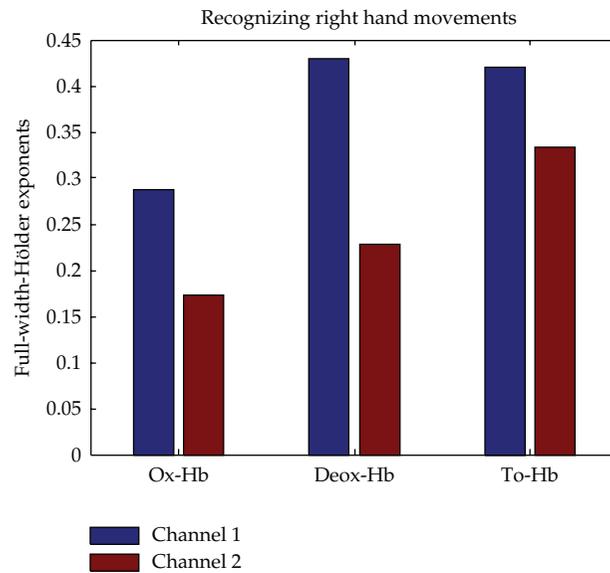


Figure 7: Average full-width Hölder exponents during right-hand motor imagery.

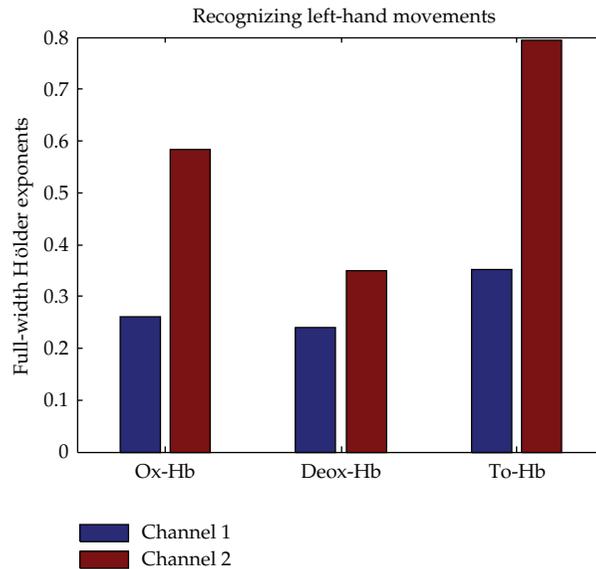


Figure 8: Average full-width-Hölder exponents during left hand motor imagery.

flexibility of human brain activities to understand brain activities. Different functional states of brain are probably governed by different degrees of multifractality. Further investigations into applications of NIRS signals could carry out meaningful contributions in medical and biological engineering.

Acknowledgements

The authors would like to acknowledge the Research Grant from the Vietnam National University in Ho Chi Minh City and the Vietnam National Foundation for Science and Technology Development (NAFOSTED) Research Grant no. 106.99-2010.11. Furthermore, They are deeply grateful for the support from our volunteers and friend.

References

- [1] M. E. Ravicz, J. R. Melcher, and N. Y.-S. Kiang, "Acoustic noise during functional magnetic resonance imaging," *Journal of the Acoustical Society of America*, vol. 108, no. 4, pp. 1683–1696, 2000.
- [2] S. C. Bunce, M. Izzetoglu, K. Izzetoglu, B. Onaral, and K. Pourrezaei, "Functional near-infrared spectroscopy," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 4, pp. 54–62, 2006.
- [3] M. Izzetoglu, K. Izzetoglu, S. Bunce et al., "Functional near-infrared neuroimaging," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 2, pp. 153–159, 2005.
- [4] M. Izzetoglu, A. Devaraj, S. Bunce, and B. Onaral, "Motion artifact cancellation in NIR spectroscopy using Wiener filtering," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 5, pp. 934–938, 2005.
- [5] K. Izzetoglu, S. Bunce, B. Onaral, K. Pourrezaei, and B. Chance, "Functional optical brain imaging using near-infrared during cognitive tasks," *International Journal of Human-Computer Interaction*, vol. 17, no. 2, pp. 211–227, 2004.
- [6] S. Fantini and M. A. Franceschini, "Frequency-domain techniques for tissue spectroscopy and imaging," in *Handbook of Optical Biomedical Diagnostics*, V. V. Tuchin, Ed., pp. 405–453, SPIE, Bellingham, Wash, USA, 2002.

- [7] A. Sassaroli, Y. Tong, F. Fabbri, B. Frederick, P. Renshaw, and S. Fantini, "Functional mapping of the human brain with near-infrared spectroscopy in the frequency-domain," in *Lasers in Surgery: Advanced Characterization, Therapeutics, and Systems XIV*, vol. 5312 of *Proceedings of SPIE*, pp. 371–377, San Jose, Calif, USA, January 2004.
- [8] R. Sitaram, H. Zhang, C. Guan et al., "Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface," *NeuroImage*, vol. 34, no. 4, pp. 1416–1427, 2007.
- [9] T. Q. D. Khoa, H. M. Thang, and M. Nakagawa, "Testing for nonlinearity in functional near-infrared spectroscopy of brain activities by surrogate data methods," *The Journal of Physiological Sciences*, vol. 58, no. 1, pp. 47–52, 2008.
- [10] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, "Fractal measures and their singularities: the characterization of strange sets," *Physical Review A*, vol. 33, no. 2, pp. 1141–1151, 1986.
- [11] J. F. Muzy, E. Bacry, and A. Arneodo, "Multifractal formalism for fractal signals: the structure-function approach versus the wavelet-transform modulus-maxima method," *Physical Review E*, vol. 47, no. 2, pp. 875–884, 1993.
- [12] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd edition, 1999.
- [13] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger et al., "Multifractality in human heartbeat dynamics," *Nature*, vol. 399, no. 6735, pp. 461–465, 1999.
- [14] L. A. N. Amaral, P. C. Ivanov, N. Aoyagi et al., "Behavioral-independent features of complex heartbeat dynamics," *Physical Review Letters*, vol. 86, no. 26, pp. 6026–6029, 2001.
- [15] K. Ohashi, L. A. N. Amaral, B. H. Natelson, and Y. Yamamoto, "Asymmetrical singularities in real-world signals," *Physical Review E*, vol. 68, no. 6, Article ID 065204, 2003.
- [16] Y. Shimizu, M. Barth, C. Windischberger, E. Moser, and S. Thurner, "Wavelet-based multifractal analysis of fMRI time series," *NeuroImage*, vol. 22, no. 3, pp. 1195–1202, 2004.
- [17] X. Yang, X. Ning, and J. Wang, "Multifractal analysis of human synchronous 12-lead ECG signals using multiple scale factors," *Physica A*, vol. 384, no. 2, pp. 413–422, 2007.
- [18] H. Weiss and V. Weiss, "The golden mean as clock cycle of brain waves," *Chaos, Solitons and Fractals*, vol. 18, no. 4, pp. 643–652, 2003.
- [19] A. K. Roopun, M. A. Kramer, L. M. Carracedo et al., "Temporal Interactions between Cortical Rhythms," *Frontiers in Neuroscience*, vol. 2, no. 2, pp. 145–154, 2008.
- [20] T. Q. D. Khoa, N. Yuichi, and N. Masahiro, "Recognizing brain motor imagery activities by identifying chaos properties of oxy-hemoglobin dynamics time series," *Chaos, Solitons and Fractals*, vol. 42, no. 1, pp. 422–429, 2009.

Research Article

Adaptive Control and Synchronization of the Shallow Water Model

P. Sangapate

Department of Mathematics, Faculty of Science, Maejo University, Chiang Mai 50290, Thailand

Correspondence should be addressed to P. Sangapate, prisayarat@mju.ac.th

Received 6 July 2011; Revised 24 September 2011; Accepted 26 September 2011

Academic Editor: Carlo Cattani

Copyright © 2012 P. Sangapate. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The shallow water model is one of the important models in dynamical systems. This paper investigates the adaptive chaos control and synchronization of the shallow water model. First, adaptive control laws are designed to stabilize the shallow water model. Then adaptive control laws are derived to chaos synchronization of the shallow water model. The sufficient conditions for the adaptive control and synchronization have been analyzed theoretically, and the results are proved using a Barbalat's Lemma.

1. Introduction

A dynamical system is a system that changes over time. Chaotic systems are dynamical systems that are highly sensitive to initial conditions. Chaos phenomena in weather models were first observed by Lorenz equation; a large number of chaos phenomena and chaos behavior have been discovered in physical, social, economical, biological, and electrical systems.

Atmosphere is a dynamical system. An atmospheric model is a set of equations that describes behavior of the atmosphere. The shallow water model is simple model for the atmosphere. Shallow water model is the set of the equations of motion that describes the evolution of a horizontal structure, hydrostatic homogeneous, and incompressible flow on the sphere [1].

The control of chaotic systems is to design state feedback control laws that stabilize the chaotic systems. Control theory is an interdisciplinary branch of engineering and mathematics that deals with the behavior of dynamical systems. The usual objective of control theory is to calculate solutions for the proper corrective action from the controller that result in system stability.

Synchronization of chaotic systems is phenomena that may occur when two or more chaotic oscillators are coupled or when a chaotic oscillator drives another chaotic oscillator, because of the butterfly effect, which causes the exponential divergence of the trajectories of two identical chaotic systems started with nearby the same initial conditions. Synchronizing two chaotic systems is seemingly a very challenging problem in chaos literature [2–6].

In 1990, Pecora and Carroll [7] introduced a method to synchronize two identical chaotic systems and showed that it was possible for some chaotic systems to be completely synchronized. From then on, chaos synchronization has been widely explored in variety of fields including physical system [8], chemical systems [9], ecological systems [10], secure communications [11, 12], and so forth.

In most of the chaos synchronization approaches, the drive-response formalism has been used. If a particular chaotic system is called the drive system and another chaotic system is called the response system, then the idea of synchronization is to use the output of the drive system to control the response system so that the output of the response system tracks the output of drive system asymptotically stable.

This paper is organized as follows. Section 2 gives notations and definitions of the stability in the chaotic system. Section 3 presents the adaptive control chaos of the shallow water model. Section 4 presents adaptive synchronization of the shallow water model. The conclusion discussion is in Section 5.

2. Notations and Definitions

X denotes an infinite dimensional Banach Space with the corresponding norm $\| \cdot \|$, R denotes the real line.

Consider a nonlinear nonautonomous differential equation of the general form

$$\begin{aligned} \dot{x}(t) &= f(t, x(t)), \quad t \geq t_0 \in R, \\ x(t_0) &= x_0, \end{aligned} \tag{2.1}$$

where the state $x(t)$ take values in X , $f(t, x) : R \times X \rightarrow X$ is a given nonlinear function and $f(t, 0) = 0$, for all $t \in R$. The stability conditions were proposed and presented in [13].

Definition 2.1. The zero solution of (2.1) is said to be *stable* if for every $\varepsilon > 0, t_0 \in R$, there exists a number $\delta > 0$ (depending upon ε and t_0) such that for any solution $x(t)$ of (2.1) with $\|x_0\| < \delta$ implies $\|x(t)\| < \varepsilon$, for all $t \geq t_0$.

Definition 2.2. The zero solution of (2.1) is said to be *asymptotically stable* if it is stable and there is a number $\delta > 0$ such that any solution $x(t)$ with $\|x_0\| < \delta$ satisfies $\lim_{t \rightarrow \infty} \|x(t)\| = 0$.

Consider the control system

$$\dot{x}(t) = f(t, x(t), u(t)), \quad t \geq 0, \tag{2.2}$$

where $u(t)$ is the external control input. The adaptive control is the control method to design state feedback control laws that stabilize the chaotic systems.

Definition 2.3. The control system (2.2) is *stabilizable* if there exists feedback control $u(t) = k(x(t))$ such that the system

$$\dot{x}(t) = f(t, x(t), k(x(t))), \quad t \geq 0, \quad (2.3)$$

is asymptotically stable.

Consider two nonlinear systems

$$\dot{x} = f(t, x(t)), \quad (2.4)$$

$$\dot{y} = g(t, y(t)) + u(t, x(t), y(t)), \quad (2.5)$$

where $x, y \in R$, $f, g \in C^r[R \times R, R]$, $u \in C^r[R \times R \times R, R]$, $r \geq 1$, R is the set of nonnegative real number. Assume that (2.4) is the drive system, (2.5) is the response system, and $u(t, x(t), y(t))$ is the control vector.

Definition 2.4. Response system and drive system are said to be *synchronic* if for any initial conditions $x(t_0), y(t_0) \in R$, $\lim_{t \rightarrow \infty} \|x(t) - y(t)\| = 0$.

Lemma 2.5 (Barbalat's lemma as used in stability). *For nonautonomous system,*

$$\dot{x}(t) = f(t, x(t)) \quad (2.6)$$

If there exists a scalar function $V(x, t)$ such that

- (1) V has a lower bound,
- (2) $\dot{V} \leq 0$,
- (3) $\dot{V}(x, t)$ is uniformly continuous in time,

then $\lim_{t \rightarrow \infty} \dot{V}(x, t) = 0$ by applying the Barbalat's Lemma to stabilize the chaotic systems.

3. Adaptive Control Chaos of the Shallow Water Model

A chaotic system has complex dynamical behaviors; those possess some special features, such as being extremely sensitive to tiny variations of initial conditions. In this section, adaptive control method is applied to control chaos shallow water model.

Shallow water model is the set of the equations of motion that describes the evolution of a horizontal structure, hydrostatic homogeneous, and incompressible flow on the sphere. Euler's equations of motion of an ideal fluid are as follows:

$$\begin{aligned} \frac{Du}{Dt} &= -\frac{1}{\rho} \frac{\partial p}{\partial x} + fv, \\ \frac{Dv}{Dt} &= -\frac{1}{\rho} \frac{\partial p}{\partial y} - fu, \\ \frac{Dw}{Dt} &= -\frac{1}{\rho} \frac{\partial p}{\partial z} - g, \end{aligned} \quad (3.1)$$

where ρ is the density of the fluid, p is the pressure, g is the gravity, and f is Coriolis parameter. Using the hydrostatic approximation,

$$\frac{\partial p}{\partial z} = -\rho g. \quad (3.2)$$

This implies $Dw/Dt = 0$. Assume the pressure p of fluid is constant, this implies that $\partial p/\partial t = 0$ and consider the continuity equation (or the incompressibility condition),

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \quad (3.3)$$

By solving for $\partial w/\partial z$ and integrating with respect to z , then w can be expressed as

$$\begin{aligned} \frac{\partial w}{\partial z} &= -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right), \\ w &= \int_0^h -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) dz = -h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right). \end{aligned} \quad (3.4)$$

The surface (of the fluid) boundary condition on w is that the fluid particles follow the surface (i.e., $Dh/Dt = w|_{\text{surface}}$). Thus

$$\frac{Dh}{Dt} = -h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right). \quad (3.5)$$

To get an expression for the pressure in the fluid, integrate the hydrostatic equation (3.2) from $p = 0$ at the top downward,

$$p(x, y, z) = \int_h^z -\rho g dz = (h - z)\rho g. \quad (3.6)$$

Take the partial derivatives of p (at the surface) with respect to x and y ,

$$\begin{aligned} \frac{\partial p}{\partial x} &= \frac{\partial}{\partial x} ((h - z)\rho g) = \rho g \frac{\partial h}{\partial x} \implies -\frac{1}{\rho} \frac{\partial p}{\partial x} = -g \frac{\partial h}{\partial x}, \\ \frac{\partial p}{\partial y} &= \frac{\partial}{\partial y} ((h - z)\rho g) = \rho g \frac{\partial h}{\partial y} \implies -\frac{1}{\rho} \frac{\partial p}{\partial y} = -g \frac{\partial h}{\partial y}. \end{aligned} \quad (3.7)$$

Taking (3.2)–(3.7) into (3.1), so the shallow water model in Cartesian coordinates is as follows:

$$\begin{aligned}\frac{Du}{Dt} &= -g\frac{\partial h}{\partial x} + fv, \\ \frac{Dv}{Dt} &= -g\frac{\partial h}{\partial y} - fu, \\ \frac{Dw}{Dt} &= -h\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right].\end{aligned}\quad (3.8)$$

In the vector form, the shallow water model is as follows:

$$\begin{aligned}\dot{V} &= -fk \times V - \nabla\Phi, \\ \dot{\Phi} &= -\Phi\nabla V,\end{aligned}\quad (3.9)$$

where $\mathbf{V} = u\vec{i} + v\vec{j}$ is the horizontal velocity, $\Phi = gh$ is the geopotential height.

Consider the controlled system of (3.9) which has the form

$$\begin{aligned}\dot{V} &= -fk \times V - \nabla\Phi + u_1, \\ \dot{\Phi} &= -\Phi\nabla V + u_2,\end{aligned}\quad (3.10)$$

where u_1, u_2 is external control input which will drag the chaotic trajectory (V, Φ) of the shallow water model to equilibrium point $E = (\bar{V}, \bar{\Phi})$ which is one of two steady states E_0, E_1 .

In this case the control law is

$$u_1 = -g(V - \bar{V}), \quad u_2 = -k(\Phi - \bar{\Phi}),\quad (3.11)$$

where k, g (estimate of k^*, g^* , resp.) are updated according to the following adaptive algorithm:

$$\begin{aligned}\dot{g} &= \mu(V - \bar{V})^2, \\ \dot{k} &= \rho(\Phi - \bar{\Phi})^2,\end{aligned}\quad (3.12)$$

where μ, ρ is adaption gains. Then the controlled systems have the following form:

$$\dot{V} = -fk \times V - \nabla\Phi - g(V - \bar{V}),\quad (3.13)$$

$$\dot{\Phi} = -\Phi\nabla V - k(\Phi - \bar{\Phi}).\quad (3.14)$$

Theorem 3.1. For $g < g^*, k < k^*$, the equilibrium point $E = (\bar{V}, \bar{\Phi})$ of the system (3.13), (3.14) is asymptotically stable.

Proof. Let us consider the Lyapunov function

$$V(\xi_1, \xi_2, \xi_3) = \frac{1}{2} \left[(V - \bar{V})^2 + (\Phi - \bar{\Phi})^2 + \frac{1}{\mu} (g - g^*)^2 + \frac{1}{\rho} (k - k^*)^2 \right]. \quad (3.15)$$

The time derivative of V is

$$\dot{V} = (V - \bar{V})\dot{V} + (\Phi - \bar{\Phi})\dot{\Phi} + \frac{1}{\mu} (g - g^*)\dot{g} + \frac{1}{\rho} (k - k^*)\dot{k}. \quad (3.16)$$

By substituting (3.13)-(3.14) in (3.16),

$$\begin{aligned} \dot{V} = & (V - \bar{V})[-fk \times V - \nabla\Phi - g(V - \bar{V})] + (\Phi - \bar{\Phi})[-\Phi\nabla V - k(\Phi - \bar{\Phi})] \\ & + \frac{1}{\mu} (g - g^*)\mu(V - \bar{V})^2 + \frac{1}{\rho} (k - k^*)\rho(\Phi - \bar{\Phi})^2. \end{aligned} \quad (3.17)$$

Let $\eta_1 = (V - \bar{V})$, $\eta_2 = (\Phi - \bar{\Phi})$. Since $(\bar{V}, \bar{\Phi})$ is an equilibrium point of the uncontrolled system (3.9), \dot{V} becomes

$$\begin{aligned} \dot{V} = & \eta_1[-fk \times V - \nabla\Phi - g(V - \bar{V})] + \eta_2[-\Phi\nabla V - k(\Phi - \bar{\Phi})] + (g - g^*)\eta_1^2 + (k - k^*)\eta_2^2 \\ = & (-fk \times V)\eta_1 - \nabla\Phi\eta_1 - g\eta_1^2 - \Phi\nabla V\eta_2 - k\eta_2^2 + (g - g^*)\eta_1^2 + (k - k^*)\eta_2^2. \end{aligned} \quad (3.18)$$

It is clear that if we choose $g < g^*$ and $k < k^*$, then \dot{V} is negative semidefinite. Since V is positive definite and \dot{V} is negative semidefinite, $\eta_1, \eta_2, g, k \in L_\infty$. From $\dot{V}(t) \leq 0$, we can easily show that the square of η_1, η_2 is integrable with respect to t , namely, $\eta_1, \eta_2 \in L_2$. From (3.13)-(3.14), for any initial conditions, we have $\eta_1, \eta_2 \in L_\infty$. By the well-known Barbalat's Lemma, we conclude that $\eta_1, \eta_2 \rightarrow (0, 0)$ as $t \rightarrow \infty$. Therefore, the equilibrium point $E = (\bar{V}, \bar{\Phi})$ of the system (3.13)-(3.14) is asymptotically stable. \square

4. Adaptive Synchronization of the Shallow Water Model

In this section, the adaptive synchronization is introduced to make two of the shallow water model. The sufficient condition for the synchronization has been analyzed theoretically, and the result is proved using a Barbalat's Lemma. Assume that there are two shallow water models such that the drive system is to control the response system. The drive and response system are given as

$$\begin{aligned} \dot{V} &= -f_1 k_1 \times V_1 - \nabla\Phi_1, \\ \dot{\Phi} &= -\Phi_1 \nabla V_1, \\ \dot{V} &= -f_2 k_2 \times V_2 - \nabla\Phi_2 - u_1, \\ \dot{\Phi} &= -\Phi_2 \nabla V_2 - u_2 \end{aligned} \quad (4.1)$$

where $u = [u_1, u_2]^T$ is the controller. We choose

$$\begin{aligned} u_1 &= k'_1 e_V, \\ u_2 &= k'_2 e_\Phi, \end{aligned} \quad (4.2)$$

where $k'_1, k'_2 \geq 0$ and e_V, e_Φ are the error states which are defined as follows

$$\begin{aligned} e_V &= V_2 - V_1, \\ e_\Phi &= \Phi_2 - \Phi_1. \end{aligned} \quad (4.3)$$

Theorem 4.1. Let $k_1, f_1, k'_1, k'_2 \geq 0$ be property chosen so that the following matrix inequalities holds:

$$P = \begin{pmatrix} k_1 f_1 + k'_1 & 0 \\ 0 & k'_2 \end{pmatrix} > 0, \quad (4.4)$$

then the two shallow water models (4.1) can be synchronized under the adaptive control (4.2).

Proof. It is easy to see from (4.1) that the error system is

$$\begin{aligned} \dot{e}_V &= -f_2 k_2 \times V_2 - \nabla \Phi_2 + f_1 k_1 \times V_1 + \nabla \Phi_1 - u_1, \\ \dot{e}_\Phi &= -\Phi_2 \nabla V_2 + \Phi_1 \nabla V_1 - u_2. \end{aligned} \quad (4.5)$$

Let $e_{kf} = k_2 f_2 - k_1 f_1$. Choose the Lyapunov function as follows:

$$V(t) = \frac{1}{2} [e_V^2 + e_\Phi^2]. \quad (4.6)$$

Then the differentiation of V along trajectories of (4.6) is

$$\begin{aligned} \dot{V}(t) &= e_V \dot{e}_V + e_\Phi \dot{e}_\Phi \\ &= e_V [-f_2 k_2 \times V_2 - \nabla \Phi_2 + f_1 k_1 \times V_1 + \nabla \Phi_1 - u_1] + e_\Phi [-\Phi_2 \nabla V_2 + \Phi_1 \nabla V_1 - u_2] \\ &= -e_V [f_2 k_2 \times V_2 + \nabla \Phi_2 - f_1 k_1 \times V_1 - \nabla \Phi_1 + u_1] - e_\Phi [\Phi_2 \nabla V_2 - \Phi_1 \nabla V_1 + u_2] \\ &= -e_V [f_2 k_2 \times V_2 - f_1 k_1 \times V_1 + f_1 k_1 \times V_2 - f_1 k_1 \times V_2] - e_V [\nabla \Phi_2 - \nabla \Phi_1] \\ &\quad - e_V u_1 - e_\Phi [\Phi_2 \nabla V_2 - \Phi_1 \nabla V_1 + \Phi_1 \nabla V_2 - \Phi_1 \nabla V_2] - e_\Phi u_2 \\ &= -e_V [e_{kf} \times V_2 + f_1 k_1 (V_2 - V_1)] - e_V \nabla (\Phi_2 - \Phi_1) - e_V k'_1 e_V \\ &\quad - e_\Phi [(\Phi_2 - \Phi_1) \nabla V_2 + \Phi_1 \nabla (V_2 - V_1)] - e_\Phi k'_2 e_\Phi \end{aligned}$$

$$\begin{aligned}
&= -e_V[e_{kf} \times V_2 + f_1 k_1 e_V] - e_V \nabla e_\Phi - e_V^2 k'_1 - e_\Phi[e_\Phi \nabla V_2 + \Phi_1 \nabla e_V] - e_\Phi^2 k'_2 \\
&= -e_V e_{kf} \times V_2 + f_1 k_1 e_V^2 - e_V \nabla e_\Phi - e_V^2 k'_1 - e_\Phi^2 \nabla V_2 - e_\Phi \Phi_1 \nabla e_V - e_\Phi^2 k'_2 \\
&\leq -f_1 k_1 e_V^2 - e_V^2 k'_1 - e_\Phi^2 k'_2 \\
&\leq -(f_1 k_1 + k'_1) e_V^2 - k'_2 e_\Phi^2 \\
&= -e^T P e,
\end{aligned} \tag{4.7}$$

where P is as in (4.4). Since $V(t)$ is positive definite and $\dot{V}(t)$ is negative semidefinite, it follows that $e_V, e_\Phi, k_1, f_1, k'_1, k'_2 \in L_\infty$. From $\dot{V}(t) \leq -e^T P e$, we can easily show that the square of e_V, e_Φ is integrable with respect to t , namely, $e_V, e_\Phi \in L_2$. From (4.5), for any initial conditions, we have $\dot{e}_{V(t)}, \dot{e}_{\Phi(t)} \in L_\infty$. By the well-known Barbalat's Lemma, we conclude that $(e_V, e_\Phi) \rightarrow (0, 0)$ as $t \rightarrow \infty$. Therefore, in the closed-loop system, $V_2(t) \rightarrow V_1(t), \Phi_2(t) \rightarrow \Phi_1(t)$ as $t \rightarrow \infty$. This implies that the two shallow water models have synchronized under the adaptive controls (4.2). \square

5. Conclusions

In this paper, we applied adaptive control theory for the chaos control and synchronization of the shallow water model. First, we designed adaptive control laws to stabilize the shallow water model based on the adaptive control theory and stability theory. Then, we derived adaptive synchronization to the shallow water model. The sufficient conditions for the adaptive control and synchronization of the shallow water model have been analyzed theoretically, and the results are proved using a Barbalat's Lemma.

References

- [1] J. Tribbia, *An Introduction to Three-Dimensional Climate Modeling*, Oxford University, New York, NY, USA, 1992.
- [2] E. Ott, C. Grebogi, and J. A. Yorke, "Controlling chaos," *Physical Review Letters*, vol. 64, no. 11, pp. 1196–1199, 1990.
- [3] J. H. Park and O. M. Kwon, "A novel criterion for delayed feedback control of time-delay chaotic systems," *Chaos, Solitons and Fractals*, vol. 23, no. 2, pp. 495–501, 2005.
- [4] T. L. Liao and S. H. Tsai, "Adaptive synchronization of chaotic systems and its application to secure communications," *Chaos, Solitons and Fractals*, vol. 11, no. 9, pp. 1387–1396, 2000.
- [5] T. Yang and L. O. Chua, "Generalized synchronization of chaos via linear transformations," *International Journal of Bifurcation and Chaos*, vol. 9, no. 1, pp. 215–219, 1999.
- [6] Y. G. Yu and S. Zhang, "Adaptive backstepping synchronization of uncertain chaotic system," *Chaos, Solitons and Fractals*, vol. 27, no. 3, pp. 643–649, 2004.
- [7] L. M. Pecora and T. L. Carroll, "Synchronization in chaotic systems," *Physical Review Letters*, vol. 64, no. 8, pp. 821–824, 1990.
- [8] M. Lakshmanan and K. Murali, *Chaos in Nonlinear Oscillators*, World Scientific Publishing, River Edge, NJ, USA, 1996.
- [9] S. K. Han, C. Kerrer, and K. Kuramoto, "Dephasing and bursting in coupled neutral oscillators," *Physical Review Letters*, vol. 75, pp. 3190–3193, 1995.
- [10] B. Blasius, A. Huppert, and L. Stone, "Complex dynamics and phase synchronization in spatially extended ecological systems," *Nature*, vol. 399, no. 6734, pp. 354–359, 1999.

- [11] K. Murali and M. Lakshmanan, "Secure communication using a compound signal from generalized synchronizable chaotic systems," *Physical Review Letters*, vol. 241, no. 6, pp. 303–310, 1998.
- [12] M. Feki, "An adaptive chaos synchronization scheme applied to secure communication," *Chaos, Solitons and Fractals*, vol. 18, no. 1, pp. 141–148, 2003.
- [13] V. Lakshmikantham, S. Leela, and A. A. Martynyuk, *Stability Analysis of Nonlinear Systems*, Marcel Dekker, New York, NY, USA, 1989.

Research Article

Photovoltaic Greenhouses: Comparison of Optical and Thermal Behaviour for Energy Savings

Maurizio Carlini, Tommaso Honorati, and Sonia Castellucci

CIRDER, Tuscia University, San Camillo de Lellis Street, 01100 Viterbo, Italy

Correspondence should be addressed to Tommaso Honorati, honorati@unitus.it

Received 29 September 2011; Accepted 15 November 2011

Academic Editor: Carlo Cattani

Copyright © 2012 Maurizio Carlini et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The production of energy from renewable sources, the diversification of the productive activities, and the development of photovoltaic technology and integrated systems have led to the development of solar greenhouses. The interest of the developers and designers is now to seek new approaches to combine the electricity and food production optimally. The interaction of factors as outside local climate, exposure, slope, soil, altitude, wind conditions, structural materials, or cultivated plant species, influences greatly the energy balance. This paper illustrates the comparison of optical and thermal behavior of a solar greenhouse and a similar glass greenhouse, devoted to the production of soil-less tomatoes in three different Italian areas, with computational aspects and methods of the TRNSYS simulation. Values of climatic parameters are obtained as a response for the feasibility of the cultivation under PV modules. The results show energy savings both for heating and cooling due to PV panels, adding a new reason for the realization of these systems.

1. Introduction

The development of solar greenhouses in Italy is due to the current trend to diversify agricultural production, energy efficiency, and farmer's specialization. In a greenhouse there are significant interactions between different factors of production, including the different structural type and climate along the Italian territory. For the various issues of this agroecosystem, a systematic and multidisciplinary approach is pursued in which the use of different available technologies can heal the conflict that still exists between food production, energy demand, environmental protection, and economic policy. In this context, an innovative strategy to reduce the impact of protected cultivation on the environment influence is one that aims to transform the emissions from the agricultural system "open" to a kind of "closed," essentially based on the reuse or reduction of waste material, and on the computerized monitoring of pests on the "soil-less" crop and the recycling of the nutrient solution.

Approximately 20–30% of the Italian greenhouses are equipped with heating and cooling systems. It was calculated that only the direct consumption of energy for the air conditioning goes whirling on the order of 140,000 TOE (tons of oil equivalent), approximately 90–95% of global energy demand for production.

However, the interaction of factors that influence the design and use of the greenhouse (outside local climate, exposure, slope, soil, altitude, wind conditions, type of glass and structural material used, cultivated plant species, etc.) influences greatly the energy balance [1]. Therefore, the focus of research and experimentation has been devoted to the study of variables in order to optimize the conditioner system [2].

To simulate a greenhouse has been proposed several studies to obtain values forecasts or simulations of influential variables for protected crops, such as ventilation, the water temperature for hydroponic systems, the control of CO₂ for carbon fertilization, the moisture budget, climate control, and heat exchange. Recently, the thermal behavior of the greenhouses was studied using dynamic thermal simulation tool TRNSYS 15.1.

Due to the actual strain of researching optimal solutions for the use of resources, it is important to create a model that includes all variable influential on greenhouse microclimate [3]. In this study, a simulation project with TRNSYS 17 software has been created for analyzing the optical and thermal behavior of a soil-less tomato crop. Values of climatic parameters are obtained comparing a glass greenhouse to a photovoltaic greenhouse in three different locations in Italy as a response for the feasibility of the cultivation under PV modules.

2. Materials and Methods

The study has been developed taking an existing greenhouse as reference and drawing it through Google SketchUp software. Then it was repeated for an analogue greenhouse with PV modules partially covered. Both the structures has been used for creating two different simulation projects considering soil-less tomato climate requirements. Projects relate to our work yet described [1], modified with new methods for solar radiation carried out with TRNSYS 17. Three different locations are considered, Turin, Rome, and Ragusa, representative, respectively for the North, Centre, and South Italy.

The choice of three different locations is due to the fact that in Italy the tomato is grown in all three areas where climatic conditions are significantly different. Soil-less tomato growing is a technique that can be carried out in all three environments and thus represents an essential link between farming and structural choices. Under these conditions the contribution of optical and thermal solar panels is really appreciated. Table 1 shows the boundary conditions for growing tomatoes.

2.1. Greenhouse Structural Description

The greenhouse considered is an Artigianfer type STO construction with steel structure prefabricated. It is covered with glass cover horizontal beam pattern and small flaps with north-south orientation.

It has a width of 25.60 m divided into two spans of 12.80 m. It is 150.107 m long and is divided into 39 sections ranging from 4.035 m. The eaves height is 4.60 m. In terms of structural elements, the greenhouse has cross doors.

Symmetric and transverse frames are stuck at the bottom and top. They are made with tubular columns 120 × 80 × 3 mm Fe 360 and horizontal beams lattice currents 80 × 40 × 3 mm

Table 1: Climatic requirements of the tomato considered in this study.

Minimum biological	°C	6
Maximum biological	°C	30
Thermal sums	°C	1800–2000
Optimal substrate	°C	15–20
Light intensity for maximum opening stomata	Klux	10
Minimum light intensity for starting photosynthesis	Klux	2
Maximum light intensity for starting photosynthesis	Klux	20
Minimum relative humidity	%	56
Maximum relative humidity	%	92
Daily optimal humidity	%	67
Nightly optimal humidity	%	84

Fe 430 tubular rods and rod wall. The roof rafters are made from the water canal collector and of pressed sheet metal. The side purlins are made of C-sections from 90 × 50 × 1,8 mm made of cool folded sheet. The glazing consists of rods 12 and 14 mm for roofs and walls. The calculation was performed in accordance with the requirements of the UNI-EN 13031-1 for greenhouses with metal structure. The maximum unit stress for steel Fe 360 of 1,600 kg/cm² for the first load cases and 1,800 kg/cm² for the other; for steel Fe 430 are of 1,900 kg/cm² for the first load cases and 2135 kg/cm² for the other. These have a corrosion protection due to the galvanizing bath.

The greenhouse consists of 8 very narrow aisles, each of 3.2 m, characterized by two sloping roof pitches of 22° degrees (40%) and exposed north-south.

The greenhouse is equipped with continuous full-stop driven by motors with rack system if the temperature inside the greenhouse exceeds a given temperature. This automated system therefore depends on measurement of a temperature sensor located near the slopes. Outside the building is also home to a wind instrument, in the case of strong wind forces the system automatically recloses.

2.2. Photovoltaic Greenhouse

The structure of the PV greenhouse is the same as the glass greenhouse with the difference that on south-facing slopes are placed photovoltaic modules, glass is used wholly within the aquifer north. The photovoltaic greenhouse modeled was designed and built through cooperation between Artigianfer and Isofotòn. It consists of a 246.16 kWp photovoltaic array that receives a fee of 0.43 €/kWh for the full architectural integration instead of glass on flap south.

The system consists of 1456 high-efficiency modules Isofoton IS-170/24 transparent laminates, unframed, allowing full integration in place of windows. The distance between cells, studied in the design stage, allows the passage of light, making possible the operation of nursery underlying coverage. Under the cover are positioned 36 inverter SMA Sunny Mini Central 7000 TL, placed on metal structures to improve the visual impact.

The PV modules produced by Isofoton are made with pseudoquadrate monocrystalline silicon cells high efficiency for energy conversion of solar radiation into DC electricity.

The cell circuit is laminated using EVA (ethylene-vinyl acetate) as encapsulating a complex of tempered glass on the front and a plastic polymer (TEDLAR) on the back,

Table 2: Details of a single photovoltaic module.

	Description
Cell type	Monocrystalline, textured, antireflective layer
Dimensions	125 mm × 125 mm
Number of cells per module	72 cells in series
Structure	(1) Tempered glass and microstructured high transmissibility (2) Cells laminated with EVA (ethylvinyl acetate) (3) Back-to-back tedlar/polyester layers

resistant to environmental agents and provided with electrical insulation. Details are shown in Table 2.

2.3. Greenhouse Model

Since TRNSYS 17 version three-dimensional geometry data created by Trnsys3d for TRNSYS can be imported into TRNBuild. Trnsys3d for TRNSYS is a plugin for Google SketchUp. This allows to create a building geometry from scratch: add zones, draw heat transfer surfaces, windows, shading surfaces, and so forth. The geometry data is divided into three groups.

- (i) Building geometry.
- (ii) External shading geometry.
- (iii) GeoPosition geometry: This data is used to define the position for comfort calculation or radiative gains.

For all surfaces of the zone three-dimensional data is provided due to importing an IDF file. For this geometry mode, detailed internal and external radiation modes have been selected. Different radiation modes for direct and diffuse shortwave radiation and longwave radiation distribution are available within a thermal zone.

2.3.1. Beam Radiation Distribution

For external windows, shading and insolation matrices are used to distribute the primary solar direct radiation entering the zone. These matrices are based on the three dimensional data of the building and shader surfaces. This mode is used for simulating highly glazed zones where the correct distribution of direct solar radiation is important. For generating shading/insolation matrices TRNBuild calls an auxiliary tool based on TRNSHD. TRNSHD subdivides the celestial hemisphere into patches based on the so-called Tregenza model. The resolution of the sky division can be set to medium (577 patches) or high (2305 patches). For each center point of a patch the sunlit fraction of external windows with three-dimensional data is calculated and saved into the shading matrix file. In addition, a diffuse radiation sunlit factor is calculated assuming an isotropic sky and written into an *.SHM file. If no external window is shaded no file is generated. In addition to sunlit fractions of external windows, TRNSHD calculates the beam sunlit fractions of the window that strike each inside surface of the zone. These distribution factors for all patches are calculated and written into a separate file (*.xxx.ISM).

2.3.2. Diffuse Radiation Distribution

For a detailed treatment of shortwave diffuse radiation including multireflection, the new radiation model applies so-called Gebhart factors. The key factor of this method is the view factor matrix. For generating the matrix TRNBuild calls an auxiliary program called TRNVFM. This mode is recommended for simulating highly glazed zones where diffuse radiation plays an important role.

2.3.3. Longwave Radiation Exchange of Surfaces within a Zone

For a detailed treatment of longwave radiation exchange including multireflection, the new radiation model applies so-called Gebhart factors. The key factor of this method is the view factor matrix. This mode is recommended for modeling the effect of low materials or detailed comfort analysis where the comfort depends on the location within the zone.

For calculating the view factor matrix TRNBuild calls an auxiliary tool called TRNVFM. For each zone with a detailed diffuse or longwave radiation mode the surface viewfactor matrix is calculated and written into a separate file (*.VFM). In addition, the view factor vector for comfort and radiative gain positions are determined and saved to the same file.

2.3.4. Direct Radiation

For a detailed treatment of shortwave beam radiation shading and distribution the multizone building model reads in the sunlit factor matrices generated by TRNBuild at the beginning of the simulation. For each time step the current sunlit fraction of surfaces are determined by a bilinear interpolation of the four nearest center points with respect to the Sun's actual position. The matrices are used for distributing primary beam radiation entering a zone through external windows.

2.3.5. Diffuse Radiation

In order to use a mathematical description similar to the longwave case, all surfaces are assumed to be transparent. That means solar radiation enters the zone from outside. The surfaces are not emitting radiation. They are assumed to be "passive" because they are only reflecting, absorbing, and/or transmitting solar radiation originating from outside of the zone. Based on this idea again a (solar) Gebhart matrix can be created. For opaque surfaces (walls) the transmitted diffuse solar radiation is zero.

The derivation of the describing equations for the detailed diffuse solar radiation heat transfer is based on the following assumptions.

- (1) All surfaces are assumed to be transparent (opaque surfaces are considered later).
- (2) Radiation leaving a surface is indicated by a positive sign of the corresponding heat flux.
- (3) ρ_{difsol} is the hemispherical solar reflectivity.

The Gebhart-factor $G_{\text{difsol},j \rightarrow k}$ for diffuse solar radiation is defined as the fraction of transmitted solar radiation through surface A_j that reaches A_k surface and is not reflected.

$G_{\text{difsol},j \rightarrow k}$ includes all the paths for reaching A_k , that is, the direct paths and paths by means of one or multiple reflections. The abbreviation “difsol” stands for diffuse solar radiation.

Using the assumptions from above, the (dimensionless) Gebhart matrix for diffuse solar radiation can be written as follows:

$$G_{\text{difsol}} = (I - F\rho_{\text{difsol}})^{-1}F(I - \rho_{\text{difsol}}), \quad (2.1)$$

where ρ_{difsol} is a diagonal matrix. The variable I describes the identity matrix, F again stands for the view factor matrix.

Using the auxiliary matrix with dimension $G_{\text{difsol}}^* [\text{m}_2]$ it can be shown that the net heat flux vector \dot{Q}_{difsol} for diffuse solar radiation in an enclosure is given by

$$\dot{Q}_{\text{difsol}} = G_{\text{difsol}}^* I_{\text{difsol}}, \quad (2.2)$$

with

$$G_{\text{difsol}}^* = -G_{\text{difsol}}^T A. \quad (2.3)$$

G_{difsol}^T is the transpose of G_{difsol} and A is the diagonal matrix describing the surface areas.

The driving force for diffuse solar radiation is the vector I_{difsol} with one component $I_{\text{difsol},k}$ for each surface A_k of the enclosure. $I_{\text{difsol},k}$ is equal to the transmitted part of the incident diffuse solar on the outside of surface A_k leaving the inner side. Walls are perfectly opaque to solar radiation and therefore $I_{\text{difsol},k}^{\text{wall}} = 0$. \dot{Q}_{difsol} is determined by a matrix multiplication with the driving force I_{difsol} .

2.3.6. Gebhart Method

The detailed model for describing the heat exchange driven by longwave radiation exchange and convection. In comparison to the standard model there is no artificial star node, because the longwave radiative heat transfer is treated separately. The derivation of the describing equations for the detailed longwave radiation heat transfer is based on the following assumptions.

- (1) Absorption of radiation on a surface is indicated by a negative sign of the corresponding heat flux whereas net emission means a positive heat flux.
- (2) All surfaces are isothermal.
- (3) All surfaces are perfect opaque for longwave radiation.
- (4) All surfaces are (diffuse) gray. This signifies that emissivity and absorptivity do not depend neither on wavelength nor on direction.
- (5) ρ_{ir} is the hemispherical longwave reflectivity.

The Gebhart-factor $G_{\text{ir},j \rightarrow k} [1, 2]$ is defined as the fraction of the emission from surface A_j that reaches surface A_k and is absorbed. $G_{\text{ir},j \rightarrow k}$ include all the paths for reaching A_k , that is, the direct paths and paths by means of one or multiple reflections. The abbreviation IR stands for “infrared,” meaning the longwave range of the radiation spectrum.

Using the assumptions from above, the (dimensionless) Gebhart matrix for longwave radiation can be written as follows:

$$G_{ir} = (I - F\rho_{ir})^{-1}F\varepsilon_{ir}, \quad (2.4)$$

where ρ_{ir} and ε_{ir} are diagonal matrices describing reflectivity and emissivity, respectively. The variable I describes the identity matrix. The view factor (written by F) is defined as the fraction of diffusely radiated energy leaving surface A that is incident on surface B .

Introducing the auxiliary matrix G_{ir}^* with dimension $[W/T^4]$, it can be shown that the net heat flux vector Q_{ir}^* longwave radiation in an enclosure is given by

$$\dot{Q}_{ir} = G_{ir}^* T^4, \quad (2.5)$$

with

$$G_{ir}^* = \left(I - G_{ir}^T \right) A \varepsilon \sigma. \quad (2.6)$$

T is the temperature vector in the enclosure, G_{ir}^T is the transpose of G_{ir} , σ the Stefan-Boltzmann constant, and A the diagonal matrix describing the surface areas. The auxiliary matrix G_{ir}^* only depends on optical (emissivity, reflectivity) and geometrical (view factor, area) properties as well as on the Stefan-Boltzmann constant. If this matrix is calculated first, the net heat flux of all surfaces in the enclosure can simultaneously be calculated by one simple matrix multiplication with the fourth power of the temperature vector.

3. Results

The project carried out with TRNSYS allows to extract all variables time-dependent, running simulations for hourly time periods established by the user, from a single hour to one year.

3.1. Simulation of Solar Radiation

Using the TRNSYS software has been run the simulation of the annual internal and external radiation in both types of structure and in the three selected locations. The results are shown in charts below (Figures 1, 2, and 3) where is possible to see the comparison between the annual internal and external solar radiation in both structures.

3.2. Simulation of Energy Demand

To compare energy consumption between PV greenhouse and glass greenhouse, the monthly energy consumptions were found for heating and cooling and compared to each other. These values were found for the three locations selected. The following tables (Tables 3, 4, and 5) show the monthly energy demand (kWh) for heating and cooling necessary for the soil-less tomato.

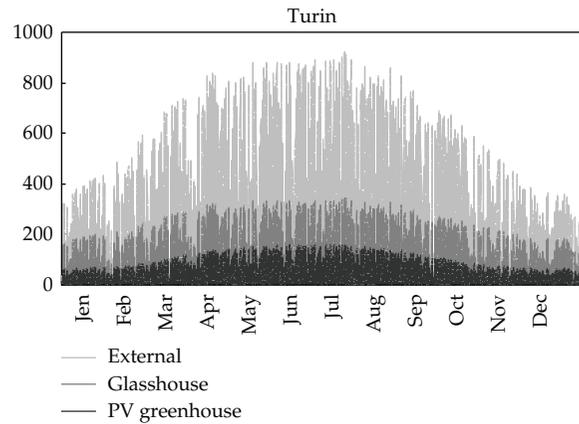


Figure 1: Hourly solar radiation (W/m^2) on external and internal soil, annual simulation for both structures (glass and PV greenhouse) in Turin.

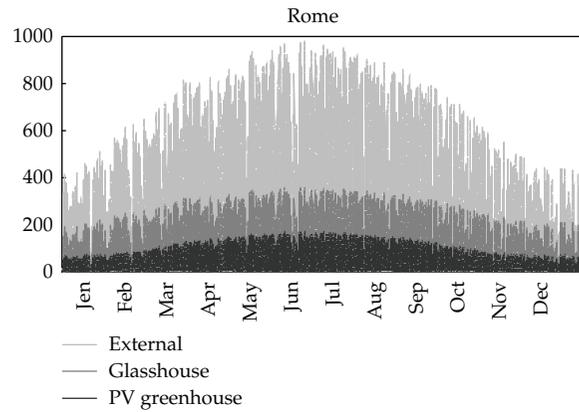


Figure 2: Hourly solar radiation (W/m^2) on external and internal soil, annual simulation for both structures (glass and PV greenhouse) in Rome.

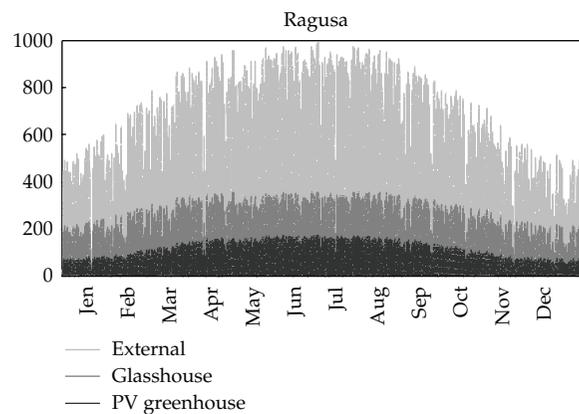


Figure 3: Hourly solar radiation (W/m^2) on external and internal soil, annual simulation for both structures (glass and PV greenhouse) in Ragusa.

Table 3: Monthly heating and cooling consumption (kWh) for soil-less tomato cultivated in PV and glass greenhouse located in Turin, with the difference (kWh) and the saving (variation percentage).

TURIN Month	Greenhouse		PV greenhouse		Difference		Variation %	
	Heating	Cooling	Heating	Cooling	Heating	Cooling	Heating	Cooling
January	53486	0	48518	0	4968	0	9.29	0.00
February	36776	0	32449	0	4327	0	11.77	0.00
March	11457	0	9725	0	1732	0	15.12	0.00
April	1658	0	850	0	807	0	48.70	0.00
May	0	20942	0	9810	0	11132	0.00	53.16
June	0	42804	0	27087	0	15717	0.00	36.72
July	0	60019	0	40326	0	19693	0.00	32.81
August	0	43565	0	25239	0	18326	0.00	42.07
September	0	16009	0	5149	0	10861	0.00	67.84
October	1276	0	939	0	337	0	26.40	0.00
November	21960	0	19382	0	2578	0	11.74	0.00
December	51478	0	45713	0	5764	0	11.20	0.00

Table 4: Monthly heating and cooling consumption (kWh) for soil-less tomato cultivated in PV and glass greenhouse located in Rome, with the difference (kWh) and the saving (variation percentage).

ROME Month	Greenhouse		PV greenhouse		Difference		Variation	
	Heating	Cooling	Heating	Cooling	Heating	Cooling	Heating	Cooling%
January	15774	0	13918	0	1856	0	11.77	0.00
February	8519	0	7755	0	765	0	8.98	0.00
March	2211	0	1723	0	488	0	22.06	0.00
April	179	13443	51	2904	128	10539	71.56	78.40
May	0	37213	0	20247	0	16966	0.00	45.59
June	0	58891	0	39235	0	19655	0.00	33.38
July	0	83357	0	57570	0	25787	0.00	30.94
August	0	70657	0	44271	0	26386	0.00	37.34
September	0	33263	0	14270	0	18993	0.00	57.10
October	14	8680	3	762	12	7918	81.17	91.22
November	3248	0	2581	0	667	0	20.53	0.00
December	11048	0	9891	0	1158	0	10.48	0.00

4. Discussion

The new version of software, TRNSYS 17, together with Google SketchUp simplifies and reduces the time for building projects and simulation to improve the quality of the information structures in high-glass surface. Simulations show that for glasshouses, the increased level of detail has a high impact on the results but increases input effort and computing time.

The simulations show that the solar radiation inside the greenhouse is on average half that of the solar greenhouse glass. This could lead to a reduction in photosynthetic efficiency of plants. It could therefore be interesting to develop a model that divides the beam of sunlight into ododa possible to assess which wavelengths are present in the structures.

Simulations of heating and cooling energy demand show how different are the energy consumptions in three locations. The integrated photovoltaic roof saves energy for air

Table 5: Monthly heating and cooling consumption (kWh) for soil-less tomato cultivated in PV and glass greenhouse located in Ragusa, with the difference (kWh) and the saving (variation percentage).

RAGUSA Month	Greenhouse		PV greenhouse		Difference		Variation %	
	Heating	Cooling	Heating	Cooling	Heating	Cooling	Heating	Cooling
January	577	0	517	0	60	0	10.33	0.00
February	149	0	111	0	38	0	25.47	0.00
March	164	7372	93	730	71	6643	43.23	90.10
April	0	18560	0	4821	0	13739	0.00	74.02
May	0	41945	0	24305	0	17641	0.00	42.06
June	0	60700	0	41294	0	19406	0.00	31.97
July	0	78880	0	54848	0	24032	0.00	30.47
August	0	72675	0	46982	0	25693	0.00	35.35
September	0	48951	0	25404	0	23547	0.00	48.10
October	0	20986	0	5563	0	15423	0.00	73.49
November	61	0	44	0	17	0	28.29	0.00
December	306	0	251	0	55	0	17.87	0.00

conditioning in all seasons: for all three locations considered it is on average of 30% for summer cooling and 11% for winter heating. In autumn and spring, these savings are much greater and in very different locations. This is due to temperatures close to those optimal for tomato and in some cases completely avoids the intervention of conditioning.

References

- [1] 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.
- [2] M. Carlini, D. Monarca, P. Biondi, T. Honorati, and S. Castellucci, "A simulation model for the exploitation of geothermal energy for a greenhouse in the viterbo province," in *Proceedings of the International Conference Ragusa SHWA, Work Safety and Risk Prevention in Agro-food and Forest Systems*, pp. 621–629, Ragusa Ibla Campus, Italy, 2010.
- [3] M. Carlini, D. Monarca, M. Cecchini et al., "Energetic and ergonomic aspects in the photovoltaic greenhouses," in *Proceedings of the International Conference Ragusa SHWA, Work Safety and Risk Prevention in Agro-food and Forest Systems*, pp. 576–629, Ragusa Ibla Campus, Italy, 2010.

Research Article

Distinguishing Stationary/Nonstationary Scaling Processes Using Wavelet Tsallis q -Entropies

**Julio Ramirez Pacheco,¹ Deni Torres Román,²
and Homero Toral Cruz³**

¹ Department of Basic Sciences and Engineering, University of Caribe, 74528 Cancún, QROO, Mexico

² Department of Electrical Engineering, CINVESTAV-IPN Unidad Guadalajara, 45010 Zapopán, JAL, Mexico

³ Department of Sciences and Engineering, University of Quintana Roo, 77019 Chetumal, QROO, Mexico

Correspondence should be addressed to Julio Ramirez Pacheco, jramirez@ucaribe.edu.mx

Received 22 July 2011; Revised 17 October 2011; Accepted 25 October 2011

Academic Editor: Carlo Cattani

Copyright © 2012 Julio Ramirez Pacheco et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Classification of processes as stationary or nonstationary has been recognized as an important and unresolved problem in the analysis of scaling signals. Stationarity or nonstationarity determines not only the form of autocorrelations and moments but also the selection of estimators. In this paper, a methodology for classifying scaling processes as stationary or nonstationary is proposed. The method is based on wavelet Tsallis q -entropies and particularly on the behaviour of these entropies for scaling signals. It is demonstrated that the observed wavelet Tsallis q -entropies of $1/f$ signals can be modeled by sum-cosh apodizing functions which allocates constant entropies to a set of scaling signals and varying entropies to the rest and that this allocation is controlled by q . The proposed methodology, therefore, differentiates stationary signals from non-stationary ones based on the observed wavelet Tsallis entropies for $1/f$ signals. Experimental studies using synthesized signals confirm that the proposed method not only achieves satisfactorily classifications but also outperforms current methods proposed in the literature.

1. Introduction

The theory of scaling processes has shown to be meaningful in several fields of applied science [1]. Some aspects of scaling behaviour have been reported in finance [2, 3], in the analysis of heart rate variability and EEGs in physiology [4, 5], in the characterization of mood and other behavioural variables in psychology [6], in the modelling of computer network traffic and delays in LANs and WANs [7, 8], and in the study of the velocity field of turbulent flows in turbulence [9–12] among others. The scaling signals studied in these fields can be modelled by a wide variety of stochastic processes, the majority of which are characterized by the single scaling index α or the associated Hurst index H . Theoretically, the

scaling index α determines not only the nature of the signal in terms of smooth, stationarity, nonstationarity, and correlations but also the selection of the methodologies employed to estimate α . The boundary $\alpha = 1$ is of special importance since processes for which $\alpha < 1$ are categorized as stationary in the sense that their statistical properties are invariant to translations and processes with $\alpha > 1$ are non-stationary. The stationary/non-stationary condition is fundamental for analysis and estimation purposes as many estimators have been devised for stationary signals while others have been formulated for non-stationary ones. The application of a non-stationary signal to an analysis/estimation technique designed to work in stationary conditions will result in an ambiguous analysis/estimate. For a review on the methodologies used to estimate α and the range of the scaling index over which they are applicable please refer to the work of Serinaldi [13], Malamud and Turcotte [14], and Gallant et al. [15]. In practice, a scaling process analyst does not know a priori the nature of the signal, and usually the estimation of α is performed somewhat arbitrarily without a stage of signal classification. Many articles in the literature, claiming that a given phenomena can be modelled by scaling signals, have performed the estimation of α without a phase of signal classification, and therefore their results remain questionable. Moreover, the scaling property of signals, in particular the long-memory characteristic, can also be caused by structural breaks in the mean, a common non-stationarity embedded in the signal [16, 17]. Because of this, the phase of signal classification is not only important but also necessary. The phase of signal classification was first recognized as important by Eke and coworkers in physiological signal analysis [4–6]. In their work, Eke et al. emphasized the importance of signal classification, the implications of omitting this phase, and the necessity of including this phase as a first step for enhancing the estimation and analysis of scaling signals. They claimed that by integrating this step in the traditional estimation and analysis methodologies, significant improvements can be achieved, and the possibility of misinterpreting the phenomena is decreased. Traditionally, signal classification has been performed by methods based on PSD [5, 6] on fractional Brownian motions, fBms, and fractional Gaussian noises, fGns. The characteristics of fBms and fGns, however, are visually different, and the signal classification process can even be performed by eye. Motivated by this, the present paper not only extends the results presented in [5] for the case of PPL signals but also proposes a methodology based on wavelet Tsallis q -entropies to differentiate scaling processes as stationary or non-stationary. The method is based on the observed sum-cosh window behaviour of these entropies which allocates constant entropies to stationary signals and varying entropies to non-stationary ones reducing the classification process to the constant/nonconstant character of the observed estimated entropies of the signals under study. Experimental and comparison studies not only confirm the capabilities of the method but also their advantages over standard methodologies based on PSD. The remainder of the paper is structured as follows. Section 2 provides a brief review of scaling processes, their definitions, and some standard results concerning its wavelet analysis. Section 3 describes wavelet entropies and its applications and derives the sum-cosh window behaviour observed in wavelet Tsallis q -entropies for signals with $1/f^\alpha$ behaviour. The techniques employed for classifying scaling signals as stationary or non-stationary as well as their advantages and disadvantages are briefly reviewed in Section 4 along with a description of the proposed methodology based on wavelet Tsallis q -entropies to discriminate signals. Section 5 describes the methodology used in the paper for testing the accuracy and robustness of the proposed technique and also details the comparative study to be performed among the techniques used for scaling signal classification. Section 6 presents the experimental results, and finally the conclusions of the paper are drawn in Section 7.

2. Wavelet Analysis of Scaling Processes

2.1. Scaling Processes

Scaling processes of parameter α , also called $1/f^\alpha$ or power-law processes, have been extensively applied and studied in the scientific literature since they model diverse phenomena [2, 3] within these fields. These processes are sufficiently characterized by the parameter α , called the scaling-index, which determines many of their properties. Various definitions have been proposed in the scientific literature, some based on their characteristics such as self-similarity or long memory, others based on the behaviour of their PSD. In this paper, a scaling process is a random process for which the associated PSD behaves as a *power law* in a range of frequencies [8, 9], that is,

$$S(f) \sim c_f |f|^{-\alpha}, \quad f \in (f_a, f_b), \quad (2.1)$$

where c_f is a constant, $\alpha \in \mathbb{R}$ the *scaling* index, and f_a, f_b represent the lower and upper bound frequencies upon which the power-law scaling holds. Depending upon f_a, f_b , and α , several particular scaling processes and behaviours can be identified. Independently of α , local regularity and band-pass power-law behaviour are observed whenever $f_a \rightarrow \infty$ and $f_b > f_a \gg 0$, respectively. When the scaling-index α is taken into consideration, long-memory behaviour is observed when both $0 < \alpha < 1$ and $f_b > f_a \rightarrow 0$. Self-similar features (in terms of distributional invariance under dilations) are observed in all the scaling-index range for all f . Scaling-index α determines not only the stationary and non-stationary condition of the scaling process but also the smoothness of their sample path realizations. The greater the scaling index α , the smoother their sample paths. As a matter of fact, as long as $\alpha \in (-1, 1)$, the scaling process is stationary (or stationary with long memory for small f and $\alpha \in (0, 1)$) and non-stationary when $\alpha \in (1, 3)$. Some transformations can make a stationary process appear non-stationary and vice versa. Outside the range $\alpha \in (-1, 3)$, several other processes can be identified, for example, the so-called extended fBms and fGns defined in the work of Serinaldi [13]. The persistence of scaling processes can also be quantified by the index α , and within this framework, scaling processes possess negative persistence as long as $\alpha < 0$, positive weak long persistence when $0 < \alpha < 1$, and positive strong long persistence whenever $\alpha > 1$. Scaling signals encompasses a large family of well-known random signals, for example, fBms, fGns [18], pure power-law processes [9], multifractal processes [8], and so forth. FBm, $B_H(t)$, comprises a family of Gaussian, self-similar processes with stationary increments, and because of the Gaussianity, it is completely characterized by its autocovariance sequence ACVS, which is given by

$$\mathbb{E}B_H(t)B_H(s) = R_{B_H}(t, s) = \frac{\sigma^2}{2} \left\{ |t|^{2H} + |s|^{2H} - |t - s|^{2H} \right\}, \quad (2.2)$$

where $0 < H < 1$ is the Hurst-index. FBm is non-stationary, and as such no spectrum can be defined on it; however, fBm possesses an average spectrum of the form $S_{\text{fBm}}(f) \sim c|f|^{-(2H+1)}$ as $f \rightarrow 0$ which implies that $\alpha = 2H+1$ [19]. FBm has been applied very often in the literature; however it is its related process, fGn, which has gained widespread prominence because of the stationarity of its realizations.

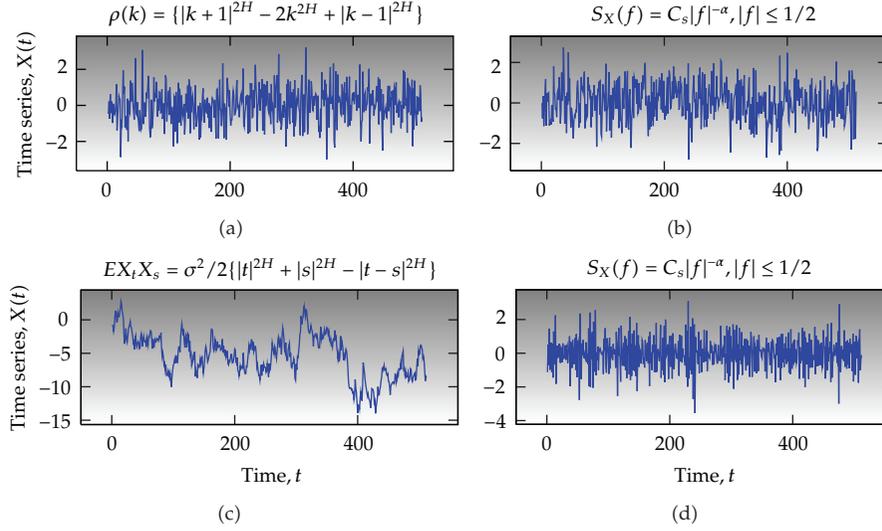


Figure 1: Sample path realizations of some scaling processes. Top left depicts a fGn with $\alpha = -0.1$, top right a PPL process with $\alpha = -0.1$, bottom left plot a fBm signal with $\alpha = 1.9$, and finally bottom right plot a PPL process with $\alpha = 1.9$.

fGn, $G_{H,\delta}(t)$, obtained by sampling a fBm process and computing increments of the form $G_{H,\delta}(t) = 1/\delta\{B_H(t+\delta) - B_H(t)\}$, $\delta \in \mathbb{Z}_+$ (i.e., by differentiating fBm), is a well-known Gaussian process. The ACVS of this process is given by

$$\mathbb{E}G_{H,\delta}(t)G_{H,\delta}(t+\tau) = \frac{\sigma^2}{2} \left\{ |\tau+\delta|^{2H} + |\tau-\delta|^{2H} - 2|\tau|^{2H} \right\}, \quad (2.3)$$

where $H \in (0, 1)$ is the Hurst-index. The associated PSD of fGn is given by [9]

$$S_{\text{fGn}}(f) = 4\sigma_X^2 C_H \sin^2(\pi f) \sum_{j=-\infty}^{\infty} \frac{1}{|f+j|^{2H+1}}, \quad |f| \leq \frac{1}{2}, \quad (2.4)$$

where σ_X is the process' variance and C_H is a constant. fGn is stationary and for large enough τ and under the restriction of $1/2 < H < 1$ possesses long-memory or long-range dependence (LRD). The scaling index α associated to fGn signals is given by $\alpha = 2H - 1$ as its PSD, given by (2.4), behaves asymptotically as $S_{\text{fGn}}(f) \sim c|f|^{-2H+1}$ for $f \rightarrow 0$. Another scaling process of interest is the family of discrete pure power-law processes (dPPL) which are defined as processes for which their PSD behaves as $S_X(f) = C_s|f|^{-\alpha}$ for $|f| \leq 1$, where $\alpha \in \mathbb{R}$ and C_s represent a constant. PPL signals are stationary when the power-law parameter $\alpha < 1$ and non-stationary whenever $\alpha > 1$. As stated in the work of Percival [9], the characteristics of these processes and those of fBms/fGns are similar; however, the differences between fBms and PPLs with $\alpha > 1$ are more evident. As a matter of fact, differentiation of stationarity/non-stationarity is far more difficult for PPL than for fBms/fGns. Figure 1 displays some realizations of fGn, fBm, and PPL processes. The scaling-index α of the PPL signals is identical to the scaling-index of the associated fGn and fBm. Note that the characteristics of the sample paths of fGn are fairly different from those of fBm. In the case of PPL processes, this differentiation

is not so evident, and as a matter of fact when the scaling indexes approach the boundary $\alpha = 1$, classification becomes complex. For further information on the properties, estimators, and analysis techniques of scaling processes please refer to [2, 3, 8–10, 13, 14].

2.2. Wavelet Analysis of Scaling Signals

Wavelets and wavelet transforms have been applied for the analysis of deterministic and random signals in almost every field of science [20–22]. The advantages of wavelet analysis over standard techniques of signal analysis have been widely reported and its potential for non-stationary signal analysis proven. Wavelet analysis represents a signal X_t in time-scale domain by the use of an analyzing or mother wavelet, $\psi_o(t)$ [23]. For the purposes of the paper, $\psi_o(t) \in L_1 \cap L_2$ and the family of shifted and dilated $\psi_o(t)$ form an orthonormal basis of $L_2(\mathbb{R})$. In addition, the finiteness of the mean average energy ($\mathbb{E} \int |X(u)|^2 du < \infty$) on the scaling process allows to represent it as a linear combination of the form:

$$X_t = \sum_{j=1}^L \sum_{k=-\infty}^{\infty} d_X(j, k) \psi_{j,k}(t), \quad (2.5)$$

where $d_X(j, k)$ is the DWT of X_t and $\{\psi_{j,k}(t) = 2^{-j/2} \psi_o(2^{-j}t - k), j, k \in \mathbb{Z}\}$ is a family of dilated (of order j) and shifted (of order k) versions of $\psi_o(t)$. The coefficients $d_X(j, k)$ in (2.5), obtained by DWT, represent a random process for every j , a random variable for fixed j and k , and as such many statistical analyses can be performed on them. Equation (2.5) represents signal X_t as a linear combination of L detail signals, obtained by means of the DWT. DWT is related to the theory of multiresolution signal representation (MRA), in which signals (or processes) can be represented at different resolutions based on the number of detail signals added to the low-frequency approximation signal. Detail random signals ($d_X(j, k)$) are obtained by projections of signal X_t into wavelet spaces \mathcal{W}_j , and approximation coefficients ($a_X(j, k)$) are obtained by projections of X_t into related approximation spaces \mathcal{U}_j . In the study of scaling processes, wavelet analysis has been primarily applied in the estimation of the wavelet variance [20, 24]. Wavelet variance or spectrum of a random processes accounts for computing variances of wavelet coefficients at each scale. Wavelet variance not only has permitted to propose estimation procedures for the scaling-index α but also to compute entropies associated with the scaling signals. Wavelet spectrum has also been used for detecting nonstationarities embedded in Internet traffic [20]. For stationary zero-mean processes, wavelet spectrum is given by

$$\mathbb{E} d_X^2(j, k) = \int_{-\infty}^{\infty} S_X(2^{-j}f) |\Psi(f)|^2 df, \quad (2.6)$$

where $\Psi(f) = \int \psi(t) e^{-j2\pi ft} dt$ is the Fourier integral of $\psi_o(t)$ and $S_X(\cdot)$ represents the PSD of X_t . Table 1 summarizes the wavelet spectrum for some standard scaling processes. For further details on the analysis, estimation, and synthesis of scaling processes please refer to the works of Abry and Veitch [23] and Bardet [25] and references therein.

Table 1: Wavelet spectrum or wavelet variance associated with different types of *scaling* processes. $\mathbb{E}(\cdot)$, $\text{Var}(\cdot)$, and $\Psi(\cdot)$ represent expectation, variance, and Fourier integral operators, respectively.

Type of scaling process	Associated wavelet spectrum or variance
Long-memory process of index α	$\mathbb{E}d_X^2(j, k) \sim 2^{j\alpha}C(\psi, \alpha)$, $C(\psi, \alpha) = c_\gamma \int f ^{-\alpha} \Psi(f) ^2 df$
Self-similar process of index H	$\mathbb{E}d_X^2(j, k) = 2^{j(2H+1)} \mathbb{E}d_X^2(0, k)$
Hssi process of index H	$\text{Var} d_X^2(j, k) = 2^{j(2H+1)} \text{Var} d_X(0, 0)$
Pure power-law process of index α	$\mathbb{E}d_X^2(j, k) = C2^{j\alpha}$

3. Wavelet-Based q -Entropies

The concept of entropy has traditionally been employed to measure the information content of random signals and systems [26, 27]. Recently, entropic functionals, such as Shannon, Rényi, and Tsallis, have been extensively applied to quantify the complexities associated with random and nonlinear phenomena [28]. Information planes, which consist of the product of positive measures of entropic functional and the Fisher information (and also of entropy/disequilibrium product), are now being applied in numerous systems (e.g., atomic, molecular, geophysical, etc.). Entropic quantities involve the calculation of functionals on probability densities or probability mass functions (pmf). Depending upon the domain in which the pmfs are obtained, entropies usually inherit their name. Entropies are called spectral entropies when entropic functionals are applied to pmfs derived from the Fourier spectrum representation of the process. When the densities are determined in the time-scale domain by discrete wavelet transformations, the associated entropy functionals are called wavelet entropies [29, 30]. If the pmf is obtained via the continuous wavelet transform, CWT, the entropy is called continuous multiresolution entropy (CMqE) [31]. Wavelet Shannon entropy, tantamount to computing a Shannon entropy functional on a *pmf* derived from the wavelet variance, has found applications in event-related potentials in neuroelectrical signals [32, 33], structural damage identification [34], segmentation of EEG signals [35], characterization of complexity in random signals [36–38] among others. Entropic measures of order q (hereafter q -entropies) generalize Shannon entropy and provide the flexibility of fine tuning to a desired behaviour with the value of q . The pmf in time-scale domain for which all entropies in this paper are computed is obtained by

$$p_j = \frac{1/N_j \sum_k \mathcal{F}(d_X(j, k))}{\sum_{i=1}^{\log_2(N)} \{1/N_i \sum_k \mathcal{F}(d_X(j, k))\}}, \quad (3.1)$$

where $\mathcal{F}(\cdot)$ represents the variance or second-order moment of the $d_X(j, k)$, N_j (resp., N_i) stands for the total number of wavelet coefficients at scale j (resp., i), and N is the length of the process. For signals with $1/f$ PSD, the so-called wavelet spectrum-based pmf is determined by direct substitution of the wavelet spectrum of the process under study (see Table 1) into (3.1), which results in

$$p_j = 2^{(j-1)\alpha} \frac{1 - 2^\alpha}{1 - 2^{\alpha M}}, \quad (3.2)$$

where $M = \log_2(N)$. The density given in (3.2) represents the probability that the energy of the scaling signal is located at scale j . The pmf of (3.2) can be used to compute numerous information theoretic functionals such as entropies, Fisher information, and information planes. Zunino and coworkers computed Shannon entropy functional on (3.2) and called it wavelet entropy. Later explicit formulas for wavelet Rényi and Tsallis entropies were derived and some applications suggested. Normalized Shannon entropy functional of scaling signals is given by

$$\widehat{\mathcal{H}}(p) = \frac{1}{\log_2(M)} \left\{ \frac{\alpha}{1-2^{-\alpha}} - \frac{\alpha M}{1-2^{-\alpha M}} - \log_2 \left(\frac{1-2^{-\alpha}}{1-2^{-\alpha M}} \right) \right\}, \quad (3.3)$$

where $M = \log_2(N)$. Wavelet Rényi q -entropies, as in the case of Shannon entropies, are extensive entropies in the sense that for any two independent random variables X_1 and X_2 , the joint entropy $\mathcal{H}_q^R(X_1, X_2) = \mathcal{H}_q^R(X_1) + \mathcal{H}_q^R(X_2)$. For scaling signals, Rényi entropy functional, $\widehat{\mathcal{H}}_q^R(p) = 1/(1-q)\log_2(\sum_j p_j^q)$, results in

$$\widehat{\mathcal{H}}_q^R(p) = \frac{q}{1-q} \left(\log_2 \left(\frac{1-2^{-\alpha}}{1-2^{-\alpha M}} \right) - \frac{1}{q} \log_2 \left(\frac{1-2^{-\alpha q M}}{1-2^{-\alpha q}} \right) \right), \quad (3.4)$$

where $q \in \mathbb{R}$ denotes the extensivity parameter. *Tsallis* q -entropies are nonextensive entropies in the sense that the extensivity property no longer holds. For a pmf p_j it is defined as

$$\widehat{\mathcal{H}}_q^T(p) = - \sum_{j=1}^M p_j^q \ln_q(p_j), \quad (3.5)$$

where $\ln_q(x) := (x^{1-q} - 1)/(1-q)$ is the q -logarithm function and $q \in \mathbb{R}$ the nonextensivity parameter. Tsallis entropies provide a valuable and interesting tool for the analysis of systems with long-range interactions, long memories, and so forth. The application of Tsallis entropies is vast, from the characterization of complexities in EEG signals [28] to the study of non-linear systems [31]. Normalized Tsallis functional applied to (3.2) results in wavelet Tsallis q -entropies which is given by

$$\widehat{\mathcal{H}}_q^T(p; \alpha) = c_{M,q} \left\{ 1 - \left(\frac{1-2^{-\alpha}}{1-2^{-\alpha M}} \right)^q \left(\frac{1-2^{-\alpha q M}}{1-2^{-\alpha q}} \right) \right\} \quad (3.6)$$

$$= c_{M,q} \left\{ 1 - \left(\frac{\sinh(\alpha \ln 2/2)}{\sinh(\alpha \ln 2M/2)} \right)^q \left(\frac{\sinh(\alpha q M \ln 2/2)}{\sinh(\alpha q \ln 2/2)} \right) \right\} \quad (3.7)$$

$$= c_{M,q} \left\{ 1 - \frac{P^{M-1} (2 \cosh(\alpha q \ln 2/2))}{(P^{M-1} (2 \cosh(\alpha \ln 2/2)))^q} \right\}, \quad (3.8)$$

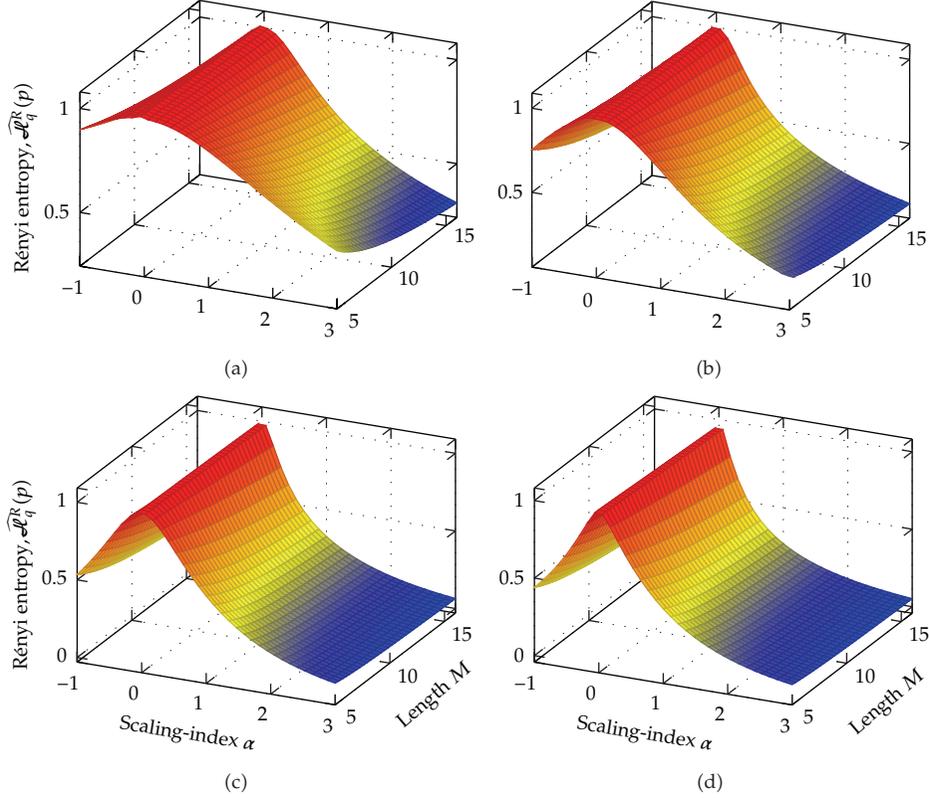


Figure 2: Wavelet Rényi q -entropies of $1/f$ signals. Top left plot computed with $q = 0.4$, top right plot with $q = 1.1$, bottom left plot with $q = 4$, and finally bottom right plot with $q = 15$.

where $c_{M,q} = 1/(1 - M^{1-q})$ is a normalizing factor and $P^{M-1}(2 \cosh u)$ is a polynomial of order $M - 1$, that is,

$$\begin{aligned}
 P^{M-1}(\cdot) &= (2 \cosh u)^{M-1} - \frac{(M-2)}{1!} (2 \cosh u)^{M-3} \\
 &+ \frac{(M-3)(M-4)}{2!} (2 \cosh u)^{M-5} - \dots .
 \end{aligned} \tag{3.9}$$

Figure 2 displays the wavelet Rényi q -entropies of scaling processes. Note that independently of signal length, wavelet Rényi entropies display a bell-shaped form for these processes. Parameter q stretches the bell-shaped form as q is varied. Parameter q has, in view of these entropy planes, no effect on the form of the observed entropies as the bell-shaped form is maintained. The maximum entropy is achieved when the scaling process is a pure white noise ($\alpha = 0$), and as the process becomes non-stationary their entropies decrease. The form and behaviour of these entropies are similar as those observed in the literature [32] and reflect the extensivity character of the entropy functionals. Note that both Shannon and Rényi entropies describe appropriately the complexities associated to $1/f$ processes: maximum for highly disordered systems and minimum for smooth signals. For further information on wavelet entropy please refer to [30, 32].

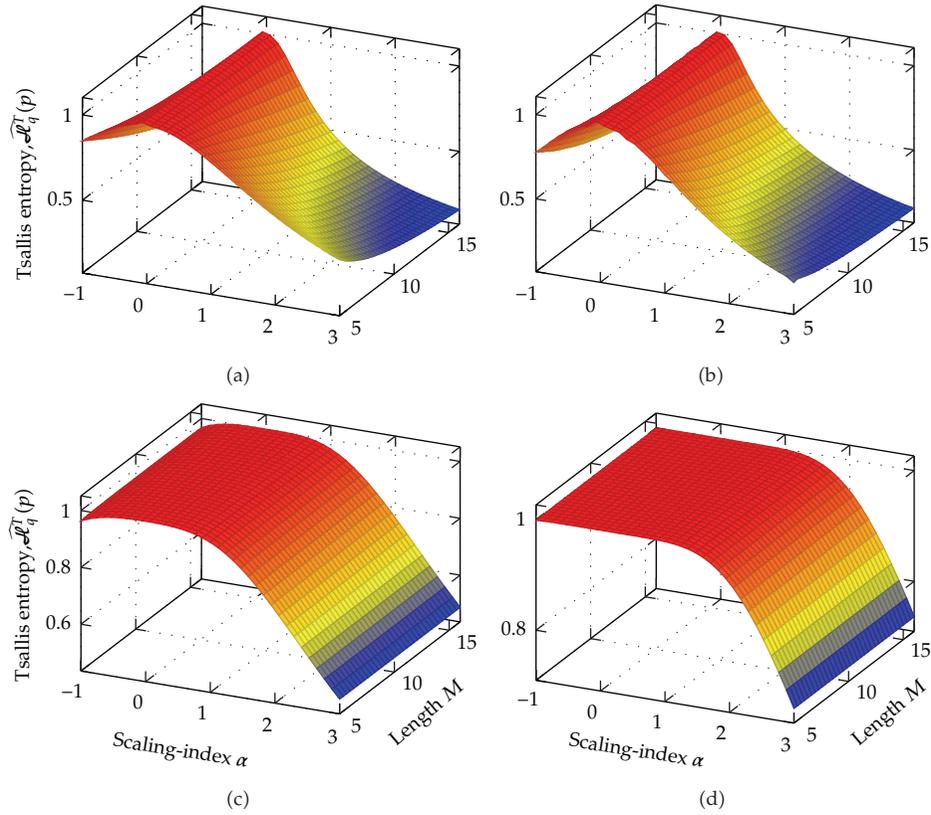


Figure 3: Wavelet Tsallis q -entropies of $1/f$ signals. Top left plot computed with $q = 0.4$, top right plot with $q = 0.95$, bottom left plot with $q = 5$, and finally bottom right plot with $q = 10$.

Figure 3 illustrates the wavelet Tsallis q -entropies for scaling processes of parameter α . Note that as long as $q < 5$, wavelet Tsallis q -entropies are identical as those observed in wavelet Rényi q -entropies (i.e., they have the same bell-shaped form). As long as $q \geq 5$, the behaviour of wavelet Tsallis q -entropies changes and differs from that of Shannon and Rényi. Wavelet Tsallis q -entropies, therefore, comprise the behaviour of wavelet Shannon and Rényi and provide greater flexibility in describing the process. Observe that, unlike Rényi entropies, Tsallis q -entropies allocate maximum (and constant) entropies to a set of scaling processes. In addition, the set of scaling signals for which this constant behaviour is observed is controlled by the nonextensivity parameter q of Tsallis entropies. The constant behaviour observed means that wavelet Tsallis q -entropies regard some set of scaling processes as totally random or disordered and, in some sense, randomizes the scaling signal under study. This particular behaviour of wavelet Tsallis q -entropies is further explored in next section, and a model for this is derived.

3.1. Sum-Cosh Window Behaviour of Wavelet Tsallis q -Entropies

Wavelet Tsallis q -entropies allocates constant entropies to a set of scaling processes and varying entropies to the rest. This particular behaviour can be modelled by the theory of

windowing or apodizing functions. This theory has been important for the design of digital filters; however, in this paper it is used to model the observed wavelet Tsallis q entropies of $1/f^\alpha$ signals. As a matter of fact, (3.8) resembles in some sense the cosh window observed in the work of Avci and Nacaroglu [39] and can be regarded as a sum-cosh window provided $\widehat{\mathcal{H}}_q^T(p; \alpha) = \widehat{\mathcal{H}}_q^T(p; -\alpha)$, $\lim_{\alpha \rightarrow 0} \widehat{\mathcal{H}}_q^T(p; \alpha) = 1$, and $\lim_{\alpha \rightarrow b} \widehat{\mathcal{H}}_q^T(p; \alpha) = 0$ conditions are satisfied. The symmetry condition ($\widehat{\mathcal{H}}_q^T(p; \alpha) = \widehat{\mathcal{H}}_q^T(p; -\alpha)$) is easily verified, and the $\lim_{\alpha \rightarrow 0} \widehat{\mathcal{H}}_q^T(p; \alpha)$ is computed based on the observation that

$$\mu_\alpha = \left(\frac{1 - 2^\alpha}{1 - 2^{\alpha M}} \right)^q = \left(\frac{-\alpha \ln 2 - (\alpha \ln 2)^2/2! - \dots}{-\alpha M \ln 2 - (\alpha M \ln 2)^2/2! - \dots} \right)^q, \quad (3.10)$$

which results in $\lim_{\alpha \rightarrow 0} \mu_\alpha = M^{-q}$. A similar reasoning for the expression to the right of μ_α in (3.6) results in

$$\lim_{\alpha \rightarrow 0} \widehat{\mathcal{H}}_q^T(p; \alpha) = c_{M,q} \left\{ 1 - (M^{-1})^q M \right\} = 1. \quad (3.11)$$

Derivation of the second limit is performed by means of the asymptotic relation $1 - 2^\alpha \approx -2^\alpha$ for large α , consequently

$$\lim_{\alpha \rightarrow b} \widehat{\mathcal{H}}_q^T(p; \alpha) \approx c_{M,q} \left\{ 1 - \left(\frac{2^\alpha}{2^{\alpha M}} \right)^q \left(\frac{2^{\alpha M q}}{2^{\alpha q}} \right) \right\} = 0, \quad (3.12)$$

as $b \gg 1$. The above demonstrates that wavelet Tsallis q -entropies can be modelled by sum-cosh windowing functions which in turn implies that for particular q , rectangular-like behaviour can be observed. The quasirectangular behaviour implies that constant regions of entropies are observed for a range of scaling processes and varying for the rest. The set of scaling processes for which constant wavelet Tsallis q -entropies are observed is controlled by the non-extensivity parameter q of Tsallis entropies. Figure 4 displays the shape of the wavelet Tsallis q -entropies for fixed length and different values of the non-extensivity parameter. For the cases $q = 0.999$ and $q = 3$, a bell-shaped form is observed which in some sense is identical to the ones observed for wavelet Shannon and Rényi entropies. Note from the figure that as $q = 8$, constant entropies are assigned to scaling processes in a symmetric range of α . This quasirectangular form can be set up to allocate constant entropies to stationary scaling processes and varying entropies to non-stationary ones. As a matter of fact, constant wavelet Tsallis entropies can be obtained for stationary signals and varying entropies to non-stationary ones as long as $q \approx 8$. This behaviour is important since a potential application of this feature is on the classification of scaling processes as stationary or non-stationary.

4. Classification of Scaling Signals

The classification of scaling signals as stationary or non-stationary has already been recognized as an important and unresolved problem in many areas of signal analysis [5, 12, 40–42]. Signal classification not only enhances the estimation process (i.e., estimation of the scaling index α) but also provides a correct interpretation of the phenomena, which in turn eases the application of a given technique in the process under study. Much of the literature

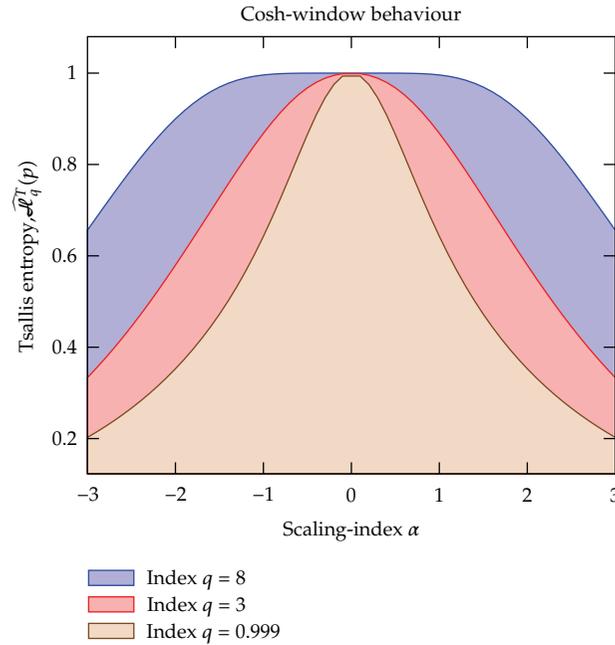


Figure 4: Cosh-window modelling of wavelet Tsallis q -entropies. Variation of window form on q .

on self-similar, long-memory, and fractal processes lack a step of signal classification, the parameters were estimated under the assumption of stationary, and therefore their results remain questionable. The process of signal classification becomes harder as we approach the boundary of stationarity and non-stationarity, that is, when $\alpha \rightarrow 1$. The reason for this is that as $\alpha \rightarrow 1$, stationary signals incorporate some features of non-stationarity and viceversa. Signal classification techniques often fail to distinguish fractal noises from motions within this boundary. The signal classification phase is sometimes more straightforward in some families of scaling signals than in others. For example, fBms and fGns are visually different, and the classification is simpler than that for the case of PPLs which are more difficult to classify. In this respect, any signal classification procedure must differentiate scaling signals independently of signal family and also provide meaningful classifications in the boundary $\alpha = 1$. Classification of scaling signals has traditionally been accomplished by using standard methodologies based on the PSD. PSD and PSD-based signal summation conversion (SSC) were recently proposed as methodologies for distinguishing fractal noises and motions in [5] by using synthesized fBms and fGns. In that work, fGns and fBms were generated in the interval $\alpha \in (-1, 3)$ and with sufficiently large lengths. The present paper proposes a methodology based on wavelet Tsallis q -entropies and its sum-cosh window behaviour. In the following, we briefly review current techniques employed to perform the signal classification phase and describe the proposed methodology based on wavelet Tsallis q -entropies.

4.1. Power Spectral Density

Spectral density function (SDF) characterizes stationary random signals in frequency domain. According to the work of Eke and coworkers [5, 41], SDF can be used to classify $1/f^\alpha$

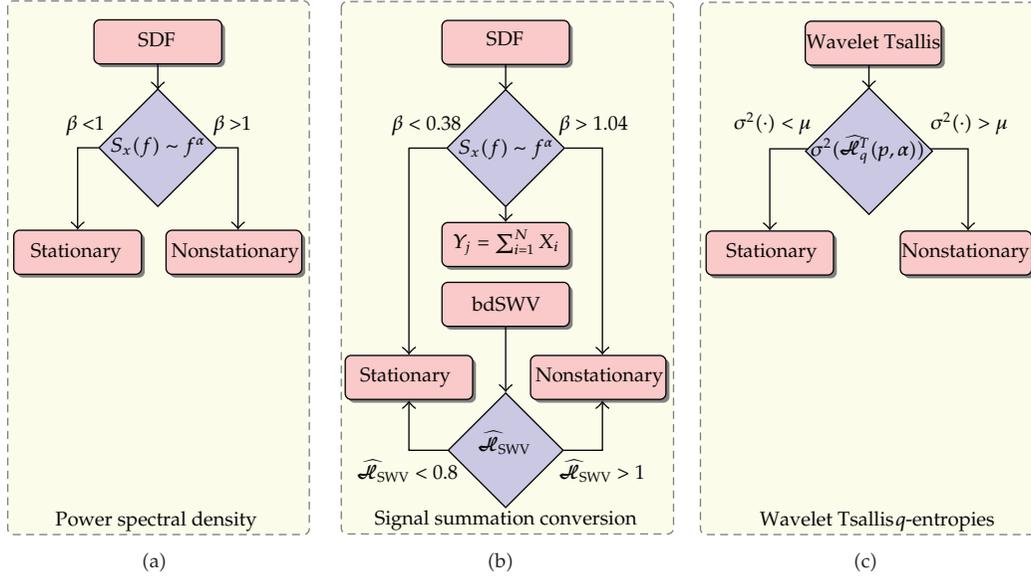


Figure 5: Algorithms for classifying scaling signals as stationary or nonstationary. Leftmost diagram displays the steps required in PSD, middle diagram the ones for SSC, and rightmost plot displays the steps of the proposed technique based on wavelet Tsallis q -entropies.

signals based on the fact that the observed PSD of $1/f$ processes follows a power-law dependence ($S_X(f) \sim f^{-\alpha}$). When the estimated parameter of power-law dependence ($\hat{\alpha}$) is less than 1 ($\hat{\alpha} < 1$), the process is stationary; on the other hand if $\hat{\alpha} > 1$, the process is non-stationary. Signal classification in the SDF framework is therefore accomplished by first estimating the SDF of the scaling process under study using some standard methodology (e.g., Periodogram), plot $\log(S_x(f))$ versus f , fit a line, compute the slope which corresponds to the estimated $\hat{\alpha}$, and finally based on the observed slope determine the nature of the process. Authors in [5] reported on the classification properties of the SDF method using synthesized signals of the fGn/fBm type. Eke and coauthors [5, 41] claimed that PSD performs satisfactorily when the process' scaling parameter lies in the intervals $-1 < \alpha < 0.38$ and $1.04 < \alpha < 3$ but misclassifies signals in the range $0.38 < \alpha < 1.04$. Because of this, they proposed a methodology specially designed to enhance the classification of signals in the interval $\alpha \in (-1, 3)$. Figure 5 displays the algorithm based on PSD to classify scaling signals.

4.2. Signal Summation Conversion

As stated in the previous section, PSD offers limited classification when the scaling process studied has a scaling index lying in the interval $\alpha \in (0.38, 1.04)$. The work of Eke et al. [5] not only identified this limitation but also proposed a solution based on the cumulative sum operation. The technique, called signal summation conversion, is only necessary whenever the estimated scaling index of the process lies in $\alpha \in (0.38, 1.04)$. The solution posed by Eke was to classify the process in the non-stationarity domain by the use of some standard non-stationarity technique. The use of the cumulative sum technique allowed the conversion of a stationary process into a non-stationary one and also maintaining the non-stationarity condition in a non-stationary process. Once the process to be classified is transformed to

exhibit non-stationarity features, the following step is to estimate the Hurst index of this process by using some standard technique, for example, bridge-detrended scaled-windowed variance (bdSWV). Depending upon the estimated Hurst index obtained by bdSWV (\widehat{H}_{SWV}), a process is classified as stationary whenever $\widehat{H}_{\text{SWV}} < 0.8$ and non-stationary when $\widehat{H}_{\text{SWV}} > 1$. If the estimated \widehat{H}_{SWV} lies outside this interval, the scaling process is regarded as unclassifiable. Eke and coworkers showed that the SSC enhances the classification observed in PSD at the expense of higher computational time. Even though SSC enhanced the classification of processes, many disadvantages can be identified in this technique. First, extended fGn cannot be classified as stationary within this framework as its cumulative sum is still stationary; secondly, SSC is based on PSD, a technique which has traditionally been attached to stationary signals. In addition, SSC has not been tested on signals displaying more complex behaviour such as PPLs, and the signals used to perform the classification are long. Figure 5 displays the algorithms for performing scaling signal classification in the PSD and SSC framework.

4.3. Wavelet Tsallis q -Entropies

Section 3 demonstrated that wavelet Tsallis q -entropies can be modelled by sum-cosh apodizing functions which among other properties display constant regions of entropies and regions of decreasing entropies (i.e., quasirectangular behaviour). The length of the constant region, which usually lies in a symmetric range of the scaling index α , can be controlled by the non-extensivity parameter q of Tsallis entropies. If the constant region of entropies lies in the interval $\alpha \in (-1, 1)$, then, every stationary scaling process will present maximum wavelet Tsallis entropy ($\mathcal{H} = 1$). On the other hand if the process has a scaling index α outside this range it will present fluctuations of entropy. The above suggest that wavelet Tsallis q -entropies can be used to differentiate scaling signals as stationary or non-stationary based on the observed entropies. If the estimated entropies are constant, then the scaling process is stationary, otherwise it is non-stationary.

Figure 6 captures the rationale behind the signal classification procedure based on wavelet Tsallis q -entropies. As long as $q \geq 5$, constant regions of entropies are observed for scaling signals with $\alpha < \alpha_{\text{coff}}(q)$ and varying for $\alpha > \alpha_{\text{coff}}(q)$. If $\alpha_{\text{coff}}(q) = 1$, then classification of fractional noises and motions can be accomplished, and when $\alpha_{\text{coff}}(q) = 3$, classification of fractional motions from extended fractional motions is accomplished. Therefore, wavelet Tsallis q -entropies not only allows distinguishing stationary from non-stationary but also non-stationary from non-stationary.

5. Methodology

In [5], a comparison of PSD and SSC was performed by using synthesized signals of length $N = 2^{17}$. SSC was reported to present better classifications of fBms (as true fBms) and fGns (as true fGns). The present paper extends the results reported in [5] to PPL signals, which are known to present more complex behaviour than fBms and fGns, and proposes a novel methodology for scaling signal classification based on wavelet Tsallis q -entropies. The paper uses PPL signals with length $N = 2^{14}$, which are more realistic in the sense that many studies, reported in the literature with measured data, claimed that the nature of the phenomena does not permit to obtain higher signal lengths [6]. Also, estimation techniques often increase their MSE for short-length signals. Therefore, the present study not only proposes a methodology for classifying scaling processes as stationary or non-stationary but also compares the

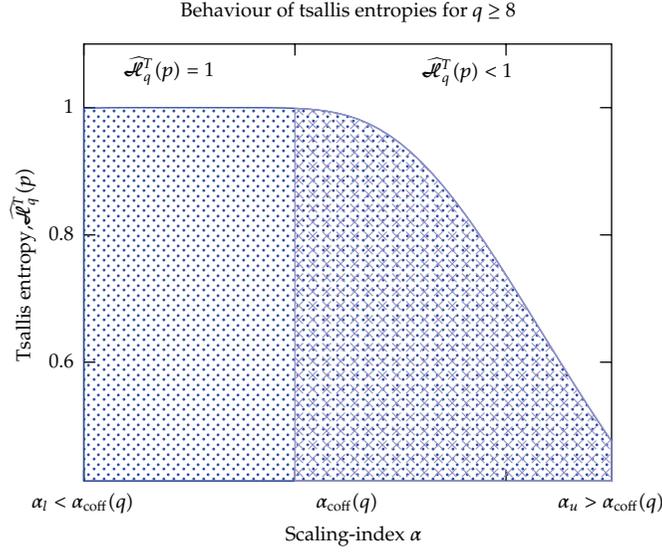


Figure 6: Dependence of the constant entropies on the nonextensivity parameter of Tsallis entropies.

techniques for signal classification in non-standard conditions (i.e., by using complex signals with short lengths). PPL signals were synthesized by using the *R* package `fractal`, which simulates signals using the circular embedding algorithm of Davies and Harte [43]. To test the performance of each technique, PPL signals were generated in the range $.01 < \alpha < 1.99$ in steps of $.01$. For each α (in the range $.01 < \alpha < 1.99$), 100 traces were simulated; therefore, a total of 19900 traces were studied. The selection of the range: $.01 < \alpha < 1.99$ is because of the fact that techniques of signal classification often fail in the limit $\alpha \rightarrow 1$ and perform better outside this range. SSC often considers a signal as unclassifiable; however, for the purposes of comparison, an unclassifiable signal is regarded as misclassified in this paper. SSC was implemented in *R* using the PSD algorithm of the `fractal` package. Wavelet entropy was implemented in *R* (and also in MATLAB), and the classification of signals was based on fluctuations of entropy (by computing wavelet entropy in sliding windows). To study the fluctuations, subsets of the original scaling signal, $X(t)$, were taken in sliding windows of the form:

$$X(m; \omega, \Delta) = X(t_k) \Pi\left(\frac{t - m\Delta}{\omega} - \frac{1}{2}\right), \quad (5.1)$$

where $m = 0, 1, 2, \dots, m_{\max}$, Δ is the sliding factor, and $\Pi(\cdot)$ is the standard rectangular function. Once the signals were classified, their results were summarized by plotting

$$\{\mathcal{N}(m; j), j\}_{j=.01}^{1.99}, \quad (5.2)$$

where $\mathcal{N}(m; j)$ stand for the number of signals classified correctly for the technique m for signals with scaling index j .

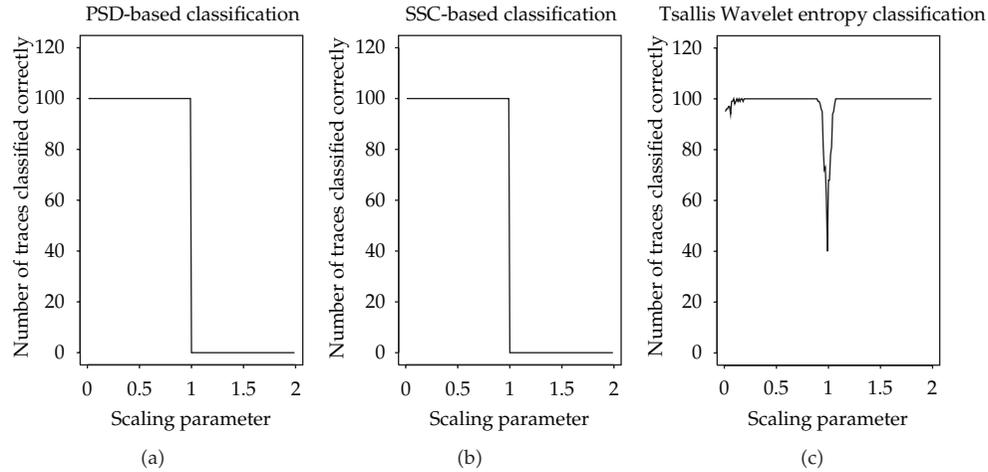


Figure 7: Classification of correct *power-law* signals. Recall that $\alpha < 1$ indicates the presence of a fractal noise while $\alpha > 1$ designates a non-stationary fractal motion. Left plot shows classification for PSD method, middle plot for the so-called SSC (signal summation conversion), and right plot to the novel wavelet Tsallis q -entropies-based method with $q = 20$.

6. Experimental Results

Figure 7 displays the results of the experimental study detailed (methodology) detailed in previous section. Note that for PSD, stationary PPL signals are classified correctly (i.e., classified as stationary). This was expected, since as previously stated, PSD was primarily designed to work for time-invariant (stationary) random signals. For non-stationary signals, PSD classifies non-stationary signals as stationary, misclassifying every non-stationary PPL process. PSD, therefore, do not provide reliable classifications of PPL signals, and it is not recommended for use in a signal classification scheme. In [5], SSC was shown to enhance the classification of scaling signals for the range $\hat{\alpha} \in (0.38, 1.04)$. Note, however, that SSC enhances the classification mostly for stationary signals. Moreover, SSC is only applicable as the estimated scaling index lies in $\hat{\alpha} \in (0.38, 1.04)$, otherwise only the PSD is applied. Based on this, it is expected to have a similar behaviour of classifications as the PSD for the SSC technique. Middle plot of Figure 7 displays the classifications of the SSC technique for PPL signals of length $N = 2^{14}$. As expected, SSC presents identical behaviour as that of PSD and, as the case of PSD, is not recommended as a signal classification tool for signals with PPL behaviour. The results of PSD and those of SSC differ from those presented in [5]. Note, however, that the signals studied in this paper are of different nature than those studied in [5]. First, the signals studied in the work of Eke et al. [5] are fBms and fGns, which in some sense are more easily classifiable since their smoothness properties are visually different. Finally, the length of the signals studied in the work of Eke are longer than the length of the signals considered in this paper. It is well known that as the length of the studied signals increases, the MSE of the estimated $\hat{\alpha}$ decreases. Thus, the synthesized signals studied in this paper possess not only higher complexities but also shorter length. Rightmost plot of Figure 7 presents the classifications of PPL signals using the methodology proposed in this paper based on wavelet Tsallis q -entropies. Note that the classifications of the proposed technique are better than those observed in PSD and SSC. The proposed technique based on wavelet Tsallis q -entropies classifies correctly stationary as well as non-stationary PPL signals and that this classification

is somewhat unacceptable in the limit of $\alpha \rightarrow 1$. The technique based on wavelet Tsallis q -entropies is fast enough and can also classify extended fGns from fGns and fBms from extended fBms. The classifications of signals with these characteristics are not supported by either the PSD and SSC techniques. In performing the classifications with the technique based on wavelet Tsallis entropies, the entropies were computed in sliding windows, and the boundary of fluctuations was taken as $\mu = 3e - 09$. The nonextensivity parameter q was set to $q = 10$ but similar results are observed for $q \geq 10$.

7. Conclusions

This paper presented a novel methodology for classifying scaling signals as stationary or non-stationary based on wavelet Tsallis q -entropies. It was shown that the sum-cosh window behaviour of wavelet Tsallis q -entropies allocated constant entropies to a set of scaling signals and varying to the rest and that the length of the constant region is controlled by q , the non-extensivity parameter of Tsallis entropies. It was also shown that by setting the constant regions to the range of stationary scaling signals, the problem of signal classification can be reduced to the observation of constant/nonconstant entropies. The classification properties of the PSD and SSC were extended to signals with pure power-law behaviour with length $N = 2^{14}$, and a comparison procedure was performed among PSD, SSC, and the technique based on wavelet Tsallis q -entropy. The results not only confirm that the technique based on wavelet Tsallis q -entropies provides meaningful classification but also outperforms PSD and SSC techniques. The results presented in this paper are meaningful in many areas of scaling signal analysis since many estimation/analysis results presented in the literature have been performed without a phase of signal classification.

Acknowledgments

The present paper was jointly funded by the National Council of Science and Technology (CONACYT) under Grant 47609, FOMIX-COQCYT Grant no. 126031, and University of Caribe internal funds.

References

- [1] J. Beran, "Statistical methods for data with long-range dependence (with discussion)," *Statistical Science*, vol. 7, pp. 404–416, 1992.
- [2] J. Beran, *Statistics for Long-Memory Processes*, vol. 61 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, New York, NY, USA, 1994.
- [3] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*, Stochastic Modeling, Chapman & Hall, New York, NY, USA, 1994.
- [4] M. J. Cannon, D. B. Percival, D. C. Caccia, G. M. Raymond, and J. B. Bassingthwaighte, "Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series," *Physica A*, vol. 241, no. 3-4, pp. 606–626, 1997.
- [5] A. Eke, P. Hermán, J. B. Bassingthwaighte et al., "Physiological time series: distinguishing fractal noises from motions," *Pflügers Archiv European Journal of Physiology*, vol. 439, no. 4, pp. 403–415, 2000.
- [6] D. Delignieres, S. Ramdani, L. Lemoine, K. Torre, M. Fortes, and G. Ninot, "Fractal analyses for "short" time series: a re-assessment of classical methods," *Journal of Mathematical Psychology*, vol. 50, no. 6, pp. 525–544, 2006.
- [7] 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.
- [8] I. W. C. Lee and A. O. Fapojuwo, "Stochastic processes for computer network traffic modeling," *Computer Communications*, vol. 29, no. 1, pp. 1–23, 2005.

- [9] D. B. Percival, "Stochastic models and statistical analysis for clock noise," *Metrologia*, vol. 40, no. 3, pp. S289–S304, 2003.
- [10] S. B. Lowen and M. C. Teich, "Estimation and simulation of fractal stochastic point processes," *Fractals*, vol. 3, no. 1, pp. 183–210, 1995.
- [11] S. Thurner, S. B. Lowen, M. C. Feurstein, C. Heneghan, H. G. Feichtinger, and M. C. Teich, "Analysis, synthesis, and estimation of fractal-rate stochastic point processes," *Fractals*, vol. 5, no. 4, pp. 565–595, 1997.
- [12] D. C. Caccia, D. Percival, M. J. Cannon, G. Raymond, and J. B. Bassingthwaite, "Analyzing exact fractal time series: evaluating dispersional analysis and rescaled range methods," *Physica A*, vol. 246, no. 3-4, pp. 609–632, 1997.
- [13] F. Serinaldi, "Use and misuse of some Hurst parameter estimators applied to stationary and non-stationary financial time series," *Physica A*, vol. 389, no. 14, pp. 2770–2781, 2010.
- [14] B. D. Malamud and D. L. Turcotte, "Self-affine time series: measures of weak and strong persistence," *Journal of Statistical Planning and Inference*, vol. 80, no. 1-2, pp. 173–196, 1999.
- [15] J. C. Gallant, I. D. Moore, M. F. Hutchinson, and P. Gessler, "Estimating fractal dimension of profiles: a comparison of methods," *Mathematical Geology*, vol. 26, no. 4, pp. 455–481, 1994.
- [16] W. Rea, M. Reale, J. Brown, and L. Oxley, "Long memory or shifting means in geophysical time series?" *Mathematics and Computers in Simulation*, vol. 81, no. 7, pp. 1441–1453, 2011.
- [17] C. Cappelli, R. N. Penny, W. S. Rea, and M. Reale, "Detecting multiple mean breaks at unknown points in official time series," *Mathematics and Computers in Simulation*, vol. 78, no. 2-3, pp. 351–356, 2008.
- [18] B. B. Mandelbrot and J. W. van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, pp. 422–437, 1968.
- [19] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, part 2, pp. 910–917, 1992.
- [20] S. Stoev, M. S. Taqqu, C. Park, and J. S. Marron, "On the wavelet spectrum diagnostic for Hurst parameter estimation in the analysis of Internet traffic," *Computer Networks*, vol. 48, no. 3, pp. 423–445, 2005.
- [21] L. Hudgins, C. A. Friehe, and M. E. Mayer, "Wavelet transforms and atmospheric turbulence," *Physical Review Letters*, vol. 71, no. 20, pp. 3279–3282, 1993.
- [22] A. Cohen and A. J. Kovačević, "Wavelets: the mathematical background," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 514–522, 1996.
- [23] 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.
- [24] H. Shen, Z. Zhu, and T. C. M. Lee, "Robust estimation of the self-similarity parameter in network traffic using wavelet transform," *Signal Processing*, vol. 87, no. 9, pp. 2111–2124, 2007.
- [25] J.-M. Bardet, "Statistical study of the wavelet analysis of fractional Brownian motion," *IEEE Transactions on Information Theory*, vol. 48, no. 4, pp. 991–999, 2002.
- [26] J. A. Bonachela, H. Hinrichsen, and M. A. Muñoz, "Entropy estimates of small data sets," *Journal of Physics A*, vol. 41, no. 20, Article ID 202001, 2008.
- [27] U. Kumar, V. Kumar, and J. N. Kapur, "Normalized measures of entropy," *International Journal of General Systems*, vol. 12, no. 1, pp. 55–69, 1986.
- [28] M. T. Martin, A. R. Plastino, and A. Plastino, "Tsallis-like information measures and the analysis of complex signals," *Physica A*, vol. 275, no. 1, pp. 262–271, 2000.
- [29] R. G. Baraniuk, P. Flandrin, A. J. E. M. Janssen, and O. J. J. Michel, "Measuring time-frequency information content using the Rényi entropies," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1391–1409, 2001.
- [30] D. G. Pérez, L. Zunino, M. Garavaglia, and O. A. Rosso, "Wavelet entropy and fractional Brownian motion time series," *Physica A*, vol. 365, no. 2, pp. 282–288, 2006.
- [31] M. M. Añino, M. E. Torres, and G. Schlotthauer, "Slight parameter changes detection in biological models: a multiresolution approach," *Physica A*, vol. 324, no. 3-4, pp. 645–664, 2003.
- [32] L. Zunino, D. G. Pérez, M. Garavaglia, and O. A. Rosso, "Wavelet entropy of stochastic processes," *Physica A*, vol. 379, no. 2, pp. 503–512, 2007.
- [33] R. Quian Quiroga, O. A. Rosso, E. Başar, and M. Schürmann, "Wavelet entropy in event-related potentials: a new method shows ordering of EEG oscillations," *Biological Cybernetics*, vol. 84, no. 4, pp. 291–299, 2001.
- [34] W. X. Ren and Z. S. Sun, "Structural damage identification by using wavelet entropy," *Engineering Structures*, vol. 30, no. 10, pp. 2840–2849, 2008.

- [35] H. A. Al-Nashash, J. S. Paul, W. C. Ziai, D. F. Hanley, and N. V. Thakor, "Wavelet entropy for subband segmentation of EEG during injury and recovery," *Annals of Biomedical Engineering*, vol. 31, no. 6, pp. 653–658, 2003.
- [36] O. A. Rosso, L. Zunino, D. G. Pérez et al., "Extracting features of Gaussian self-similar stochastic processes via the Bandt-Pompe approach," *Physical Review E*, vol. 76, no. 6, Article ID 061114, 6 pages, 2007.
- [37] L. Zunino, D. G. Pérez, M. T. Martín, A. Plastino, M. Garavaglia, and O. A. Rosso, "Characterization of Gaussian self-similar stochastic processes using wavelet-based informational tools," *Physical Review E*, vol. 75, no. 2, Article ID 021115, 10 pages, 2007.
- [38] D. G. Pérez, L. Zunino, M. T. Martín, M. Garavaglia, A. Plastino, and O. A. Rosso, "Model-free stochastic processes studied with q-wavelet-based informational tools," *Physics Letters, Section A*, vol. 364, no. 3-4, pp. 259–266, 2007.
- [39] K. Avci and A. Nacaroglu, "Cosh window family and its application to FIR filter design," *AEU*, vol. 63, no. 11, pp. 907–916, 2009.
- [40] P. Castiglioni, G. Parati, A. Civijian, L. Quintin, and M. D. Rienzo, "Local scale exponents of blood pressure and heart rate variability by detrended fluctuation analysis: effects of posture, exercise, and aging," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, Article ID 4633671, pp. 675–684, 2009.
- [41] A. Eke, P. Hermán, L. Kocsis, and L. R. Kozak, "Fractal characterization of complexity in temporal physiological signals," *Physiological Measurement*, vol. 23, no. 1, pp. R1–R38, 2002.
- [42] G. L. Gebber, H. S. Orer, and S. M. Barman, "Fractal noises and motions in time series of presympathetic and sympathetic neural activities," *Journal of Neurophysiology*, vol. 95, no. 2, pp. 1176–1184, 2006.
- [43] R. B. Davies and D. S. Harte, "Tests for hurst effect," *Biometrika*, vol. 74, no. 1, pp. 95–101, 1987.

Research Article

Grouping Optimization Based on Social Relationships

Rong-Chang Chen

*Department of Distribution Management, National Taichung University of Science and Technology,
Taichung 404, Taiwan*

Correspondence should be addressed to Rong-Chang Chen, rchens@nutc.edu.tw

Received 1 July 2011; Revised 23 September 2011; Accepted 27 September 2011

Academic Editor: Sheng-yong Chen

Copyright © 2012 Rong-Chang Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grouping based on social relationships is a complex problem since the social relationships within a group usually form a complicated network. To solve the problem, a novel approach which uses a combined sociometry and genetic algorithm (CSGA) is presented. A new nonlinear relation model derived from the sociometry is established to measure the social relationships, which are then used as the basis in genetic algorithm (GA) program to optimize the grouping. To evaluate the effectiveness of the proposed approach, three real datasets collected from a famous college in Taiwan were utilized. Experimental results show that CSGA optimizes the grouping effectively and efficiently and students are very satisfied with the grouping results, feel the proposed approach interesting, and show a high repeat intention of using it. In addition, a paired sample *t*-test shows that the overall satisfaction on the proposed CSGA approach is significantly higher than the random method.

1. Introduction

Grouping optimization based on social relationships attracts more and more attention as social networks [1–11] have been growing rapidly in recent years. A social network is a special structure made of individuals (or organizations). It includes the ways in which individuals are connected through various social familiarities [1]. Through the analysis of relational network and the measurement of social relationships, grouping optimization is obtainable and can achieve a group objective.

In dealing with the grouping optimization problem concerning social relationships, some important issues should be addressed. Firstly, there is a deep discrepancy between the official and the secret behaviour of members [12]. A good approach to the grouping optimization should be able to find out the hidden information behind a group. Relations can be conceptualized at three levels of social complexity: individual, dyad, and group [13]. The approach should be able to disclose the hidden information in these levels. In addition,

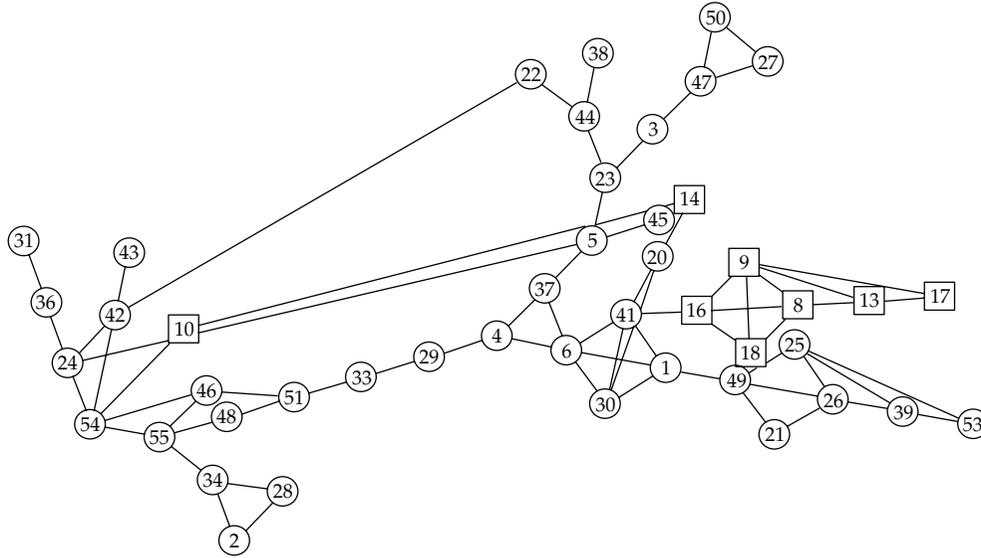


Figure 1: A social network structure.

the approach should be cost-effective. Consequently, a simple method which can explore indeed thinking of individuals is needed. Another issue is that most social networks such as Facebook employ a linear model to measure the links between group members. They use the number of connections that a node has to evaluate degree, betweenness, closeness, and network centralization [6]. However, a nonlinear relation model should be used since the real relationship tends to be nonlinear. For example, suppose that two nodes have the same number of connections. The connections for the first node are all its first choices, while for the second one the connections are its, for example, second and third choices, respectively. The disappointment in being grouped with a second choice partner over a first is unlikely to be linearly related to that of being grouped with a fourth choice over a third. A nonlinear model should be developed to suitably describe the relative importance of relationships. Thirdly, the approach should be able to help organizations improve their performances. Finally, but not the last, many grouping optimization problems subject to some constraints are known as NP-hard [14–16]. In addition, the structure of social relations usually forms a complex network, causing the problem to be very difficult to deal with. Another cause to make the grouping optimization more complicated is the constraints that require each group to include some different attributes, such as ability, specialty, gender, and position.

To tackle the above-mentioned complex problem, a new approach based on a combined sociometry [17–21] and genetic algorithm (GA) [22–28] is employed. For convenience, we call this novel approach CSGA in short. To measure the social relationships between members and to explore the hidden information behind a group, a choice-making method [29] is employed. Group members are first asked to indicate their choices (or preferences) to other members in an index system that a one stands for the first choice, two the second choice, and the like until an allowed maximum number of choices is reached. Based on the choices, a sociogram can be drawn by using a sociometry tool [17–21], as illustrated in Figure 1. Previous studies showed that sociometric choices do tend to predict such performance criteria as productivity, combat effectiveness, training ability, and leadership [30]. Moreover, some indices of social status can be calculated. In this paper, a new nonlinear model for measuring social status is presented. From the sociogram and social

status indices, the socially vulnerable or the potential bullying members can be found. Then the grouping optimization can be done according to the social status of individuals and the objective of the organization. Results from this study show that the proposed approach is effective in dealing with the grouping problem mentioned above. Furthermore, experiments founded on real data also show high satisfaction and repeat intention with this method.

The remainder of this paper is structured as follows. The grouping problem is briefly introduced in Section 2. In the subsequent section, the sociometry is briefly introduced and a new modified model for measuring relations between members is presented. Then proposed approach is presented in Section 4. Subsequently, results and discussion are presented. Finally, concluding remarks are drawn in Section 6.

2. The Grouping Problem

The grouping problem belongs to a large family of problems which partitions a set of items U into a collection of mutually disjoint subsets U_i of U , such that $\cup U_i = U$ and $U_i \cap U_j = \emptyset$, $i \neq j$. In this study, the grouping is to optimally assign m members (individuals) to g subgroups with n members in each subgroup. Let $M = \{1, 2, \dots, m\}$ be a set of members, $G = \{1, 2, \dots, g\}$ a set of subgroups, $P = \{1, 2, \dots, n-1\}$ a set of partners of a member, $A = \{1, 2, \dots, a_l\}$ a set of attributes (or position), and $C = \{1, 2, \dots, N_p\}$ a set of choices, where m is the number of members in a group, g is the total number of subgroups, n is the number of members in each subgroup, and a_l is the number of attribute l . The members can list a number of preferred members with whom they would like to work with or sit beside by an index system that a "1" stands for the first choice, "2" for the second choice, and the like, up to a preassigned maximum integer N_p . Note that a member cannot select himself or herself as his or her own partner. Slavin suggested that the best group size is from two to six [31]. For $i \in M$, $k \in P$, define c_{ik} as the preference given by member i to being grouped with his/her partner k and c_{ki} as the preference given by partner k to being grouped with member i . If partner k is the first choice of member i , c_{ik} is equal to one, the second choice, c_{ik} is equal to two, and the like. A lower coefficient of preference means that a member has more preference with his/her partner. If member i does not include his/her partner k in their list of preferences, then c_{ik} is assigned a relatively large penalty value b . The mathematical formulation, hence, can be expressed as follows:

$$\min F = \sum_{i \in M} \sum_{j \in G} w_i R_i x_{ij}, \quad (2.1)$$

$$\text{subject to } \sum_{j \in G} x_{ij} = 1 \quad \forall i \in M, \quad (2.2)$$

$$\sum_{i \in M} x_{ij} = n \quad \forall j \in G, \quad (2.3)$$

$$\sum_{i \in M} x_{ij} z_{il} = a_l \quad \forall j \in G, \quad \forall l \in A, \quad (2.4)$$

$$z_{il} = \begin{cases} 1 & \text{if member } i \text{ has an attribute } l, \\ 0, & \text{otherwise,} \end{cases} \quad (2.5)$$

$$x_{ij} = \begin{cases} 1 & \text{if member } i \text{ is assigned to group } j, \\ 0, & \text{otherwise,} \end{cases} \quad (2.6)$$

$$\forall i \in M, k \in P \quad c_{ik} \text{ or } c_{ki} = b \quad \text{if } c_{ik} \text{ or } c_{ki} \notin C, \quad (2.7)$$

where $R_i = \sum_{k \in P} \{f(c_{ik}) + f(c_{ki})\}$, for all $i \in M$. The objective is to minimize the total scoring value, as shown in (2.1). Note that the scoring value is composed of two main parts: the priority weight w_i of member i and scoring function R_i . The priority weight can be assigned by two ways: directly assigned by the manager of the organization or calculated based on the social relationships. As for the scoring function $f(\cdot)$, a popular function used is the squared function [29]. Equation (2.2) requires that each member is assigned exactly one subgroup. Equation (2.3) ensures that one subgroup is composed of n members. Equation (2.4) requires that there are exact a_l members with attribute l in a subgroup. Note that a_l depends on attribute l . For example, if a basketball team is composed of a center (attribute 1), two forwards (attribute 2), and two guards (attribute 3), then $a_1 = 1$, $a_2 = 2$, and $a_3 = 2$, respectively.

As the number of members in a group increases, the number of possible grouping outcomes, u , grows at an unacceptably rapid rate; it grows exponentially. For example, for $n = 2$, $u = 945$ when $m = 10$, $u = 6.55E + 8$ when $m = 20$, and $u = 6.19E + 15$ when $m = 30$. The optimal solution cannot be found in polynomial time. In addition, the relationship structure becomes a complex network as m increases. To deal with the optimization problems concerning a complex network, it is impractical to use an approach like the exhaustive method since it takes a huge amount of computation time. Instead, some useful algorithms such as a genetic algorithm [22–28] for finding approximate or near-optimal solutions or strategies for improving efficiency, solution quality, exactitude, and more have been presented [32–35]. In this paper, a genetic algorithm [22–28], which is proven to be very effective in dealing with complex optimization problems, is employed to find feasible solutions.

3. The Sociometric Analysis

Sociometry was developed by Moreno in 1934. Moreno defined sociometry as “the inquiry into the evolution and organization of groups and the position of individuals within them” [17]. It is a quantified method aiming at measuring and determining social relationships in groups [18]. Sociometric explorations disclose the hidden structures that give a group its form: the alliances, the subgroups, the hidden beliefs, the ideological agreements, the stars of the show, and so on. One of Moreno’s innovations in sociometry was the development of the sociogram, a systematic method for graphically representing individuals as points/nodes and the relationships between them as lines/arcs [19]. Given choices or preferences within a group, sociograms were developed and the structure and patterns of group interactions can be drawn on the basis of many different criteria [19]: social relations, channels of influence, lines of communication, and so on. Some terms in sociometry are the following:

- (1) Stars: those who have many choices;
- (2) Isolates: those with few or no choices;
- (3) Mutual Choice (MC): individuals who choose each other;
- (4) One-Way Choice: referring to individuals who choose someone but the choice is not reciprocated;
- (5) Cliques: groups of three or more people within a larger group who all choose each other.

There are many indices using values instead of categories to indicate the social status of individuals. Among the most popular indices, three indices are introduced in the following paragraphs.

3.1. Status Score Index (SSI)

One of the most critical events in the history of sociometric methods is the use of both positive and negative nominations [13]. The difference between the positive and negative nominations is called status score [20]. To easily figure out the social status of a member, a relative or a normalized score called SSI was developed and defined as

$$SSI = \frac{N^{TC} - N^{TR}}{m - 1}, \quad (3.1)$$

where N^{TC} denotes the total choice number by other members, N^{TR} denotes the total rejection number by other members, and m is the total number of members in a group. Note that a member cannot choose himself or herself. Thus, $m - 1$ rather than m is used in (3.1). The value of SSI is in the range $[-1, 1]$. A higher value of SSI means that a member is more popular within a group. However, this index considers only one-way choice, and thus one cannot understand the mutual interaction between members from this index.

3.2. Index of Sociometric Status Score (ISSS)

To consider the mutual choices between members, an index called ISSS [21] is defined as

$$ISSS = \frac{1}{2} \left(\frac{N^{TC} - N^{TR}}{m - 1} + \frac{N^{MC} - N^{MR}}{N_p} \right). \quad (3.2)$$

On the right part of the equation, N^{MC} and N^{MR} represent the number of mutual choice and the number of mutual rejection, respectively. N_p is the allowed maximum number of choices that a member can make. The value of ISSS is also in the range $[-1, 1]$. A higher value of ISSS indicates that a member is more popular within a group.

3.3. Modified Index of Sociometric Status Score (MISSS)

ISSS uses a linear function to express relationship. However, it fails to consider the relative importance of different choices. For instance, satisfaction about being grouped with a first-, a second-, or a third-choice partner should be different, and the satisfaction difference between the first to the second and the second to the third is generally quite unlike. Consequently, a modified model for measuring the status score should be presented. In consideration with

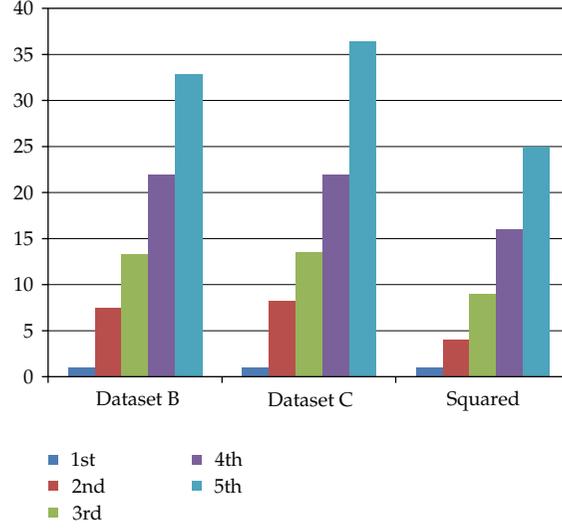


Figure 2: Characteristics of preferences among surveyed students.

this relative importance of different choices, a new index called MISSS is developed and defined as

$$\text{MISSS}_i = \frac{1}{2} \left(\frac{\sum_{p=1}^{m-1} z_{pi} w_p f(c_{pi}^{\text{TC}}) - \sum_{q=1}^{m-1} z_{qi} w_q f(c_{qi}^{\text{TR}})}{m-1} + \frac{\sum_{r=1}^{N_p} z_{ri} w_r f(c_{ri}^{\text{MC}}) - \sum_{s=1}^{N_p} z_{si} w_s f(c_{si}^{\text{MR}})}{\sum_{t=1}^{N_p} w_t f(c_{ti})} \right). \quad (3.3)$$

MISSS is composed of four parts: the contributions from the total choices, the total rejections, the mutual choices, and the mutual rejections, respectively. $z \in \{0, 1\}$ and z_{pi} , z_{qi} , z_{ri} , and $z_{si} = 1$ if a member i is chosen by other members, rejected by other members, choose mutually with other members, and reject mutually with other members, respectively. w_p , w_q , w_r , and w_s are their priority weights, respectively. The value of MISSS is in the range $[-1, 1]$. A higher value of MISSS means that a member is more popular within a group. MISSS gives a more real description of relative importance of relationships. An important issue about MISSS is how to express the relations in the equation. A scoring function $f(\cdot)$, which is a function of choice (preference), is used to measure the relations. A simple method to find out the function is the use of a questionnaire asking members to provide disappointment scores of being grouped with a second-, third-, fourth-, or fifth-choice partner relative to receiving their first-choice one. To illustrate the nature of the scoring function, choices were collected from two classrooms in a college and the values were calculated. The results are shown in Figure 2. The scoring functions appear nonlinear rather than linear. As mentioned in Section 1, a linear scoring function is unreasonable. To measure the social relationships, a nonlinear model should be used.

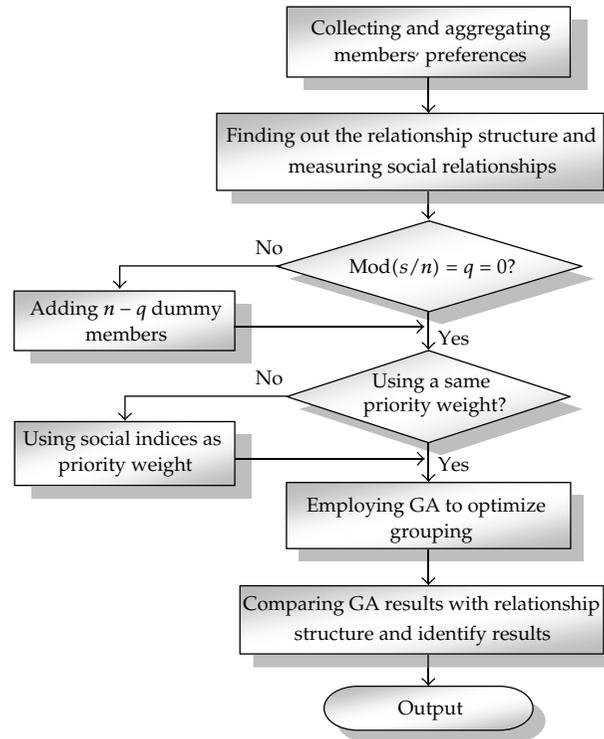


Figure 3: The proposed CSGA approach.

4. The CSGA Approach

4.1. The Procedure

The proposed CSGA approach is illustrated in Figure 3.

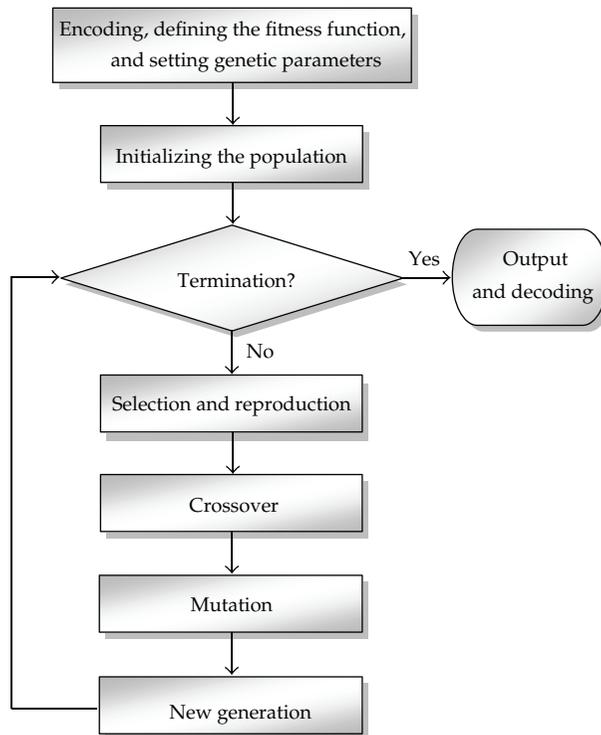
To begin with, the choices of members are collected and aggregated. A questionnaire is developed to collect the choices of the members and the reasons for choosing their partners. The questions are “whom in the group do you want to have a lunch or dinner with?” as a selection of lunch/dinner partner and “whom in the group do you want to sit beside and discuss with?” as a selection of learning partner. Table 1 illustrates a part of the questionnaire.

To be fair with all members, each member should fill in the same number of choices. Chen et al. [29] showed that a fewer number of choices filled in will have more chances to be assigned to their top choices. After aggregating the choices, a sociometric tool is employed to draw the sociogram and some social indices are measured.

If the remainder of $m/n = q \neq 0$, $n - q$ dummy members will be added and then grouping is performed using GA. Otherwise, GA is directly employed to group members. Adding dummy members does not influence the optimization since doing this only adds a constant value in the objective function. Note that the optimization can be done by different priority weights or the same priority weights. A higher value of the priority weight means that a member has a higher priority to be assigned his/her first choice. In the case of the same priority weight, all members have the same chance to be assigned to their top choices.

Table 1: An illustration of the questionnaire for collecting members' choices.

Priority	Index number of the member	Name	Index number of the main reason	Other reason (if available)
1	13	Don	8	
2	7	Mary	2	
3	2	Tom	17	good-looking
...	
N_p	1	Joe	12	

**Figure 4:** The flowchart of the GA program.

4.2. The GA Structure

The GA flowchart is illustrated in Figure 4. The details are depicted in the following paragraphs.

4.2.1. Encoding

The encoding of a chromosome is illustrated in Figure 5. Since there are m members, the number of genes is thus equal to m . Each gene is assigned a number which stands for a member. As illustrated in Figure 5, the values of the genes are in the order of 5, 2, 6, 10, ..., and 4. If $n = 2$, that is, a subgroup has two members, then member 5 and member 2 are at the same subgroup, member 6 and member 10 are at the same subgroup, and the like.

5	2	6	10	3	7	11	23	...	4
---	---	---	----	---	---	----	----	-----	---

Figure 5: Representation of a chromosome.

Table 2: A mutual choice table.

Member	Preference list	Mutual choice list
1	2, 7, 5, 4, 6	2, 6, 7
2	8, 3, 1, 7, 10	1, 7, 8, 10
3	2, 1, 9, m , 5	2, m
...
m	3, 6, 7, 8, 9	3, 6

4.2.2. Initialization of Population

The random method was employed to generate the initial solutions. Before producing the initial solutions, a mutual choice table was established. Table 2 illustrates the mutual choice. A member (say the 1st member) is first randomly selected, and the program checks with whom they choose mutually. Then from the mutual choice table, possible partners are 2, 6, and 7. Of these three possible partners, one is randomly selected to be a partner of the 1st member.

4.2.3. Evaluation of Fitness Function

To evaluate the chromosomes, the fitness value for each chromosome in the population was computed. The function used to measure the fitness is

$$\text{fitness} = \sum_{i \in S} \sum_{k \in P} w_i [f(c_{ik}) + f(c_{ki})]. \quad (4.1)$$

A lower fitness value represents a better grouping, and the optimal fitness value is kept. After producing new chromosomes, we can evaluate the chromosomes based on the members' preferences. The fitness function contains w_i , the scoring function $f(\cdot)$, c_{ik} and c_{ki} , where w_i represents the priority weight of student i , c_{ik} represents the preference indicated by member i to partner k , and the like. If a member is grouped with a partner which is not in his/her preference list, the GA program will give c_{ik} a sufficiently large value b . As for the scoring function $f(\cdot)$, a common used function is the squared function [29]. For example, if a squared scoring function is employed, $w_i = 1$, and a subgroup is composed of three members whose partners are their 2nd and 3rd, 1st, and 3rd, and 2nd and 2nd choices, respectively, then the total score is $2^2 + 3^2 + 1^2 + 3^2 + 2^2 + 2^2 = 4 + 9 + 1 + 9 + 4 + 4 = 31$. After computing the fitness of a subgroup j , the fitness of the chromosome is calculated as

$$F = \sum_{j \in G} f_j. \quad (4.2)$$

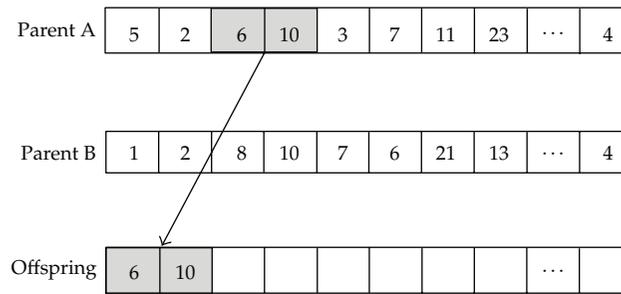


Figure 6: Crossover method.

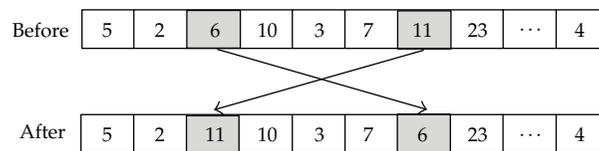


Figure 7: Mutation operation was performed by the swap method.

4.2.4. Selection

The Roulette Wheel Method [25] was employed to select fitter individuals. The fitness values of all the chromosomes in the population were first calculated and were then sorted. A fitter chromosome with a lower value of fitness has a higher probability to be selected. For example, if there are 50 chromosomes in the population. The chromosome with the lowest fitness value has a probability of $50 / (1 + 2 + 3 + \dots + 50)$ to be selected, the second lowest has a probability of $49 / (1 + 2 + 3 + \dots + 50)$ to be selected, and the like.

4.2.5. Crossover

Two chromosomes were randomly selected first. The fitness values of these two chromosomes were compared and from the better one some genes were randomly selected and placed at the beginning positions of the offspring chromosomes, as illustrated in Figure 6. The number of the selected genes is the same as the number of members per subgroup, n . The rest of the genes in the offspring chromosome are filled in the order of fitter genes by the greedy method.

4.2.6. Mutation

In this study, the swap method was employed to mutate. The method is illustrated in Figure 7. Two genes were randomly selected and then their values were interchanged. The swap method has the advantage of avoiding value duplication. Since a member is exactly assigned one subgroup, the values in the genes should be different.

Table 3: The datasets tested.

Dataset	Students	m	N_p	Grouping for	Grouping method	
					1st	2nd
A	Senior, in their 4th year	54	9	Dining partner	CSGA	CSGA
B	Junior, in their 3rd year	55	5	Learning partner	CSGA	CSGA
C	Freshman, in their 1st year	32	5	Learning partner	Random	CSGA

4.2.7. Elitism Strategy

In order to preserve the best chromosome in every generation, a simple elitism strategy [23] was employed. The best chromosome of each generation was duplicated to the next generation to ensure that it was preserved in the present population if it was the best when compared with other chromosomes of the population. This strategy assures that the best chromosome of each generation will be, even if not better, at least equal to the best chromosome of the previous generation. In addition, the elitism strategy did not lead the GA to converge prematurely since it was applied only to one chromosome of each generation.

4.2.8. Termination Condition

The termination condition we use in this study is the generation number defined by the user. The calculation will be repeated until the number of generation reaches the preassigned value. Once the termination criterion is satisfied, the solution is displayed with a chart. We can clearly know the result from the chart.

5. Results and Discussion

To numerically investigate the influences of grouping parameters such as the number of member per subgroup, the choice number, and the number of member in a group on the grouping results, a GA program was developed by Microsoft.NET 4.0. The program was run on an ASUS K52Jc Series notebook with Intel(R) Core (TM) i5 CPU M460 @2.53 GHz and 1.86 GB RAM. The operating system is Windows 7.

To evaluate the effectiveness of the proposed approach, three real datasets including three classes of undergraduate students were tested. The numbers of students are 54, 55, and 32, respectively, as shown in Table 3. For the dataset C, the method of grouping at the first time was by random and the second is by CSGA. A comparison can thus be made between these two grouping methods. For the dataset A, a maximum allowed number of choices, N_p , was set to be 9 so that the complexity of relationship network can be observed. Students were asked to show their preferences to the dining partners in dataset A and learning partners in datasets B and C.

For easy description, the following variables are defined. The population size is represented as N_{size} , the generation number as N_g , the crossover rate as R_c , and the mutation rate as R_m . In addition, we use F_{min} as the minimal fitness value in each generation, F_{best} as the best solution at each trial. In each case, the GA program performs 30 trials. We designate the best result of 30 trials as OPT, their average value as AVG, and the coefficient of variation as C_v .

Table 4: Comparison of results between Branch and Bound and GA.

m	n	Branch and Bound				GA	
		Fitness value	Runtime (min)	OPT	AVG	C_v	Average runtime, (min)
6	2	39	<0.01	39	39	0	1.29
8	2	67	0.09	67	67	0	1.31
10	2	144	8.25	144	144	0	1.52
12	2	176	>1099.71	176	176	0	1.64
14	2	NA	*	351	351	0	1.81
20	2	NA	*	323	358.73	0.15	2.43
40	2	NA	*	1360	1474.43	0.05	4.39
60	2	NA	*	2466	2718.1	0.04	6.29
6	3	85	<0.01	85	85	0	1.41
9	3	318	1.23	318	318	0	1.86
12	3	NA	*	664	664	0	2.23
15	3	NA	*	1145	1145	0	2.69
8	4	581	0.17	581	581	0	2.06
12	4	NA	*	1234	1234	0	2.69
15	5	NA	*	3016	3016	0	4.05

*a huge amount of computation time is required if compared with GA.

5.1. Validation of GA

To ensure the validation of the GA program, some tests were made. The results from GA were compared with those from the Branch and Bound (BB) method that can obtain optimal solutions. The results are shown in Table 4. When the number of members m is small, the GA program can obtain optimal solutions that are the same as those from BB, indicating that the GA program is valid. As m increases, the BB method requires a huge number of computation time. For example, for $m = 12$ and $n = 2$, the required time is about 1100 minutes or 18.3 hours. However, the average runtime for GA is about 1.64 minutes, much smaller than that for BB shown in Table 4.

5.2. The Influences of Genetic Parameters

To evaluate the influences of genetic parameters on the results, a two-stage approach was used. The generation number and population size were decided at the first stage, while at the second stage crossover rate and mutation rate were decided. The variation of minimal fitness value at each generation with different population sizes, N_{size} , is shown in Figure 8. As we can see from the figure $N_{\text{size}} = 20$, $N_g = 30000$ can obtain good results. At the second stage, the mutation rate as well as the crossover rate were changed and tested. Experimental results show that the best parameters are $R_c = 0.9$ and $R_m = 0.05$. In the following experiments, therefore, $R_c = 0.9$ and $R_m = 0.05$ were used.

5.3. The Influences of Choice Number

To easily observe the solutions, the sociograms (relationship structure) are drawn and shown in Figure 9. Note that only mutual-choice relationships are shown in the figure. A female

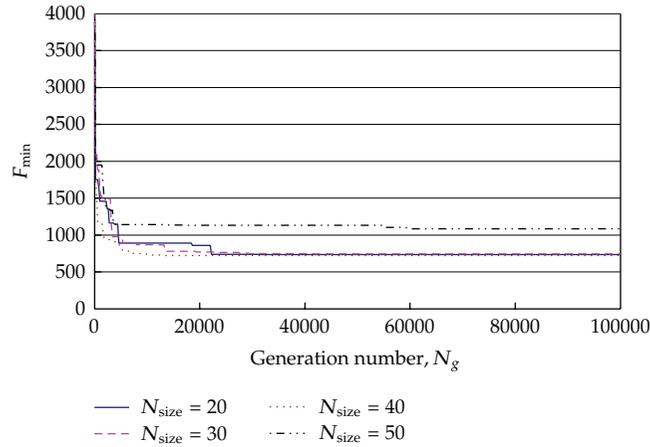


Figure 8: The variation of minimal fitness value in each generation with different population size, N_{size} .

Table 5: The influences of the number of members in a subgroup ($N_p = 5$, $N_g = 10,000$).

n	OPT	AVG	C_v	Average runtime (min)
2	801	952.57	0.10	2.09
3	2683	3170.63	0.07	2.98
4	5712	6131.40	0.04	3.97
5	9246	9548.60	0.03	5.05

member is represented with a circle, while a male member is represented with a square. From this figure we can see that the relationship network is quite complicated, especially for those with a large value of N_p . An optimal grouping is hard to be achieved by intuition or exact search methods.

5.4. The Influences of the Number of Members in a Subgroup

As the number of members in a subgroup n increases, the computation time will also increase. To observe the influences of n on the results, n was changed from two to five. Note that in Table 5 the number of choice N_p was fixed at 5, whereas in Table 6 the number of choice N_p was set to be equal to n .

5.5. The Influences of the Number of Members m in an Organization

As the number of members in a group m increases, the complexity of grouping also increases. An apparent influence is the increase of the computation time, as illustrated in Table 7. Note that the computation time increases linearly with m , showing the time efficiency of the GA program. Originally, the computation time grows exponentially. Consequently, GA is an efficient approach to solve grouping optimization problems considering social relationships.

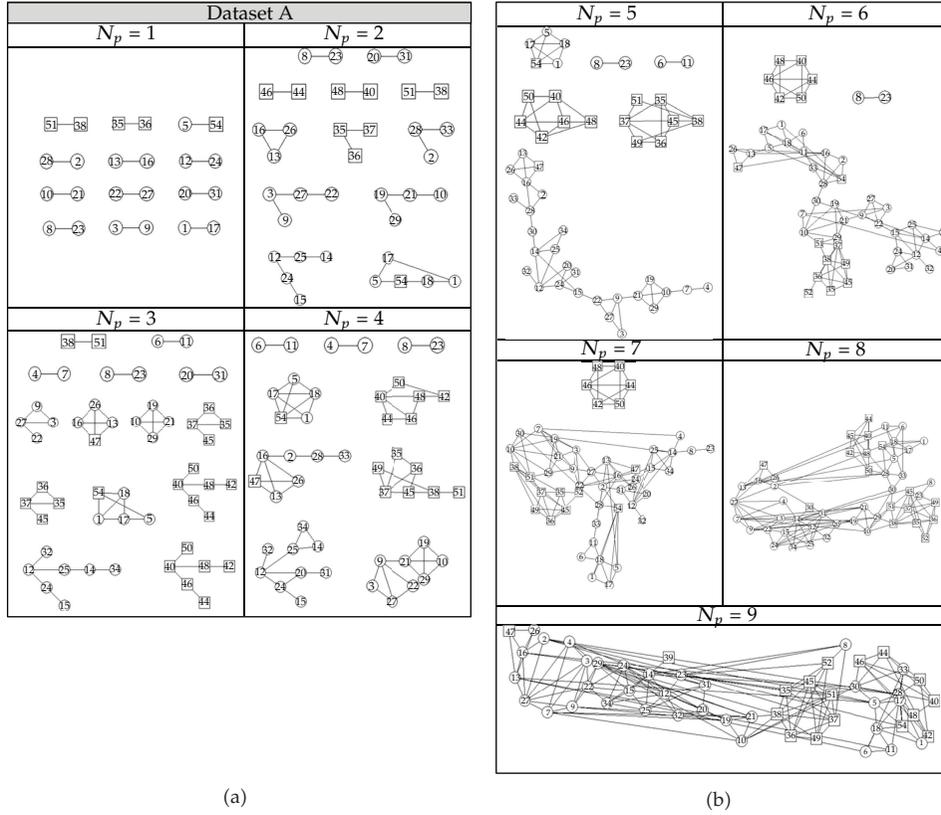


Figure 9: The sociogram of dataset A.

Table 6: The influences of the number of members in a subgroup ($N_p = n, N_g = 10,000$).

n	OPT	AVG	C_v	Average runtime (min)
2	1461	1532.90	0.05	2.14
3	3931	4223.93	0.04	3.01
4	6523	6583.10	0.05	4.05
5	9246	9548.60	0.03	5.05

Table 7: The influences of the number of members ($n = 2$).

m	C_v	Average runtime (min)
10	0	0.51
20	0.13	0.88
30	0.04	1.25
40	0.03	1.63
60	0.04	2.42

5.6. Homogeneous versus Heterogeneous Grouping

A good grouping approach should be able to meet the requirements of a group, such as a homogeneous or heterogeneous grouping. To investigate the effectiveness of the proposed

Table 8: The Influences of scoring function.

m	Scoring function	b	1	2	3	4	5	out	Average runtime (s)
		0	27.1	22.2	17.4	9.9	8.6	22.8	9.6
54	$\{1^2, 2^2, 3^2, 4^2, 5^2, 100\}$	50	28.4	22.2	16.5	10.0	8.2	22.7	22.4
		100	28.3	22.4	16.5	9.6	8.6	22.5	22.3
		200	28.1	21.5	16.9	10.1	8.7	22.7	23.7

Table 9: Overall satisfaction and repeat intention of the CSGA approach.

	Dataset B	Dataset C
Overall satisfaction	4.39	4.28
Repeat intention	4.52	4.09

approach, experiments were performed with different priority weights, which were based on MISSS. The result is illustrated in Figure 10. Members grouped into the same subgroup are circled. A solid line between two members means that members choose mutually. A dashed line with an arrow, on the other hand, represents a one-way choice, where the arrow-headed direction is the choice direction. The grouping can be done with different priority weights based on MISSS. For homogeneous grouping, MISSS can be used as a priority weight. For heterogeneous grouping, MISSS can be sorted and then divided into several parts, with the first part composed of the highest MISSS values, the second part composed of the second highest MISSS values, and the like. Using the proposed CSGA approach, homogeneous or heterogeneous grouping can be easily performed in fulfilling the requirements of a group.

5.7. The Influences of Scoring Function and Penalty Value

The penalty value b was changed to investigate its influences on the grouping results. The results are shown in Table 8. From the table we see that the effect of penalty value b is not very apparent. As for the effect of the scoring function, a squared function converges sooner than a linear function, as illustrated in Figure 11.

5.8. Overall Satisfaction and Repeat Intention

A survey was conducted to understand the overall satisfaction and repeat intention of using the proposed approach after the learning partner groups were decided. A typical five-level Likert item was employed. The question is "In overall, are you satisfied with the method?" and the format of the item is "very satisfied, satisfied, neither satisfied nor dissatisfied, dissatisfied and very dissatisfied." Most of the students are satisfied or very satisfied with the new method. The average points are 4.39 and 4.28, respectively, in dataset B and dataset C, indicating that the students have high satisfaction with the proposed method (see Table 9). As for repeat intention, the question is "Do you agree to use the new method to select learning partners next time?" and anchored by "strongly agree," "agree," "neither agree nor disagree," "disagree," and "strongly disagree." The average points of this question are 4.52 and 4.09, respectively. The high values indicate that the repeat intention of using the proposed method is also very high. The high overall satisfaction and the high repeat intention, in part, reflect the effectiveness of the present approach.

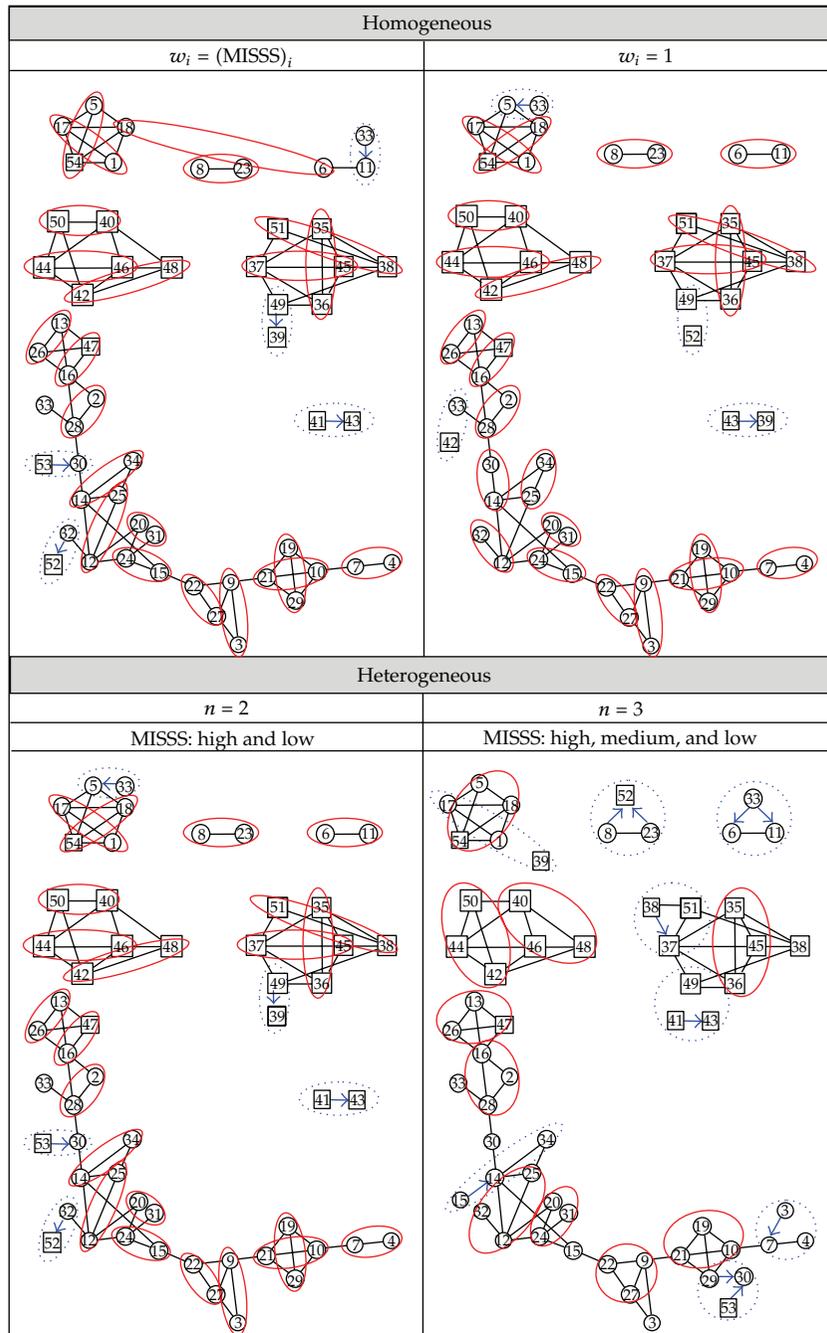


Figure 10: The grouping results are illustrated in a sociogram.

Table 10: Paired samples statistics.

Method	Mean	m	Std. deviation	Std. error mean
Random	3.50	32	.71842	.12700
CSGA	4.28	32	.68318	.12077

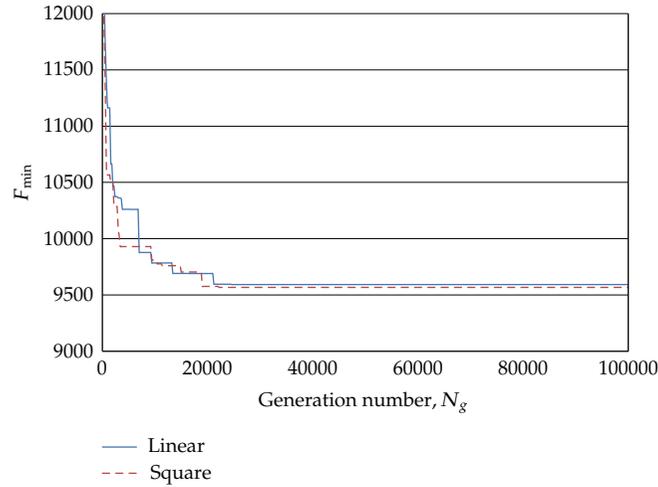


Figure 11: The variation of F_{\min} with generation at different scoring functions.

Table 11: Paired sample t -test.

		Paired Differences				
		Mean	Std. deviation	Std. error mean	95% confidence interval of the difference	
					Lower	Upper
Pair 1	Random—CSGA	-.78125	.87009	.15381	-1.09495	-.46755

Table 12: Significance of paired sample t -test.

		t	df	Sig. (2-tailed)
Pair 1	Random—CSGA	-5.079	31	.000

A paired sample t -test was performed on dataset C to see if there is no difference between these two methods.

Hypothesis 1. There is no difference in satisfaction between using the CSGA approach and using the random method.

The result shows that the overall satisfaction of using the proposed CSGA approach is significantly higher than using the random method, as shown in Tables 10–12.

6. Conclusions

In this paper, we have employed a novel approach to group members based on their social relationships. Members are first asked to show their preferences (choices) to partners with whom they would like to be together in an index system, where a one stands for the first choice, two for the second choice, and so on up to a preassigned number. Subsequently, the choices of members are aggregated and a combined sociometry and genetic algorithm (CSGA) scheme is employed to optimize the grouping. To investigate the effectiveness of the

proposed approach, three real datasets collected from a university were used. In addition, a nonlinear model is developed to calculate the social scores of individuals. Experimental results show that CSGA can optimize the grouping effectively and efficiently and students are very satisfied with the grouping results and show high repeat intention of using it. Moreover, a paired sample *t*-test shows that the overall satisfaction on using the proposed CSGA approach is significantly higher than using the random method.

For further studies, it is recommended to apply the combined sociometry and genetic algorithm approach to solve other kinds of grouping optimization problems. Methods of rating or paired comparisons as well as peer nominations can be used to measure social relationships.

Acknowledgments

The author wishes to express appreciation to the reviewers for their valuable comments and suggestions on this paper. Thanks are also extended to both Ms. Jui-Hui Chen, at the Department of Social and Public Affairs, Taipei Municipal University of Education, and Professor Tung-Shou Chen, at the Department of Computer Science and Information Engineering, National Taichung Institute of Technology, Taiwan, for their suggestions on the paper. Appreciation is also extended to Lindy Chen, Gordon Lee, Jyun-You Fan, Shu-Ping Suen, and Jyun-Yang Li for their help during the course of writing this paper. This work was supported by the National Science Council under Grant no. 100-2221-E-025-016.

References

- [1] E. Turban, D. King, J. Mckey, J. Lee, and D. Viehland, *Electronic Commerce: A Managerial Perspective*, Prentice Hall, New York, NY, USA, 2010.
- [2] P. J. Lamberson, "Social learning in social networks," *The B.E. Journal of Theoretical Economics*, vol. 10, no. 1, article 36, 2010.
- [3] D. Centola, "The spread of behavior in an online social network experiment," *Science*, vol. 329, no. 5996, pp. 1194–1197, 2010.
- [4] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.
- [5] R. S. Burt, *Structural Holes: The Social Structure of Competition*, Harvard University Press, Cambridge, Mass, USA, 1992.
- [6] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Harvard University Press, Cambridge, Mass, USA, 1994.
- [7] B. Wellman, *Networks in the Global Village: Life in Contemporary Communities*, Westview Press, Boulder, Colo, USA, 1999.
- [8] L. C. Freeman, *The Development of Social Network Analysis: A Study in the Sociology of Science*, Empirical Press, Vancouver, Canada, 2004.
- [9] N. C. Mullins, *Theories and Theory Groups in Contemporary American Sociology*, Harper and Row, New York, NY, USA, 1973.
- [10] S. D. Berkowitz, *An Introduction to Structural Analysis: The Network Approach to Social Research*, Butterworth, Toronto, Canada, 1982.
- [11] P. J. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*, Cambridge University Press, New York, NY, USA, 2005.
- [12] J. L. Moreno, *The Sociometry Reader*, The Free Press, Glencoe, Ill, USA, 1960.
- [13] A. H. N. Cillessen, "Sociometric methods," in *Handbook of Peer Interactions, Relationships, and Groups*, p. 83, 2011.
- [14] T. Lappas, K. Liu, and E. Terzi, "Finding a team of experts in social networks," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 467–476, July 2009.

- [15] E. Falkenauer, *Genetic Algorithms and Grouping Problems*, John Wiley & Sons, New York, NY, USA, 1998.
- [16] M. Gallego, M. Laguna, R. Martí, and A. Duarte, "Adaptive memory programming for the maximally diverse grouping problem," in *Proceedings of the 9th Metaheuristic International Conference*, 2011.
- [17] J. L. Moreno, *Who shall survive? A New Approach to Problem of Human Interrelations*, Nervous and Mental Disease Publishing, Washington, DC, USA, 1934.
- [18] A. H. N. Cillessen, "Chapter 5-Sociometric Methods," in *Handbook of Peer Interactions, Relationships, and Groups*, K. H. Rubin, W. M. Bukowski, and B. Laursen, Eds., 2011.
- [19] H. H. Jennings, *Sociometry in Group Relations*, Greenwood, Westport, Conn, USA, 2nd edition, 1987.
- [20] J. D. Coie, K. A. Dodge, and H. Coppotelli, "Dimensions and types of social status: a cross-age perspective," *Developmental Psychology*, vol. 18, no. 4, pp. 557–570, 1982.
- [21] C. H. Ho, *A study of the effects on emotional experience, peer relationship and teacher-student relationship of the therapeutic approach to art activities in class in an elementary school setting*, M.S. thesis, National Pingtung University of Education, Taiwan, 2002.
- [22] J. H. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, UK, 1975.
- [23] D. E. Goldberg, *Genetic Algorithm in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1989.
- [24] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, UK, 1996.
- [25] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Design*, John Wiley & Sons, New York, NY, USA, 1997.
- [26] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Optimization*, John Wiley & Sons, New York, NY, USA, 2000.
- [27] D. A. Coley, *An Introduction to Genetic Algorithms for Scientists and Engineers*, World Scientific, Singapore, 1999.
- [28] M. Gen, R. Cheng, and L. Lin, *Network Models and Optimization: Multiobjective Genetic Algorithm Approach*, Springer, New York, NY, USA, 2008.
- [29] R.-C. Chen, M.-R. Huang, R.-G. Chung, and C.-J. Hsu, "Allocation of short-term jobs to unemployed citizens amid the global economic downturn using genetic algorithm," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7537–7543, 2011.
- [30] V. K. Maheshwari, "Introduction to sociometry," <http://www.vkmaheshwari.com/WP/?p=50>, 2011.
- [31] R. E. Slavin, *Cooperative Learning: Theory, Research, and Practice*, Allyn & Bacon, Boston, Mass, USA, 2th edition, 1995.
- [32] S. Chen, W. Huang, and C. Cattani, "Traffic dynamics on complex networks: a survey," *Mathematical Problems in Engineering*, vol. 2012, Article ID 732698, 23 pages, 2012.
- [33] S. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
- [34] S. 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.
- [35] S. Chen, Y. Wang, and C. Cattani, "Key issues in modeling of complex 3D structures from video sequences," *Mathematical Problems in Engineering*, vol. 2011, Article ID 856523, 12 pages, 2011.

Research Article

Synchronization of Complex Networks with Time-Varying Delayed Dynamical Nodes via Pinning Control

Shuguo Wang,^{1,2} Hongxing Yao,² and Qiuxiang Bian^{2,3}

¹ Department of Mathematics and Physics, Hohai University, Changzhou Campus, Jiangsu, Changzhou 213022, China

² Faculty of Science, Jiangsu University, Jiangsu, Zhenjiang 212013, China

³ Faculty of Mathematics and Physics, Jiangsu University of Science and Technology, Jiangsu, Zhenjiang 212013, China

Correspondence should be addressed to Shuguo Wang, wsg97@163.com

Received 17 August 2011; Accepted 19 October 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Shuguo Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper investigates the pinning synchronization of nonlinearly coupled complex networks with time-varying coupling delay and time-varying delay in dynamical nodes. Some simple and useful criteria are derived by constructing an effective control scheme and adjusting automatically the adaptive coupling strengths. To validate the proposed method, numerical simulation examples are provided to verify the correctness and effectiveness of the proposed scheme.

1. Introduction

In the past few years, the analysis and controllability of complex networks have attracted lots of attention [1–5]. One of the main reasons for that is its wide applications in biology [6], physics, and engineering [7]. So far, many different types of synchronization have been investigated, such as complete synchronization [8], generalized synchronization [9], phase synchronization [10], lag synchronization [11], projection synchronization [12, 13], and so forth. Meanwhile, many control approaches have been developed to synchronize complex networks such as adaptive control [14], pinning control [15, 16], impulsive control [17–20] intermittent control [21, 22], and so on.

For the complexity of the dynamical network, it is difficult to realize the synchronization by adding controllers to all nodes, such as [23]. To reduce the number of the controllers, a natural way is using pinning control method. Guo et al. [24] investigated the

global synchronization of the complex networks with nondelayed and delayed coupling based on the pinning controllers. We can obtain some sufficient conditions for the global synchronization by adding linear and adaptive feedback controllers to some part of nodes from his works. In [25], the authors investigated pinning control for linearly coupled networks and found that one can pin the coupled networks by introducing fewer locally negative feedback controllers. They also compared two different pinning strategies: randomly pinning and selective pinning based on the connection degrees and found out that the pinning strategy based on highest connection degree has better performance than totally randomly pinning. In [26], the authors pinned the complex network to the synchronization manifold by controlling one single node. And in [27], the authors investigated the pinning synchronization of delayed dynamical networks via periodically intermittent control. However, not much has been done on the synchronization of nonlinearly coupled complex networks with time-varying coupling delay and time-varying delay in dynamical nodes by pinning part of nodes and adjusting time-varying coupling strengths.

Motivated by the above discussions, in this paper, we work on the pinning synchronization of nonlinearly coupled complex networks with time-varying coupling delay and time-varying delay in dynamical nodes. The main contributions of this paper are threefold. (1) This paper deals with the synchronization problem for nonlinearly coupled complex networks with time-varying coupling delay and time-varying delay in dynamical nodes. The method used is adaptive pinning controlling method. By adjusting time-varying coupling strengths, some sufficient conditions for the synchronization are derived by constructing an effective control scheme, which are different from the methods used in [28]. (2) By using the Lyapunov stability theorem and a linear matrix inequality (LMI), we prove that the dynamical network can be made to be synchronous with one isolated node by adding controllers to only a small subset of the nodes. Compared with some similar designs, our pinning adaptive controllers are very simple. In addition, the pinning nodes can be randomly selected. It indeed provides some new insights for the future practical engineering design. (3) The Synchronization criteria are independent of time delay. Numerical examples are also provided to demonstrate the effectiveness of the theory.

The rest of this paper is organized as follows. The network model is introduced, and some necessary definitions, lemmas, and hypotheses are given in Section 2. The synchronization of the coupled complex networks is discussed in Section 3. Examples and their simulations are obtained in Section 4. Finally, conclusions are drawn in Section 5.

2. Networks Models and Mathematical Preliminaries

2.1. Model Description

The network with time-varying coupling delay and adaptive coupling strengths can be described by

$$\begin{aligned} \dot{x}_i(t) = & f(t, x_i(t), x_i(t - \tau(t))) + \sigma(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) \\ & + c(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))), \quad i = 1, 2, \dots, N, \end{aligned} \quad (2.1)$$

where $x_i(t) = (x_{i1}(t), x_{i2}(t), \dots, x_{in}(t))^T \in R^n$ is the state vector of the i th dynamical node, $f : R^+ \times R^n \rightarrow R^n$ is an vector value function, $\tau(t) > 0$ is time-varying delay, $\sigma(t) > 0$ and $c(t) > 0$ are time-varying coupling strengths, $H_1(\cdot)$ and $H_2(\cdot)$ are nonlinear functions, and $A = (a_{ij})_{N \times N}$ and $B = (b_{ij})_{N \times N}$ are the weight configuration matrices; if there is a connection from node i to node j ($j \neq i$), then, $a_{ij} = a_{ji} > 0$, $b_{ij} = b_{ji} > 0$, otherwise, $a_{ij} = a_{ji} = 0$, $b_{ij} = b_{ji} = 0$, and the diagonal elements of matrix A, B are defined by

$$a_{ii} = - \sum_{j=1, j \neq i}^N a_{ji}, \quad b_{ii} = - \sum_{j=1, j \neq i}^N b_{ji}, \quad i = 1, 2, \dots, N. \quad (2.2)$$

2.2. Mathematical Preliminaries

Definition 2.1. Assume that $s(t) \in R^n$ is any smooth dynamics. The controlled complex network (2.1) is said to be synchronized on to the homogeneous state $s(t)$ if the solution satisfies $\lim_{t \rightarrow \infty} \|x_i(t) - s(t)\| = 0$ for any initial conditions.

Let $C([- \tau, 0], R^n)$ be the Banach space of continuous functions that map the interval $[- \tau, 0]$ into R^n with norm $\|\phi\| = \sup_{-\tau \leq \theta \leq 0} \|\phi(\theta)\|$. The initial conditions of the functional differential equation (2.1) are given by $x_i(t) = \phi_i(t) \in C([- \tau, 0], R^n)$. It is assumed that (2.1) has a unique solution for these initial conditions.

Suppose that $s(t)$ is a solution of the uncoupled system $\dot{s}(t) = f(s(t))$. In order to pin the system (2.1) onto the synchronization manifold $(s(t), s(t), \dots, s(t))$, we will add the controllers $u_i(t) = k_i(H_1(x_i(t)) - H_1(s(t)))$ to a certain selection of the nodes where k_i will be defined in the following section. Without loss of generality, we add the controllers to the first l nodes ($0 \leq l \leq N$) so that we have

$$\begin{aligned} \dot{x}_i(t) &= f(t, x_i(t), x_i(t - \tau(t))) + \sigma(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) \\ &\quad + c(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))) - u_i(t), \quad i = 1, 2, \dots, l, \\ \dot{x}_i(t) &= f(t, x_i(t), x_i(t - \tau(t))) + \sigma(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) \\ &\quad + c(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))), \quad i = l + 1, l + 2, \dots, N. \end{aligned} \quad (2.3)$$

Noting that $e_i(t) = x_i(t) - s(t)$, we have

$$\begin{aligned} \dot{e}_i(t) &= \tilde{f}(t, x_i(t), x_i(t - \tau(t))) + \sigma(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) \\ &\quad + c(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))) - u_i(t), \quad i = 1, 2, \dots, l, \end{aligned}$$

$$\begin{aligned} \dot{x}_i(t) = & \tilde{f}(t, x_i(t), x_i(t - \tau(t))) + \sigma(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) \\ & + c(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))), \quad i = l + 1, l + 2, \dots, N, \end{aligned} \quad (2.4)$$

where $\tilde{f}(t, x_i(t), x_i(t - \tau(t))) = f(t, x_i(t), x_i(t - \tau(t))) - f(t, s(t), s(t - \tau(t)))$.

We now introduce some definitions, assumptions, and lemmas that will be required throughout this paper.

Lemma 2.2 (see [29]). *Assuming that $A = (a_{ij}) \in \mathbb{R}^{N \times N}$ satisfies the following conditions:*

- (1) $A = (a_{ij}) \in \mathbb{R}^{N \times N}$, $a_{ij} = a_{ji} > 0$, $a_{ii} = -\sum_{j=1, j \neq i}^N a_{ij}$, $i = 1, 2, \dots, N$,
- (2) A is irreducible.

Then we have that

- (i) the real parts of the eigenvalues of A are all negative except an eigenvalue 0 with multiplicity 1.
- (ii) A has a right eigenvector $(1, 1, \dots, 1)^T$ corresponding to the eigenvalue 0.

Lemma 2.3 (see [24]). *If $G = (g_{ij})_{N \times N}$ is an irreducible matrix that satisfies $g_{ij} = g_{ji} \geq 0$ for $i \neq j$, and $g_{ii} = -\sum_{j=1, i \neq j}^N g_{ij}$, for $i = 1, 2, \dots, N$, then all the eigenvalues of the matrix*

$$\tilde{G} = \begin{pmatrix} g_{11} - \varepsilon_1 & \cdots & g_{1N} \\ \vdots & \ddots & \vdots \\ g_{N1} & \cdots & g_{NN} - \varepsilon_N \end{pmatrix} \quad (2.5)$$

are negative, where $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ are nonnegative constants and $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N > 0$.

Lemma 2.4 (see [30]). *For any two vectors x and y , a matrix $S > 0$ with compatible dimensions, one has: $2x^T y \leq x^T S x + y^T S^{-1} y$.*

Lemma 2.5 (see [31]). *Let Q and R be two symmetric matrices, and matrix S has suitable dimension. Then*

$$\begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} < 0 \quad (2.6)$$

if and only if both $R < 0$ and $Q - SR^{-1}S^T < 0$.

Lemma 2.6 (see [32]). For the vector-valued function $f(t, x_i(t), x_i(t - \tau(t)))$, assuming that there exist positive constants $\gamma > 0$, $\theta > 0$ such that f satisfies the semi-Lipschitz condition

$$\begin{aligned} & (x_i(t) - y_i(t))^T (f(t, x_i(t), x_i(t - \tau(t))) - f(t, y_i(t), y_i(t - \tau(t)))) \\ & \leq \gamma (x_i(t) - y_i(t))^T (x_i(t) - y_i(t)) + \theta (x_i(t - \tau(t)) - y_i(t - \tau(t)))^T (x_i(t - \tau(t)) - y_i(t - \tau(t))), \end{aligned} \quad (2.7)$$

for all $x, y \in R^n$ and $t \geq 0$, $i = 1, 2, \dots, N$.

Assumption 2.7 (see [9] (Global Lipschitz Condition)). Suppose that there exist nonnegative constants α, β , for all $t \in R_+$, such that for any time-varying vectors $x(t), y(t) \in R^n$

$$\|H_1(x) - H_1(y)\| \leq \alpha \|x - y\|, \quad \|H_2(x) - H_2(y)\| \leq \beta \|x - y\|, \quad (2.8)$$

where $\|\cdot\|$ denotes the 2 norm throughout the paper.

Assumption 2.8 (see [9]). $\tau(t)$ is differential function with $0 \leq \dot{\tau}(t) \leq \varepsilon \leq 1$. Clearly, this assumption is certainly to ensure that the delay $\tau(t)$ is constant.

3. Main Results

In this section, a control scheme is developed to synchronize a delayed complex network with time-varying delay dynamical nodes to any smooth dynamics $s(t)$. The local controllers are designed as follows:

$$u_i(t) = k_i(H_1(x_i(t)) - H_1(s(t))), \quad (3.1)$$

we let

$$\tilde{A} = \begin{pmatrix} a_{11} - \frac{k_1}{\mu_1} & \dots & a_{1l} & a_{1,l+1} & \dots & a_{1N} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ a_{l1} & \dots & a_{ll} - \frac{k_l}{\mu_l} & a_{l,l+1} & \dots & a_{lN} \\ a_{l+1,1} & \dots & a_{l+1,l} & a_{l+1,l+1} & \dots & a_{l+1,N} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ a_{N1} & \dots & a_{Nl} & a_{N,l+1} & \dots & a_{NN} \end{pmatrix}, \quad (3.2)$$

where k_1, k_2, \dots, k_l are positive constants. It follows from Lemma 2.3 that if A is symmetric and irreducible, then \tilde{A} is negative definite. We can therefore prove the following theorem.

For convenience in later use, we denote

$$\begin{aligned}\tilde{x}_i(t) &= (x_{1i}(t), x_{2i}(t), \dots, x_{N_i}(t))^T, & \tilde{e}_i(t) &= (e_{1i}(t), e_{2i}(t), \dots, e_{N_i}(t))^T, \\ H_1(\tilde{x}_i(t)) &= (H_1(x_{1i}(t)), H_1(x_{2i}(t)), \dots, H_1(x_{N_i}(t)))^T, & (3.3) \\ H_2(\tilde{x}_i(t)) &= (H_2(x_{1i}(t)), H_2(x_{2i}(t)), \dots, H_2(x_{N_i}(t)))^T, & \tilde{e}_i = \tilde{x}_i(t) - \tilde{s}_i(t).\end{aligned}$$

Theorem 3.1. *Suppose that Assumptions 2.7 and 2.8 hold, the adaptive coupling strength*

$$\sigma(t) = \hat{\rho}(t) + \mu_1, \quad c(t) = \hat{\eta}(t) + \mu_2, \quad (3.4)$$

where $\mu = \max(\mu_1, \mu_2)$, $\mu_1 > 0$, $\mu_2 > 0$, and $\hat{\rho}(0) > 0$, $\hat{\eta}(0) > 0$. If $\hat{\rho}(t)$, $\hat{\eta}(t)$ satisfies the following update law: $\dot{\hat{\rho}}(t) = -\sum_{i=1}^N e_i^T(t)e_i(t)$, $\dot{\hat{\eta}}(t) = -\sum_{i=1}^N [e_i^T(t)e_i(t) + e_i^T(t - \tau(t))e_i(t - \tau(t))]$ and the following condition hold:

$$\mu_1 \alpha \tilde{A} + \frac{1}{2(1-\varepsilon)} I_N + \frac{1}{2} \beta^2 \mu_2^2 B^T B < 0. \quad (3.5)$$

then the controlled dynamical network (2.1) is synchronization.

Proof. Construct the following Lyapunov function:

$$\begin{aligned}V(t) &= \frac{1}{2} \sum_{i=1}^N e_i^T(t)e_i(t) + \frac{\theta}{(1-\varepsilon)} \sum_{i=1}^N \int_{t-\tau(t)}^t e_i^T(\xi)e_i(\xi) d\xi \\ &+ \frac{1}{2(1-\varepsilon)} \sum_{j=1}^n \int_{t-\tau(t)}^t \tilde{e}_j^T(\xi)\tilde{e}_j(\xi) d\xi + \frac{\tilde{\rho}^2}{2\beta_1} + \frac{\tilde{\eta}^2}{2\beta_2},\end{aligned} \quad (3.6)$$

where $\tilde{\rho}(t) = -\beta_1 \sum_{i=1}^N e_i^T(t)e_i(t)$, $\tilde{\eta}(t) = -\beta_2 \sum_{i=1}^N [e_i^T(t)e_i(t) + e_i^T(t - \tau(t))e_i(t - \tau(t))]$, $\tilde{\rho}(t) = (\gamma + (\theta/(1-\varepsilon))) + \alpha \lambda_1 \hat{\rho}(t)$, $\tilde{\eta}(t) = (1/2)\beta \lambda_2 \hat{\eta}(t)$, $\lambda_1 = \max(|\lambda_1(A)|, |\lambda_2(A)|, \dots, |\lambda_N(A)|)$, $\beta_1 = \alpha \lambda_1$,

$\lambda_2 = \max(|\lambda_1(B)|, |\lambda_2(B)|, \dots, |\lambda_N(B)|)$, and $\beta_2 = (1/2)\beta\lambda_2$. In virtue of Assumptions 2.7 and 2.8, the time derivative of $V(t)$ along the trajectory of system (2.4) is derived by

$$\begin{aligned}
\dot{V}(t) &= \sum_{i=1}^N e_i^T(t) \left(\tilde{f}(t, x_i(t), x_i(t - \tau(t))) + \sigma(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) + c(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))) - u_i(t) \right) \\
&\quad + \frac{\theta}{(1 - \varepsilon)} \sum_{i=1}^N \left(e_i^T(t) e_i(t) - (1 - \dot{\tau}(t)) e_i^T(t - \tau(t)) e_i(t - \tau(t)) \right) \\
&\quad + \frac{1}{2(1 - \varepsilon)} \sum_{j=1}^n \left(\tilde{e}_j^T(t) \tilde{e}_j(t) - (1 - \dot{\tau}(t)) \tilde{e}_j^T(t - \tau(t)) \tilde{e}_j(t - \tau(t)) \right) + \tilde{\rho}(t) \dot{\tilde{\rho}}(t) + \tilde{\eta}(t) \dot{\tilde{\eta}}(t) \\
&= \sum_{i=1}^N e_i^T(t) \tilde{f}(t, x_i(t), x_i(t - \tau(t))) + \sum_{i=1}^N e_i^T(t) \sigma(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) \\
&\quad + \sum_{i=1}^N e_i^T(t) c(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))) - \sum_{i=1}^l k_i e_i^T(t) (H_1(x_i(t)) - H_1(s(t))) \\
&\quad + \frac{\theta}{(1 - \varepsilon)} \sum_{i=1}^N e_i^T(t) e_i(t) - \frac{\theta(1 - \dot{\tau}(t))}{(1 - \varepsilon)} \sum_{i=1}^N e_i^T(t - \tau(t)) e_i(t - \tau(t)) \\
&\quad + \frac{1}{2(1 - \varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{e}_j(t) - \frac{1 - \dot{\tau}(t)}{2(1 - \varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t - \tau(t)) \tilde{e}_j(t - \tau(t)) + \tilde{\rho}(t) \dot{\tilde{\rho}}(t) + \tilde{\eta}(t) \dot{\tilde{\eta}}(t) \\
&\leq \sum_{i=1}^N \left[\gamma e_i(t)^T e_i(t) + \theta e_i^T(t - \tau(t)) e_i(t - \tau(t)) \right] + \sum_{i=1}^N e_i^T(t) \hat{\rho}(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) \\
&\quad + \mu_1 \sum_{i=1}^N e_i^T(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) + \sum_{i=1}^N e_i^T(t) \hat{\eta}(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))) \\
&\quad + \mu_2 \sum_{i=1}^N e_i^T(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))) - \sum_{i=1}^l k_i e_i^T(t) (H_1(x_i(t)) - H_1(s(t))) \\
&\quad + \frac{\theta}{(1 - \varepsilon)} \sum_{i=1}^N e_i^T(t) e_i(t) - \frac{\theta(1 - \dot{\tau}(t))}{(1 - \varepsilon)} \sum_{i=1}^N e_i^T(t - \tau(t)) e_i(t - \tau(t)) \\
&\quad + \frac{1}{2(1 - \varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{e}_j(t) - \frac{1 - \dot{\tau}(t)}{2(1 - \varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t - \tau(t)) \tilde{e}_j(t - \tau(t)) + \tilde{\rho}(t) \dot{\tilde{\rho}}(t) + \tilde{\eta}(t) \dot{\tilde{\eta}}(t) \\
&\leq \gamma \sum_{i=1}^N e_i(t)^T e_i(t) + \theta \sum_{i=1}^N e_i(t - \tau(t))^T e_i(t - \tau(t)) + \mu_1 \sum_{i=1}^n \tilde{e}_j^T(t) \tilde{A} (H_1(\tilde{x}_j(t)) - H_1(\tilde{s}(t)))
\end{aligned}$$

$$\begin{aligned}
& + \mu_2 \sum_{i=1}^n \tilde{e}_j^T(t) B (H_2(\tilde{x}_j(t - \tau(t))) - H_2(\tilde{s}_j(t - \tau(t)))) \\
& + \frac{\theta}{(1 - \varepsilon)} \sum_{i=1}^N e_i^T(t) e_i(t) - \frac{\theta(1 - \dot{\tau}(t))}{(1 - \varepsilon)} \sum_{i=1}^N e_i^T(t - \tau(t)) e_i(t - \tau(t)) \\
& + \frac{1}{2(1 - \varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{e}_j(t) - \frac{1 - \dot{\tau}(t)}{2(1 - \varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t - \tau(t)) \tilde{e}_j(t - \tau(t)) + \tilde{\rho}(t) \dot{\hat{\rho}}(t) + \tilde{\eta}(t) \dot{\hat{\eta}}(t) \\
& + \sum_{i=1}^N e_i^T(t) \hat{\rho}(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) + \sum_{i=1}^N e_i^T(t) \hat{\eta}(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))) \\
& \leq \left(\gamma + \frac{\theta}{(1 - \varepsilon)} \right) \sum_{i=1}^N e_i(t)^T e_i(t) + \mu_1 \alpha \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{A} \tilde{e}_j(t) + \mu_2 \beta \sum_{j=1}^n \tilde{e}_j^T(t) B \tilde{e}_j(t - \tau(t)) \\
& + \frac{1}{2(1 - \varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{e}_j(t) - \frac{1}{2} \sum_{j=1}^n \tilde{e}_j^T(t - \tau(t)) \tilde{e}_j(t - \tau(t)) + \tilde{\rho}(t) \dot{\hat{\rho}}(t) + \tilde{\eta}(t) \dot{\hat{\eta}}(t) \\
& + \sum_{i=1}^N e_i^T(t) \hat{\rho}(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) + \sum_{i=1}^N e_i^T(t) \hat{\eta}(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))).
\end{aligned} \tag{3.7}$$

Let $e(t) = (e_1^T(t), e_2^T(t), \dots, e_N^T(t))^T \in R^{nN}$,

$$\begin{aligned}
\dot{V}(t) & \leq \left(\gamma + \frac{\theta}{(1 - \varepsilon)} \right) \sum_{i=1}^N e_i(t)^T e_i(t) + \mu_1 \alpha \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{A} \tilde{e}_j(t) + \mu_2 \beta \sum_{j=1}^n \tilde{e}_j^T(t) B \tilde{e}_j(t - \tau(t)) \\
& + \frac{1}{2(1 - \varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{e}_j(t) - \frac{1}{2} \sum_{j=1}^n \tilde{e}_j^T(t - \tau(t)) \tilde{e}_j(t - \tau(t)) + \tilde{\rho}(t) \dot{\hat{\rho}}(t) + \tilde{\eta}(t) \dot{\hat{\eta}}(t) \\
& + \alpha \hat{\rho}(t) e^T(t) A e(t) + \beta \hat{\eta}(t) e^T(t) B e(t - \tau(t)) \\
& \leq \mu_1 \alpha \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{A} \tilde{e}_j(t) + \mu_2 \beta \sum_{j=1}^n \tilde{e}_j^T(t) B \tilde{e}_j(t - \tau(t)) + \frac{1}{2(1 - \varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{e}_j(t) \\
& - \frac{1}{2} \sum_{j=1}^n \tilde{e}_j^T(t - \tau(t)) \tilde{e}_j(t - \tau(t)) + \tilde{\eta}(t) \dot{\hat{\eta}}(t) + \frac{1}{2} \beta \lambda_2 \hat{\eta}(t) (e^T(t) e(t) + e^T(t - \tau(t)) e(t - \tau(t)))
\end{aligned}$$

$$\begin{aligned}
&\leq \mu_1 \alpha \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{A} \tilde{e}_j(t) + \mu_2 \beta \sum_{j=1}^n \tilde{e}_j^T(t) B \tilde{e}_j(t - \tau(t)) + \frac{1}{2(1-\varepsilon)} \sum_{j=1}^n \tilde{e}_j^T(t) \tilde{e}_j(t) \\
&\quad - \frac{1}{2} \sum_{j=1}^n \tilde{e}_j^T(t - \tau(t)) \tilde{e}_j(t - \tau(t)) \\
&\leq \sum_{j=1}^n \tilde{e}_j^T(t) \left(\mu_1 \alpha \tilde{A} + \frac{1}{2(1-\varepsilon)} I \right) \tilde{e}_j(t) + \beta \mu_2 \sum_{j=1}^n \tilde{e}_j^T(t) B \tilde{e}_j(t - \tau(t)) \\
&\quad - \frac{1}{2} \sum_{j=1}^n \tilde{e}_j^T(t - \tau(t)) \tilde{e}_j(t - \tau(t)) \\
&\leq \sum_{j=1}^n \left(\tilde{e}_j^T(t), \tilde{e}_j^T(t - \tau(t)) \right) Z_j \left(\tilde{e}_j(t), \tilde{e}_j(t - \tau(t)) \right),
\end{aligned} \tag{3.8}$$

where

$$Z_j = \begin{pmatrix} \mu_1 \alpha \tilde{A} + \frac{1}{2(1-\varepsilon)} I_N & \frac{1}{2} \beta \mu_2 B \\ \frac{1}{2} \beta \mu_2 B^T & -\frac{1}{2} I_N \end{pmatrix}. \tag{3.9}$$

It follows from Lemma 2.5 that if

$$\mu_1 \alpha \tilde{A} + \frac{1}{2(1-\varepsilon)} I_N + \frac{1}{2} \beta^2 \mu_2^2 B^T B < 0, \tag{3.10}$$

we obtain $\dot{V}(t) < 0$. The proof is completed. \square

Remark 3.2. Compared with the other control methods in the literature, pinning controller is relatively simple and is easy to implement. As we know now, the real-world complex networks normally have a large number of nodes. Therefore, it is usually difficult to control a complex network by adding the controllers to all nodes. To reduce the number of the controllers, a natural approach is to control a complex network by pinning part of nodes. In this paper, we designed controllers to ensure that the special networks could get synchronization. The pinning nodes can be randomly selected. It indeed provides some new insights for the future practical engineering design.

Remark 3.3. Synchronization criteria have been given in Theorem 3.1. Compared with some similar designs, our pinning adaptive controllers are very simple. In particular, the synchronization criteria are independent of time delays.

4. Illustrative Examples

In this section, a numerical example will be given to demonstrate the validity of the synchronization criteria obtained in the previous sections. Considering the following network:

$$\begin{aligned}
 \dot{x}_i(t) &= f(t, x_i(t), x_i(t - \tau(t))) + \sigma(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) \\
 &\quad + c(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))) - u_i(t), \quad i = 1, 2, \dots, l, \\
 \dot{x}_i(t) &= f(t, x_i(t), x_i(t - \tau(t))) + \sigma(t) \sum_{j=1}^N a_{ij} H_1(x_j(t)) \\
 &\quad + c(t) \sum_{j=1}^N b_{ij} H_2(x_j(t - \tau(t))), \quad i = l + 1, l + 2 \dots N,
 \end{aligned} \tag{4.1}$$

where $f(t, x_i(t), x_i(t - \tau(t))) = Dx_i(t) + h_1(x_i(t)) + h_2(x_i(t - \tau(t)))$ here $\tau(t) = e^t/100(1 + e^t)$, $x_i(t) = (x_{i1}(t), x_{i2}(t), x_{i3}(t))^T$, $h_1(x_i) = (0, -x_{i1}x_{i3}, x_{i1}x_{i2})^T$, $h_2(x_i) = (0, 0.2x_{i2}, 0)^T$, $\mu = 5$, $N = 4$, $l = 1$, $H_1(x) = \sin x + 2x$, $H_2(x) = \cos x + 3x$, and

$$\begin{aligned}
 D &= \begin{bmatrix} -10 & 10 & 0 \\ 28 & -1 & 0 \\ 0 & 0 & -\frac{8}{3} \end{bmatrix}, \\
 A = B &= \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -3 & 1 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 1 & 1 & -2 \end{bmatrix}.
 \end{aligned} \tag{4.2}$$

5. Conclusions

The problems of synchronization and pinning control for the nonlinearly coupled complex networks with time-varying coupling delay and time-varying delay in dynamical nodes are

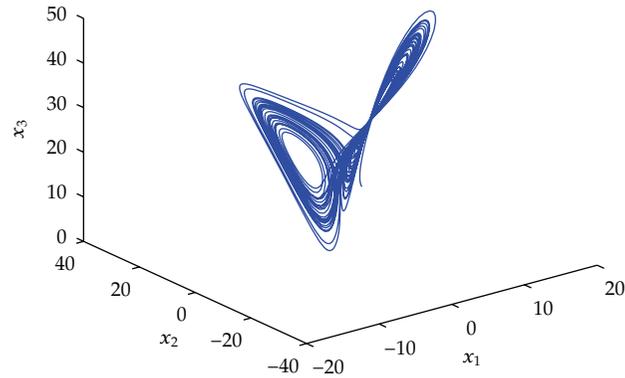


Figure 1: The chaotic behavior of time-delayed Lorenz system.

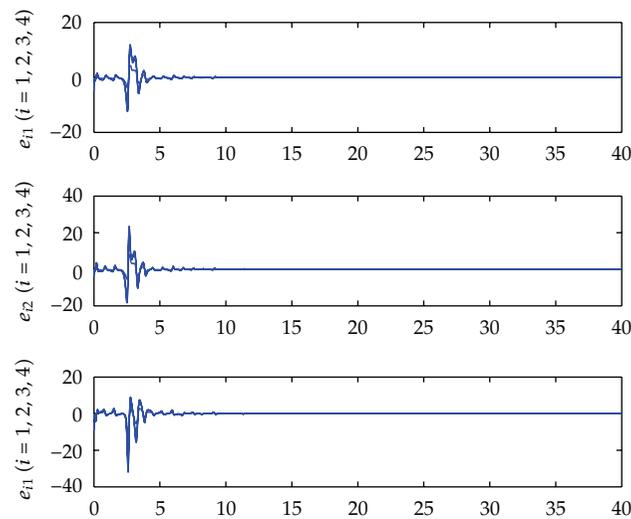


Figure 2: Time evolution of the synchronization errors.

investigated. It is shown that synchronization can be realized via adjusting time-varying coupling strengths. The study showed that the use of simple control law helps to derive sufficient criteria which ensure that the synchronization of the network model is derived. In addition numerical simulations were performed to verify the effectiveness of the theoretical results. Compared with existing results, our synchronization is still very useful when the existing methods become invalid (Figures 1 and 2).

Acknowledgments

Research is partially supported by the National Nature Science Foundation of China (no. 70871056) and by the Six Talents Peak Foundation of Jiangsu Province. The authors are also grateful to Miss Mary Opokua Ansong, at the Department of Computer Science and Technology, Jiangsu University, for making time to go through this work.

References

- [1] F. Sorrentino, "Effects of the network structural properties on its controllability," *Chaos*, vol. 17, no. 3, Article ID 033101, 2007.
- [2] M. Porfiri and M. D. Bernardo, "Criteria for global pinning-controllability of complex networks," *Automatica*, vol. 44, no. 12, pp. 3100–3106, 2008.
- [3] S. Y. Chen, W. Huang, and C. Cattani, "Traffic dynamics on complex networks: a survey," *Mathematical Problems in Engineering*, vol. 2012, Article ID 732698, 23 pages, 2012.
- [4] S. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
- [5] F. Sorrentino, M. Di Bernardo, F. Garofalo, and G. Chen, "Controllability of complex networks via pinning," *Physical Review E*, vol. 75, no. 4, Article ID 046103, 2007.
- [6] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press, Oxford, UK, 2003.
- [7] P. M. Gade, "Synchronization of oscillators with random nonlocal connectivity," *Physical Review E*, vol. 54, no. 1, pp. 64–70, 1996.
- [8] J. Lu, D. W. C. Ho, and M. Liu, "Globally exponential synchronization in an array of asymmetric coupled neural networks," *Physics Letters, Section A*, vol. 369, no. 5-6, pp. 444–451, 2007.
- [9] H. Liu, J. Chen, J. A. Lu, and M. Cao, "Generalized synchronization in complex dynamical networks via adaptive couplings," *Physica A*, vol. 389, no. 8, pp. 1759–1770, 2010.
- [10] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "Phase synchronization of chaotic oscillators," *Physical Review Letters*, vol. 76, no. 11, pp. 1804–1807, 1996.
- [11] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "From phase to lag synchronization in coupled chaotic oscillators," *Physical Review Letters*, vol. 78, no. 22, pp. 4193–4196, 1997.
- [12] R. Zhang, Y. Yang, Z. Xu, and M. Hu, "Function projective synchronization in drive-response dynamical network," *Physics Letters, Section A*, vol. 374, no. 30, pp. 3025–3028, 2010.
- [13] S. Zheng, "Adaptive-impulsive control of the projective synchronization in drive-response complex dynamical networks with time-varying coupling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 501843, 12 pages, 2012.
- [14] J. Zhou, J. Lu, and J. Lü, "Adaptive synchronization of an uncertain complex dynamical network," *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, vol. 51, no. 4, pp. 652–656, 2006.
- [15] L. Pan, W. Zhou, J. Fang, and D. Li, "Synchronization and anti-synchronization of new uncertain fractional-order modified unified chaotic systems via novel active pinning control," *Communications in Nonlinear Science and Numerical Simulation*, vol. 15, no. 12, pp. 3754–3762, 2010.
- [16] S. Zheng and Q.S. Bi, "Synchronization analysis of complex dynamical networks with delayed and non-delayed coupling based on pinning control," *Physica Scripta*, vol. 84, Article ID 025008, 2011.
- [17] G. Liu and W. Ding, "Impulsive synchronization for a chaotic system with channel time-delay," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 2, pp. 958–965, 2011.
- [18] L. X. Guo, M. F. Hu, and Z. Y. Xu, "Impulsive synchronization and control of directed transport in chaotic ratchets," *Chinese Physics B*, vol. 19, no. 2, Article ID 020512, 2010.
- [19] S. Zheng, "Adaptive-impulsive projective synchronization of drive-response delayed complex dynamical networks with time-varying coupling," *Nonlinear Dynamics*. In press.
- [20] S. G. Wang and H. X. Yao, "Impulsive synchronization of two coupled complex networks with time-delayed dynamical nodes," *Chinese Physics B*, vol. 20, no. 9, Article ID 090513, 2011.
- [21] A. A. Syntetos and J. E. Boylan, "On the stock control performance of intermittent demand estimators," *International Journal of Production Economics*, vol. 103, no. 1, pp. 36–47, 2006.
- [22] S. M. Cai, J. Hao, Q. He, and Z. Liu, "Exponential synchronization of complex delayed dynamical networks via pinning periodically intermittent control," *Physics Letters, Section A*, vol. 375, no. 19, pp. 1965–1971, 2011.
- [23] T. Liu, J. Zhao, and D. J. Hill, "Synchronization of complex delayed dynamical networks with nonlinearly coupled nodes," *Chaos, Solitons and Fractals*, vol. 40, no. 3, pp. 1506–1519, 2009.
- [24] W. Guo, F. Austin, S. Chen, and W. Sun, "Pinning synchronization of the complex networks with non-delayed and delayed coupling," *Physics Letters, Section A*, vol. 373, no. 17, pp. 1565–1572, 2009.
- [25] X. F. Wang and G. Chen, "Pinning control of scale-free dynamical networks," *Physica A*, vol. 310, no. 3-4, pp. 521–531, 2002.
- [26] T. Chen, X. Liu, and W. Lu, "Pinning complex networks by a single controller," *IEEE Transactions on Circuits and Systems. I*, vol. 54, no. 6, pp. 1317–1326, 2007.

- [27] W. W. Yu, G. Chen, and J. H. Lue, "On pinning synchronization of complex dynamical networks," *Automatica*, vol. 45, no. 2, pp. 429–435, 2009.
- [28] W. W. Yu, J. Cao, and J. Lü, "Global synchronization of linearly hybrid coupled networks with time-varying delay," *SIAM Journal on Applied Dynamical Systems*, vol. 7, no. 1, pp. 108–133, 2008.
- [29] W. Lu and T. Chen, "New approach to synchronization analysis of linearly coupled ordinary differential systems," *Physica D.*, vol. 213, no. 2, pp. 214–230, 2006.
- [30] Z. Zuo, C. Yang, and Y. Wang, "A unified framework of exponential synchronization for complex networks with time-varying delays," *Physics Letters, Section A*, vol. 374, no. 19-20, pp. 1989–1999, 2010.
- [31] J. Zhou and T. Chen, "Synchronization in general complex delayed dynamical networks," *IEEE Transactions on Circuits and Systems. I*, vol. 53, no. 3, pp. 733–744, 2006.
- [32] C. Hu, J. Yu, H. Jiang, and Z. Teng, "Synchronization of complex community networks with nonidentical nodes and adaptive coupling strength," *Physics Letters, Section A*, vol. 375, no. 5, pp. 873–879, 2011.

Research Article

Parameterization Method on B-Spline Curve

H. Haron,¹ A. Rehman,² D. I. S. Adi,¹ S. P. Lim,¹ and T. Saba¹

¹ Faculty of Computer Sciences and Information Systems, Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru, Malaysia

² College of Computer and Information Science, Al-Imam Muhammad Bin Saud Islamic University, Riyadh 11432, Saudi Arabia

Correspondence should be addressed to T. Saba, tanzilasaba@yahoo.com

Received 18 July 2011; Revised 30 September 2011; Accepted 7 October 2011

Academic Editor: Carlo Cattani

Copyright © 2012 H. Haron et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The use of computer graphics in many areas allows a real object to be transformed into a three-dimensional computer model (3D) by developing tools to improve the visualization of two-dimensional (2D) and 3D data from series of data point. The tools involved the representation of 2D and 3D primitive entities and parameterization method using B-spline interpolation. However, there is no parameterization method which can handle all types of data points such as collinear data points and large distance of two consecutive data points. Therefore, this paper presents a new parameterization method that is able to solve those drawbacks by visualizing the 2D primitive entity of scanned data point of a real object and construct 3D computer model. The new method has improved a hybrid method by introducing exponential parameterization method in the beginning of the reconstruction process, followed by computing B-spline basis function to find maximum value of the function. The improvement includes solving a linear system of the B-spline basis function using numerical method. Improper selection of the parameterization method may lead to the singularity matrix of the system linear equations. The experimental result on different datasets show that the proposed method performs better in constructing the collinear and two consecutive data points compared to few parameterization methods.

1. Introduction

The usage of computer graphics allows real life to be modeled in a computer model. A computer model makes the work much easier and more effective. At the beginning, the computer model is visualized as wire-frame model then rigged and rendered at the end. In addition, animation of the computer model can be added to bring a computer model close to the original model. There are many applications in engineering, scientific visualization, medical imaging, image processing, design, and entertainment industry that use the aid of computer graphics. The sophisticated computer graphics techniques, which involve the geometric modeling of curves and surfaces, differential geometry, and so forth, are needed

for realistic 3D modeling such as virtual reality and computer-aided design (CAD). One of example of emerging application in CAD is the reverse engineering process in manufacturing industry.

The reverse engineering process is the creation process of computer model from its original real-life model or its prototypes. Basically, the reverse engineering process is started from a point which could be obtained from a scanner device then locally modeled to be interpolated or approximated by using surface patches. The interpolation and approximation are done by performing parameterization over the data points at the beginning. This paper presents new parameterization method on B-spline curve. Accordingly, this paper is further organized into five sections. Section 2 presents state of the art on parameterization methods; Section 3 presents proposed techniques and methodology; Section 4 exhibits analysis and discussion of results along with detail pictures of experimented data. Finally, conclusion and future work are presented in Section 5.

2. Background

2.1. Parameterization Methods

The parameterization of 2D or 3D data points is one of the fundamental steps in many CAGD applications such as texture mapping, remeshing, typography, and curve and surface fitting [1–3]. For example, in [4], they applied a new method to enhance the accuracy of image fitting which is based on boundary moment invariants. While for [5], they had derived an equation to compute the volume of the NURBS shape and obtained more accurate results.

The parameter value reflects the distribution of the data points. Some previous research works have been done on parameterization method starting from the easiest method such as uniform [6], intermediate method, for instance, exponential [7], universal [8], and hybrid [9], until complex calculation such as rational chord length parameterization [10]. The most common method used in solving the parameterization problems was centripetal method [11]. Most of the works were focusing on general curves while for the closed curve is as explained in [12]. There are some works based on best parameterization method in [2]. However, the best approximation method is still depending on the designer. There are many popular parameterization methods such as uniform, chord length, and centripetal. However, these parameterizations presented that the state of the art contains pros and cons as stipulated in Table 1.

2.2. B-spline

As Bezier curves cannot be modified locally and the movement of the control points will affect the whole curve shape [13], hence, the researchers try to solve the problems. Finally, B-spline method has been generalised from Bezier method to overcome the problems faced and it can be up to C^2 continuity [14].

As discussed in [13, 15, 16], a d th-degree B-spline curve can be defined as

$$P(u) = \sum_{i=0}^{n+1} P_i N_i, d(u) \quad 0 \leq u \leq u_{\max}, \quad (2.1)$$

Table 1: Comparison among parameterization methods.

Parameterization Method	Advantages	Disadvantages
Uniform [6]	The parameter value is easy to calculate	In many cases, interpolation between edit points is not as good and can lead to unpredictable stretching of textures during rendering.
Chord length [10]	Minimizes stretching and squeezing of textures	A large bulge and twists away from the data point polygon may be generated.
Centripetal [11]	In some cases, may make smoother objects. On surface of very evenly spaced control points, it may give slightly better texture mapping results	A small loop may be produced.
Exponential [7]	Can freely select the resulting curve, in other word, generalization of uniform, chord length, and centripetal method	Difficult to find the suitable value of α for better curve result
Universal [8]	Affine invariant	Similar to uniform parameterization method
Hybrid [9]	Give better accuracy	Unstable and more computation effort due to the calculation of centripetal and chord length method.

and the Basis Function are recursively defined as [15]

$$N_{i,d}(u) = \frac{u - u_i}{u_{i+d-1} - u_i} N_{i,d-1}(u) + \frac{u_{i+d} - u}{u_{i+d} - u_{i+1}} N_{i+1,d-1}(u), \quad (2.2)$$

$$N_{i,1}(u) = \begin{cases} 1, & \text{if } u_i \leq u \leq u_{i+1}, \\ 0, & \text{otherwise,} \end{cases} \quad (2.3)$$

where P_i are the control points, for example, P_0, \dots, P_n , $N_{i,d}$ are the d th-degree B-spline basis function, U is the knot vector with the range $U = \{0, \dots, u_{\max}\}$.

As mentioned in [15], B-spline curve can be defined as a polynomial spline function of order d (degree $d-1$) and composed of $n-d+2$ [13] when it satisfies the condition: $P(u)$ with degree $d-1$ polynomial on each interval $u_i \leq u \leq u_{i+1}$ and $P(u)$ with the order of $1, \dots, d-2$ derivatives are continuously on the curve.

Beside that, as discussed in [17], when the degree is lower, the closer the curve follows the control polyline. In the work [18], the authors mentioned that B-spline is the useful tool in computer performance. Some previous work has been studied on the topic related to B-spline. In [19], the authors used B-spline to build several models of LV inside wall and the models were calculated by B-spline integral and the curves were fitted in the cardiac cycle. While in [20], the authors maintained the B-spline basis and retaining full approximation power of it at the same time. For [21], their algorithm is applied in the least-square sense by generating suitable control points of the fitting B-spline curve.

In general, the parameterization problem for B-spline curve is started with a set of 3D data points P_i , $i = 0, \dots, n$. Then, by finding a corresponding nondecreasing sequence of

Table 2: The data points sets used in this study, each data points set consists of 3D data points (x, y, z) .

i	x		y	z
Dataset 1				
1	0.58072		2.08688	0
2	3.50755		2.05734	0
3	5.36585		3.24051	0
4	7.64228		6.74273	0
5	9.40767		7.79716	0
6	11.1963		7.94949	0
7	13.9837		6.73456	0
8	15.08595		4.83752	0
9	16.7247		2.71041	0
10	19.2799		2.03701	0
Dataset 2				
1	1.83761		1.33271	0
2	0.726496		2.6504	0
3	4.82906		7.40482	0
4	5.12821		7.15185	0
5	8.4188		1.62069	0
6	8.84615		2.08095	0
7	9.44444		6.46124	0
Dataset 3				
1	0		9	0
2	0		5	0
3	0		3	0
4	0		1	0
5	1		0	0
6	3		0	0
7	5		0	0
8	9		0	0
Dataset 4				
1	32		14	0
2	34		10	0
3	36		9	0
4	38		12	0
5	38.2		11.7	0
6	42		14	0
7	44		12	0
8	46		13	0
9	48		7	0
10	50		6	0
11	52		10	0
12	54		10	0
13	56		10	0
14	58		10	0
Dataset 5				
1	1.40152		2.75472	0
2	4.84848		8.41509	0

Table 2: Continued.

<i>i</i>	<i>x</i>	<i>y</i>	<i>z</i>
3	4.84848	9.4	0
4	6.21212	7.0566	0
5	5	4.22642	0
6	9.01515	1.32075	0
Dataset 6			
1	-64.9885	-3.47E - 15	10
2	-64.9885	1.17851	10
3	-64.9885	4.71172	10
4	-64.9885	8.23099	10
5	-64.9885	11.7234	10
6	-64.9885	15.177	10
7	-64.9885	18.5811	10
8	-64.9885	21.9267	10
9	-64.8629	25.1835	10
10	-64.3195	27.1986	10
11	-63.7437	28.4339	10
12	-63.0155	29.5658	10
13	-62.1502	30.5815	10
14	-61.1627	31.4684	10
15	-60.0687	32.2155	10
16	-58.8839	32.8123	10
17	-57.6241	33.2485	10
18	-56.3068	33.5147	10
19	-54.9499	33.602	10
20	-51.8667	33.602	10
21	-48.783	33.602	10
22	-45.6989	33.602	10
23	-42.6143	33.602	10
24	-39.5292	33.602	10
25	-36.4438	33.602	10
26	-33.3579	33.602	10
27	-30.2717	33.602	10
28	-27.1852	33.602	10
29	-24.0983	33.602	10
30	-21.0111	33.602	10
31	-17.9236	33.602	10
32	-14.8359	33.602	10
33	-11.7479	33.602	10
34	-8.65971	33.602	10
35	-5.57133	33.602	10
36	-2.48278	33.602	10
37	0.605909	33.602	10
38	3.69473	33.602	10
39	6.78365	33.602	10
40	9.87265	33.602	10
41	12.9617	33.602	10
42	16.0508	33.602	10

Table 2: Continued.

i	x	y	z
43	19.1399	33.602	10
44	22.229	33.602	10
45	25.3181	33.602	10
46	28.4071	33.602	10
47	31.496	33.602	10
48	34.5849	33.602	10
49	37.6736	33.602	10
50	40.7622	33.602	10
51	43.8506	33.602	10
52	46.9388	33.602	10
53	50.0218	33.5576	10
54	52.0115	33.1629	10
55	53.2678	32.6832	10
56	54.4462	32.0405	10
57	55.5291	31.2451	10
58	56.4991	30.3076	10
59	57.3395	29.2401	10
60	58.0337	28.0564	10
61	58.5653	26.7706	10
62	58.9177	25.3996	10
63	59.074	23.9625	10
64	59.0805	20.619	10
65	59.0805	17.2107	10
66	59.0805	13.7467	10
67	59.0805	10.2369	10
68	59.0805	6.69301	10
69	59.0805	3.12756	10

value $u_0 < u_1 < \dots < u_n$, a function $x, y, z : [u_0, u_n] \rightarrow \mathbb{R}^3$, such that $\begin{bmatrix} x(u_i) \\ y(u_i) \\ z(u_i) \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}$, $i = 0, \dots, n$.

3. The Proposed Methodology

3.1. Experimental Design

The hybrid parameterization has better resulting curve in term of fairness and smoothness compared to uniform, chord length, and centripetal method [9]. This is because based on the comparison on the previous section, this method produces better accuracy on results compared to other methods and hence has been chosen to further explore this research. However, the hybrid parameterization method has a drawback. Consider (3.1):

$$u_0 = 0, \quad u_n = 1, \quad u_i = u_{i-1} + \frac{\sqrt{|B_i - B_{i-1}|}}{\sum_{j=1}^n \sqrt{|B_j - B_{j-1}|}}, \quad i = 1, 2, \dots, n-1. \quad (3.1)$$

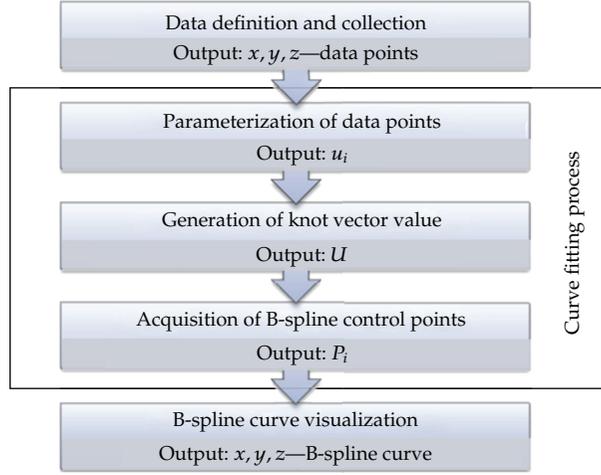


Figure 1: The Framework.

The value of the numerator $\sqrt{|B_i - B_{i-1}|}$ will close to zero as $B_i \approx B_{i-1}$, thus, $\sqrt{|B_i - B_{i-1}|} / \sum_{j=1}^n \sqrt{|B_j - B_{j-1}|} \approx 0$. As a result, the matrix of system of linear equation will be singular as $u_i \approx u_{i-1}$. To overcome this shortcoming, the authors propose a new parameterization method on B-spline curve presented in the following section. Figure 1 shows the framework design for this research.

3.2. New Proposed Parameterization Method

The proposed method is introduced to overcome the weakness of hybrid parameterization. Assume P_i are $n + 1$ number of given data points, and d is the degree of expected curve, thus, the determination of the parameter value is calculated as follows.

- (1) Let u_i be the initial parameter value, then find u_i by performing the exponential parameterization method with $\alpha = 0.8$. The value of α is determined based on experiment result.
- (2) Generate the knot vector, U , by using averaging knot vector method.
- (3) Compute the B-spline basis function, $N_{i,d}(u)$, and find the maximum value of $N_{i,d}(u)$, B_i on each span. After B_i are obtained, determine the new parameter value, θ_i , which is associated with B_i by solving the system of linear equations of B-spline basis function, $N'_{i,d}(u) = 0$, or by using numerical method such as Newton-Raphson method to search u value:

$$B_i = \max\{N_{i,d}(u)\}. \quad (3.2)$$

- (4) Calculate the median of both u_i and θ_i value and then compute the difference between them. Let s_i , $i = 1, \dots, n/2$, n is even, $i = 1, \dots, (n-1)/2$, n is odd be the difference between them, thus the new parameter value, u_i , is determined as follow:

```

INPUT:
  Pi: (x, y, z)-3D data points
  n: Number of data points
  d: Degree of the curve
OUTPUT:
  ui: Parameter value
  U: Knot vector
THE ALGORITHM:
  //Calculate initial parameter value
  ui → Exponential Parameterization (Pi, 0.8);
  U → Averaging Knot Vector (ui);
  For i = 1 ⋯ n
    Ni,d(u) → B spline Basis Function (u, U, n, d);
    Bi → max{Ni,d(u)};
    θi → get Param Val (Bi);
  end
  if (n mod 2 = 0)
    for i = 1 ⋯ n/2
      mθi → (θ2i-1 + θ2i)/2;
      mui → (u2i-1 + u2i)/2;
      si → mθi - mui;
      u2i-1 → u2i-1 + si;
      u2i → u2i + si;
    end
  else
    for i = 1 ⋯ (n - 1)/2
      mθi → (θ2i-1 + θ2i)/2;
      mui → (u2i-1 + u2i)/2;
      si → mθi - mui;
      u2i-1 → u2i-1 + si;
      u2i → u2i + si;
    end
  end
  u1 → 0.0;
  un → 1.0;
END

```

Algorithm 1

$$u_{2i-1} = u_{2i-1} + s_i, \quad \begin{cases} i = 1, \dots, \frac{n}{2}, & n \text{ is even,} \\ i = 1, \dots, \frac{(n-1)}{2}, & n \text{ is odd,} \end{cases} \quad (3.3)$$

$$u_{2i} = u_{2i} + s_i, \quad \begin{cases} i = 1, \dots, \frac{n}{2}, & n \text{ is even,} \\ i = 1, \dots, \frac{(n-1)}{2}, & n \text{ is odd.} \end{cases} \quad (3.4)$$

(5) Use u_i and U as final parameter value and knot vector.

By using above method, the resulting curve is expected to overcome the drawback of hybrid parameterization method. This is due to the methods proposed by previous

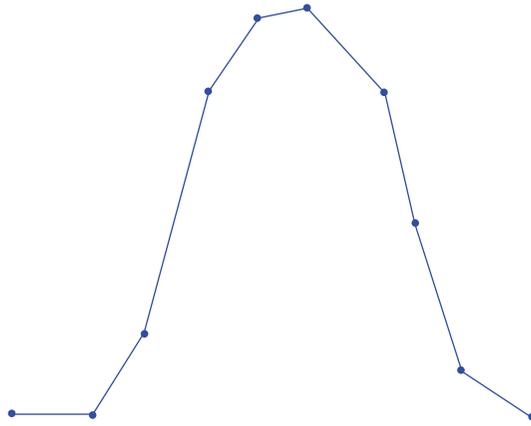


Figure 2: Dataset 1.

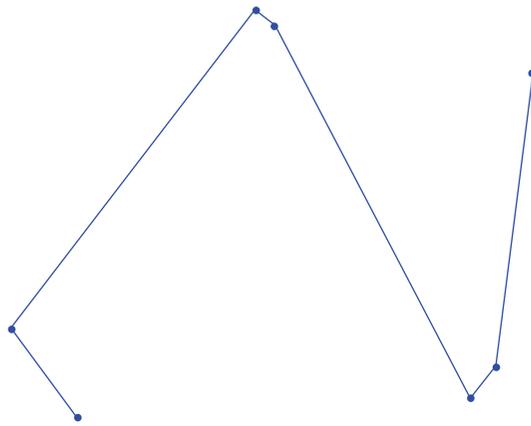


Figure 3: Dataset 2.

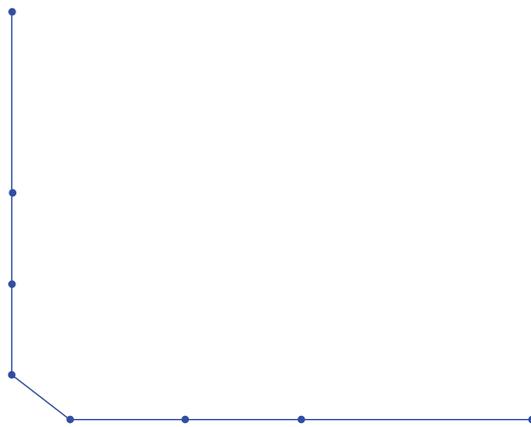


Figure 4: Dataset 3.

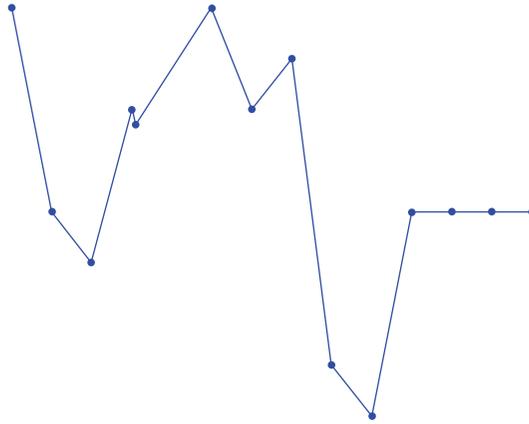


Figure 5: Dataset 4.

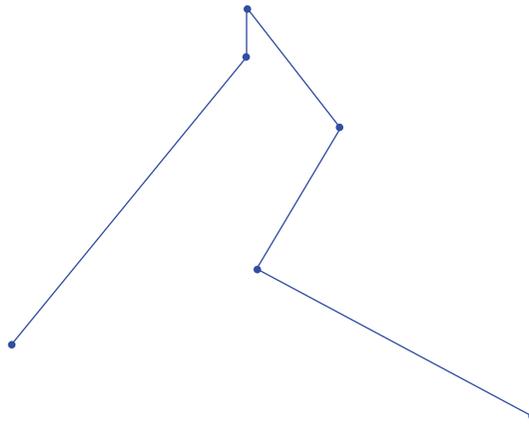


Figure 6: Dataset 5.

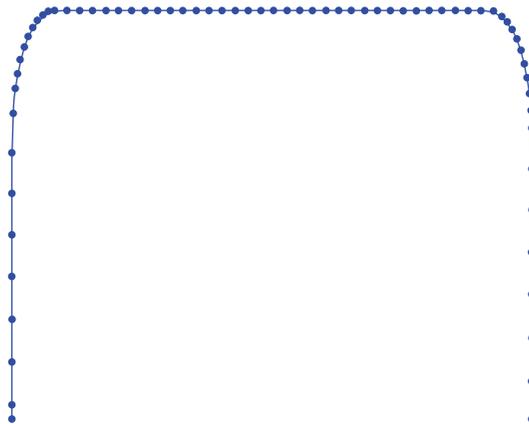


Figure 7: Dataset 6.

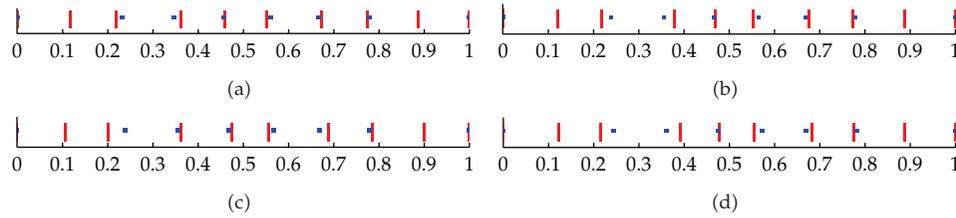


Figure 8: Parameter and knot distribution of (a) centripetal, (b) exponential, (c) proposed parameterization method, and (d) chord length.

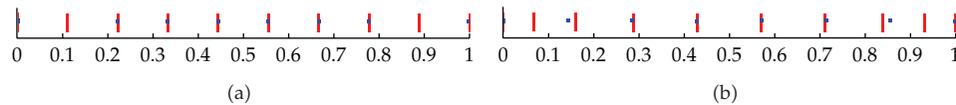


Figure 9: Parameter and knot distribution of (a) uniform and (b) universal parameterization method.

researchers sometimes may not produce a better result if compared to other methods. Maybe it can only produce a better result if using other datasets. Hence, this method is expected to handle all kinds of data and generate better curves. Moreover, the steps above were illustrated and presented in the algorithm form in the following section.

3.3. Implementation of the Proposed Parameterization Method

Algorithm 1 is the implementation of the proposed method in the pseudocode form. The 3D data point is an input while parameter value and knot vector are the output which can be used to build the basis value and construct the B-spline curve at the end.

4. Result and Discussion

This section presents experimental data, results, and analysis of the proposed method in comparison with the previous methods reported in literature. Both the proposed and previous methods are compared using the same data points.

4.1. Data Definition and Collection

Datasets which are used in this study are taken from the previous studies. There are six datasets. The datasets 1, 2, 3, and 5 are obtained from [22–25], respectively. In addition, dataset 6 which is one of the cross-sectional curve data of brush handle model obtained from [26] is also used for testing, while the dataset 4 is generated points data. The complete data points are listed in Table 2.

Each dataset has different characteristics. For instance, dataset 1 is simple data points. There are no collinear points and no two consecutive data points. The dataset 2 is one example of data points set which has two consecutive and large distances between two data points. Dataset 3 has collinear points at the start and the end of the data. The dataset 4 is the combination of the characteristic of previous datasets. That has two consecutive, collinear,

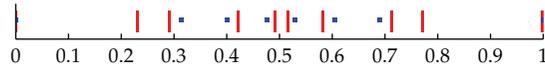


Figure 10: Parameter and knot distribution of old hybrid parameterization method.

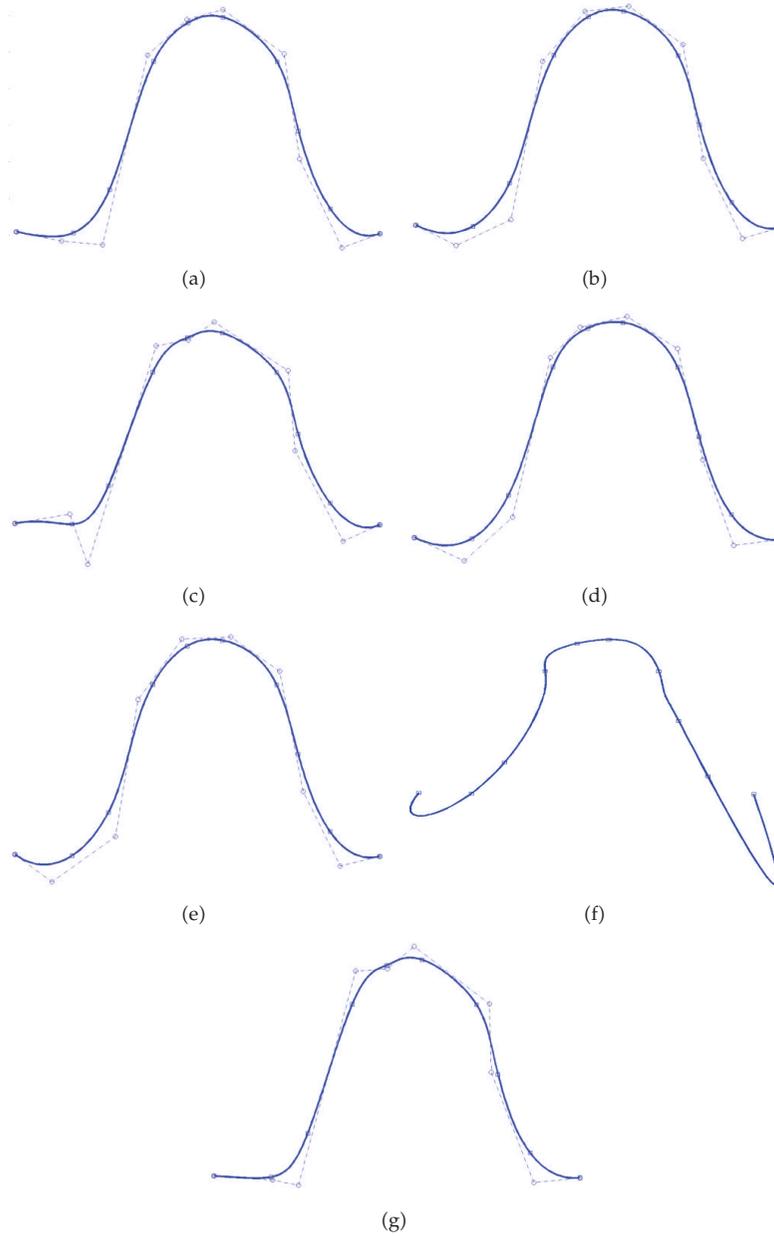


Figure 11: Curve interpolation for Dataset 1 using (a) centripetal, (b) exponential, (c) uniform, (d) proposed, (e) chord length, (f) hybrid, and (g) universal parameterization method.

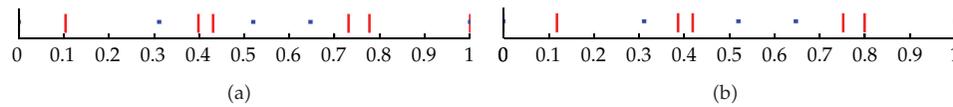


Figure 12: Parameter and knot distribution of (a) exponential and (b) proposed parameterization method.

and large distances between two data points. The dataset 5 has similarity to the dataset 2 but with collinear data points near the top. In addition, the dataset 6 has at most data points compared to the others (Figures 2, 3, 4, 5, 6, and 7).

4.1.1. Dataset 1

The dataset 1 is shown in Figure 8. It has 10 data points which are marked with a circle. Figures 11(a)–11(g) show the interpolation curve result of dataset 1. The data points are marked with a square while the control points are marked with a circle. The results show that centripetal, exponential, proposed, and chord length methods in Figures 11(a), 11(b), 11(d), and 11(e) have similar results. This is also strengthened by examining the distribution of the parameter value as depicted in Figures 8(a)–8(d). The blue dots represent a knot value while the red bar is parameter value. The figures show that centripetal, exponential, proposed method, and chord length have similar parameter value distribution.

Similar results are also shown by uniform and universal parameterization in Figures 11(c) and 11(g). However, it has inflection point at near the top of the curve and universal parameterization has wider distance parameter value compared to uniform parameterization as depicted in knot and parameter distribution shown in Figures 9(a) and 9(b).

The only different result is given by hybrid parameterization as illustrated in Figure 11(f). It has bulky curve at the start and end of data point due to a big distance at the start and end of the generated parameter value as presented in Figure 10.

4.1.2. Dataset 2

Figures 13(a)–13(g) are the resulting curve of dataset 2 which has been constructed by various parameterization methods. Figure 13(a) depicts the interpolation curve which is constructed by centripetal method while Figure 13(e) is constructed by using chord length method. As seen in Figure 13(e), the chord length method has bulky result at two neighbor points with longer distance compared to centripetal, exponential, and proposed method in Figures 13(a), 13(b), and 13(d). The hybrid method generates unpredicted result in Figure 13(f) since it cannot handle two adjacent points as also shown with the uniform and universal parameterization method which produces small loop as shown in Figures 13(c) and 13(g).

For this dataset, the proposed method has similar result with the exponential since the proposed method is an extension of exponential method as shown in Figures 13(d) and 13(b). This is also strengthened by the similarity of the parameter and knot distribution which is depicted in Figures 12(a) and 12(b).

4.1.3. Dataset 3

In dataset 3, the proposed method has better result as shown in Figure 16(d). It shows that the proposed method can handle collinear data. Similar results are also achieved by uniform and

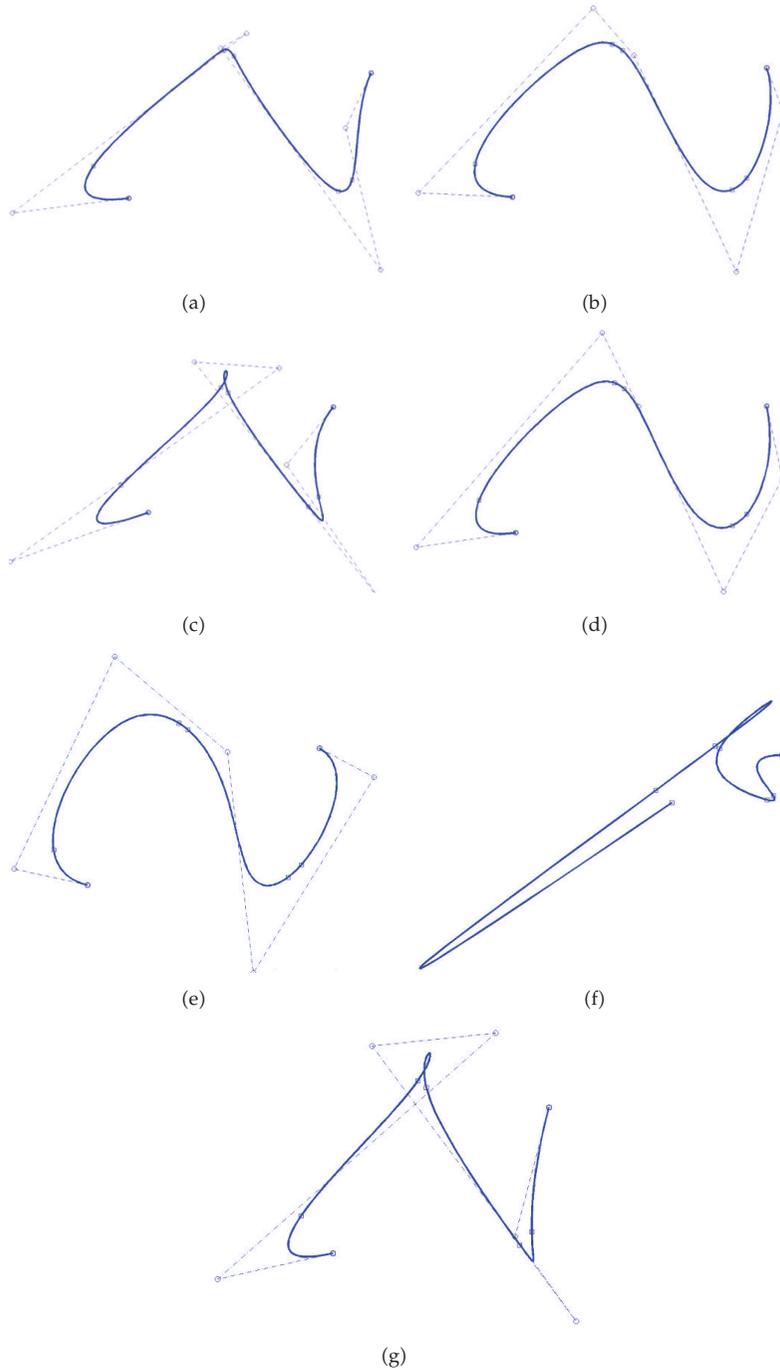


Figure 13: Curve interpolation for dataset 2 using (a) centripetal, (b) exponential, (c) uniform, (d) proposed, (e) chord length, (f) hybrid, and (g) universal parameterization method.

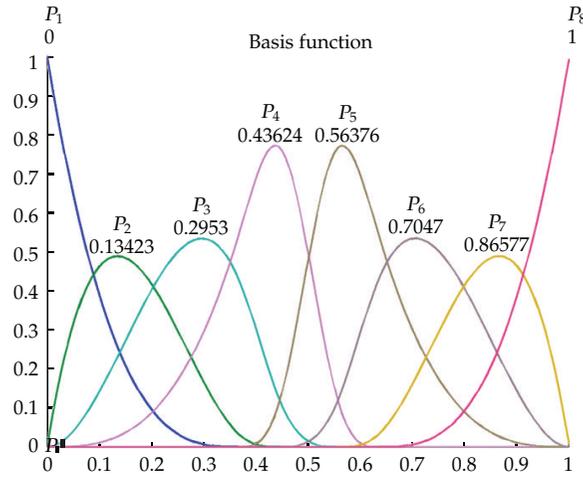


Figure 14: B-spline basis function of old hybrid parameterization method.

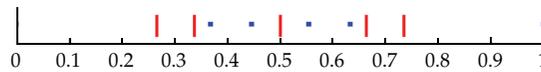


Figure 15: Knot and parameter value distribution of old hybrid parameterization method.

universal method in Figures 16(c) and 16(g) as well as centripetal and exponential method in Figures 16(a) and 16(b), respectively. In addition, Figure 16(e) shows that the chord length parameterization has bulky result compared to the others.

Figure 16(f) shows the unexpected result that is constructed by hybrid parameterization method. This may occur due to the similarity value of two consecutive basis values as shown in Figure 14. Since there is similarity of two basis function value, thus, this leads to the similarity of the parameter value ($u_3 \approx u_4 = 0.5$) which can cause the singularity of the matrix of linear equations (Figure 15).

4.1.4. Dataset 4

The interpolation results for dataset 4 are shown in Figures 17(a)–17(g). The centripetal and exponential methods, which are illustrated in Figures 17(a) and 17(b), have similar results of each other. It is shown at the beginning of data points and at the end of the data points. The interpolation result on collinear data points has few bulky result on the curve while chord length, which is shown in Figure 17(e), has twisty results at the beginning of the curve but performs better in handling collinear data points at the end. Moreover, the proposed method has similar result shown in Figure 17(d). These all are shown in Figures 17(a), 17(b), 17(d), and 17(e).

The hybrid result is depicted in Figure 17(f). It shows the hybrid parameterization method generating unwanted curve while interpolating two adjacent and collinear data points. Furthermore, the uniform and universal method produces small cusp in the two adjacent data points in Figures 17(c) and 17(g).

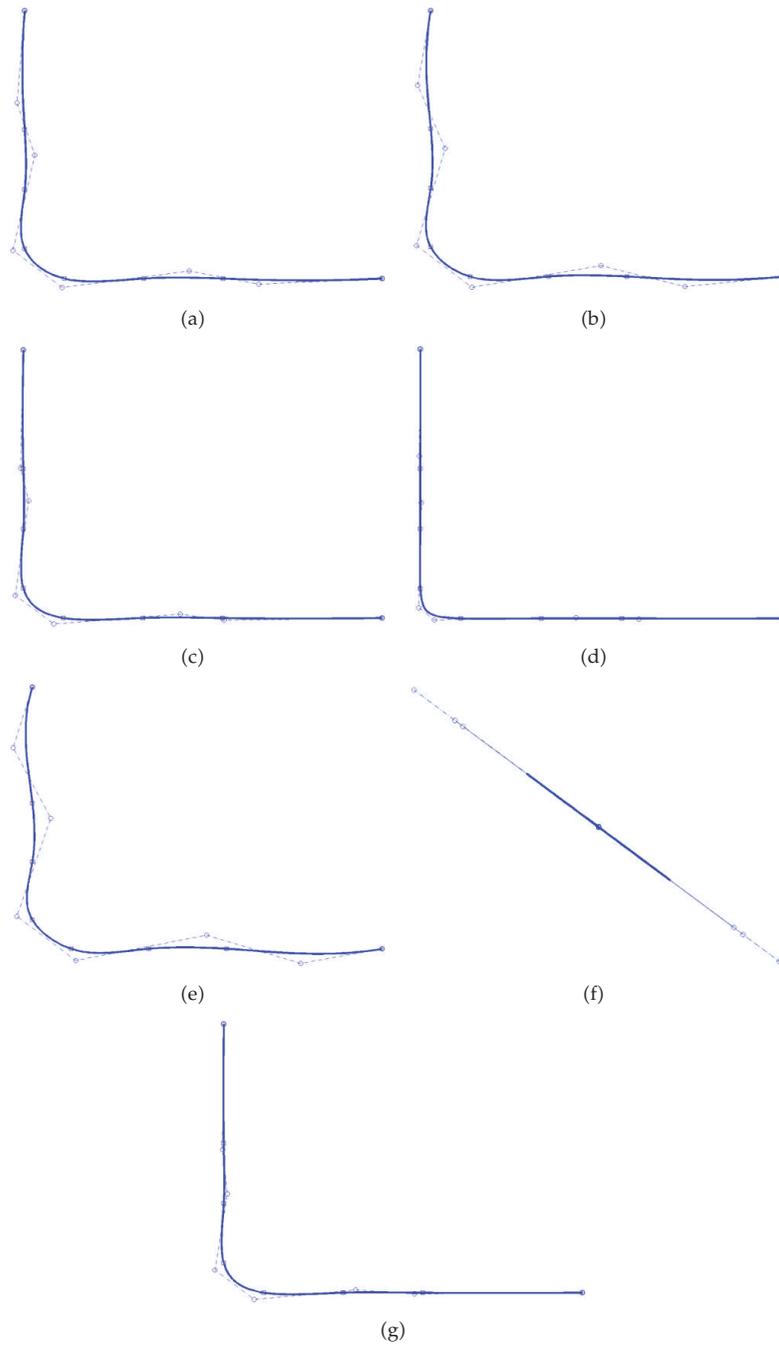


Figure 16: Curve interpolation for dataset 3 using (a) centripetal, (b) exponential, (c) uniform, (d) proposed, (e) chord length, (f) hybrid, and (g) universal parameterization method.

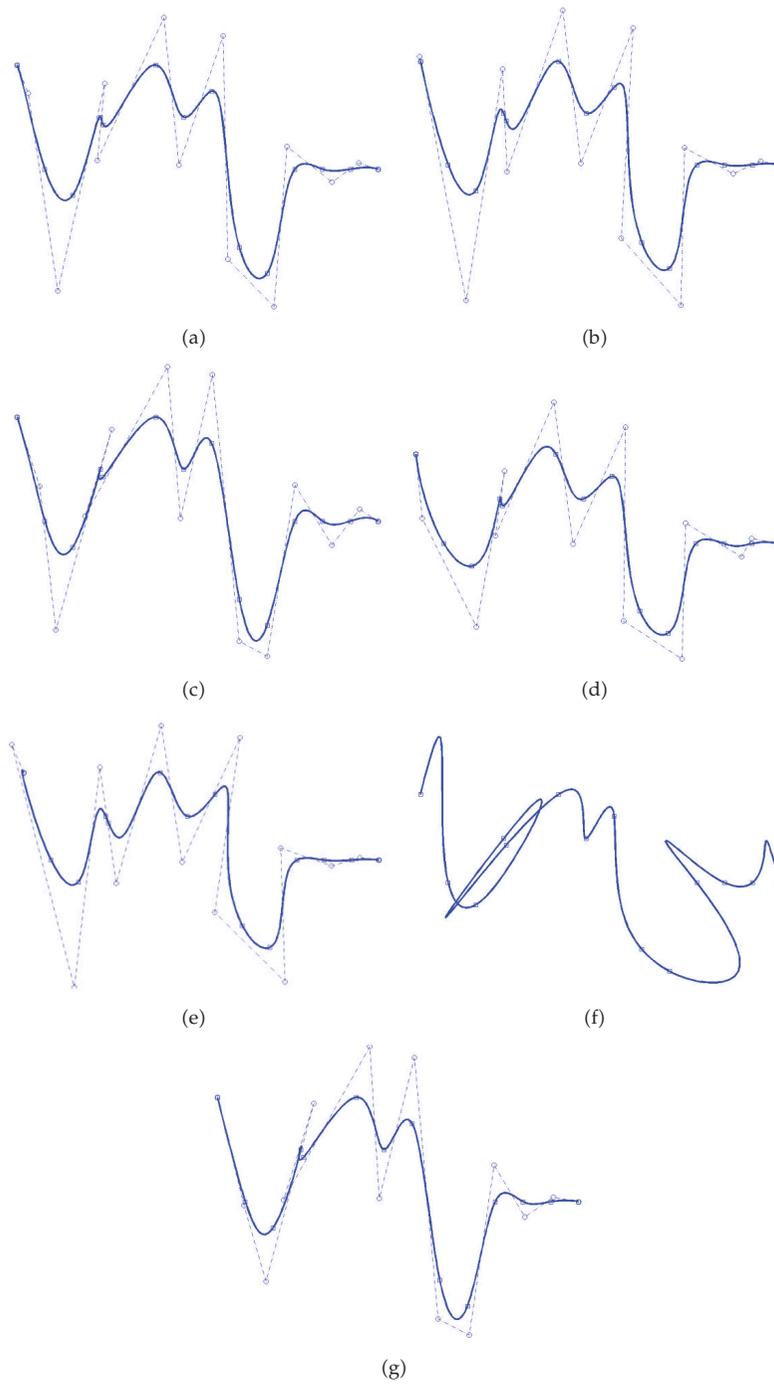


Figure 17: Curve interpolation for dataset 4 using (a) centripetal, (b) exponential, (c) uniform, (d) proposed, (e) chord length, (f) hybrid, and (g) universal parameterization method.

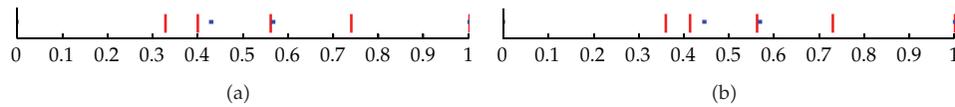


Figure 18: Parameter and knot distribution of (a) exponential and (b) chord length parameterization method.

4.1.5. Dataset 5

Figures 19(a)–19(g) present the interpolation results for dataset 5. The centripetal and proposed method has better result compared to the others. It is shown in Figures 19(a) and 19(d), respectively. The exponential and chord length methods have similar result, they produce bulky result for two consecutive data point with a long distance (Figures 19(b) and 19(e)). This is also strengthened by the similarity of the parameter and knot distribution which is shown in Figures 18(a) and 18(b).

A small loop is generated on the collinear data points by using uniform and universal method as illustrated in Figures 19(c) and 19(g). Moreover, the resulting curve constructed by using old hybrid parameterization is illustrated in Figure 19(f).

4.1.6. Dataset 6

Figures 21(a)–21(g) have depicted the interpolation curve of dataset 6. For this dataset, all the parameterization methods work well in constructing the curve, except the hybrid parameterization which has cusp near the corner as shown in Figures 20(a)–20(g). It also can be seen that all the parameterization methods have similarity of parameter and knot distribution which is illustrated in Figures 21(a)–21(f).

5. Conclusion and Future Work

The generation of the B-spline curve interpolation is started by parameterizing the data points followed by generating knot vector and finished by solving the system of linear equations. Each parameterization methods have different properties that cause varied shape of the curve. There are many parameterization methods that have been developed by previous researchers. In this study, the new parameterization method is developed.

The experiment on proposed method is conducted on various datasets which include collinear data points, two adjacent data points, and a long distance of consecutive data points. In conclusion, based on the analysis of the experiment results, in average, the proposed method has the better result compare to the others in Datasets 3 and 5. However, this method shows the best result in Dataset 3 where the curve roughly maps all points on the line if compared with the centripetal method. It shows that the proposed method works well in handling collinear data points, two adjacent data points both vertical and horizontal, and a long distance between two data points. The methods maybe not as good as others for certain experiments, but based on the experiments that have been carried out, it shows that this method can produce all good and better results compared to hybrid parameterization method. While for hybrid parameterization, it does not show that it is better compared to other methods in producing the results for this research. This can be shown by the results

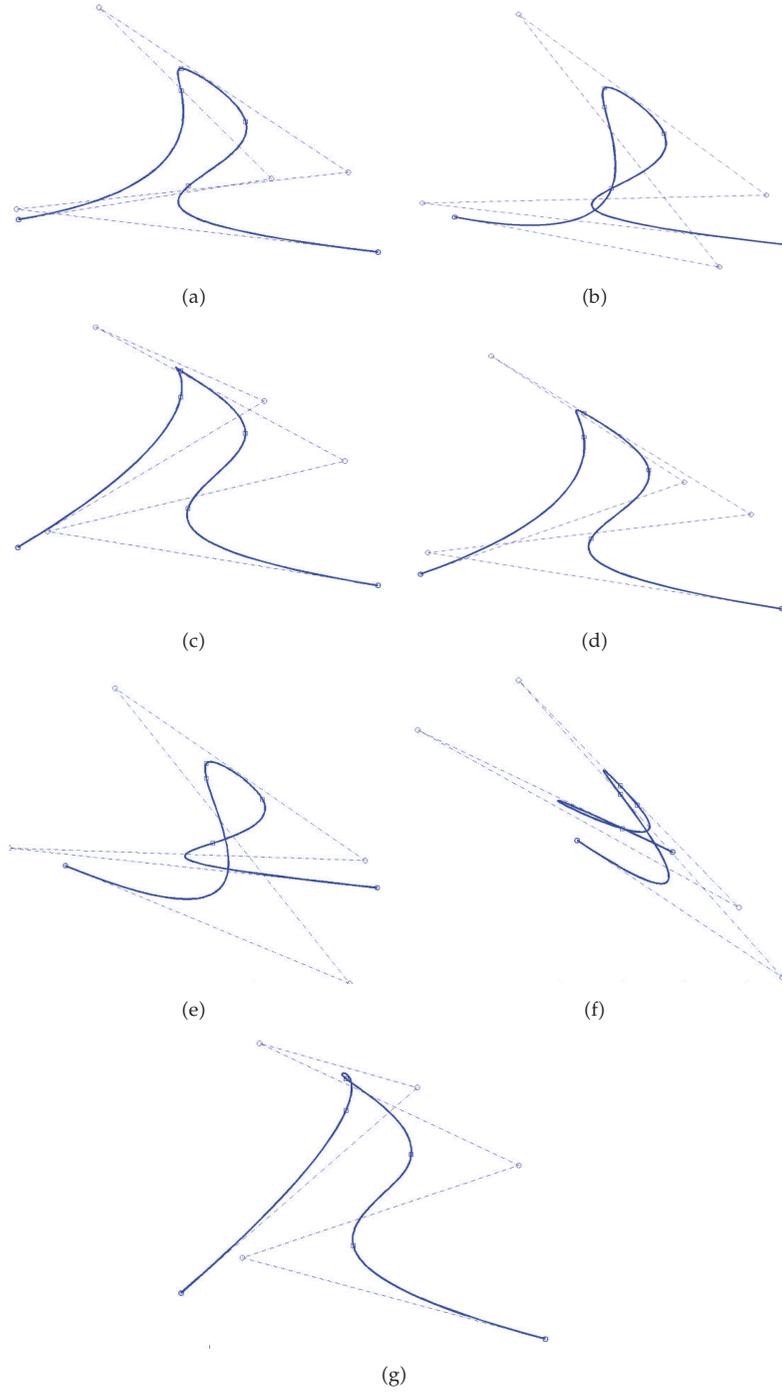


Figure 19: Curve interpolation for dataset 5 using (a) centripetal, (b) exponential, (c) uniform, (d) proposed, (e) chord length, (f) hybrid, and (g) universal parameterization method.

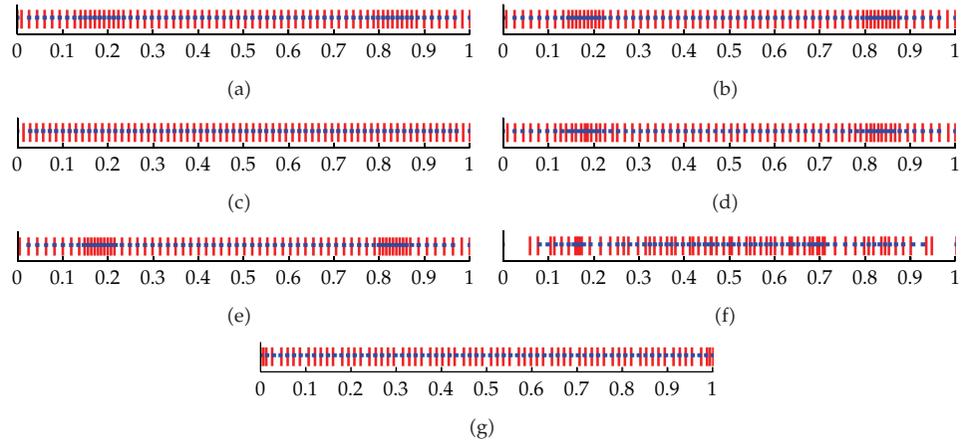


Figure 20: Parameter and knot distribution of (a) centripetal, (b) exponential, (c) uniform, (d) proposed, (e) chord length, (f) hybrid, and (g) universal parameterization method.

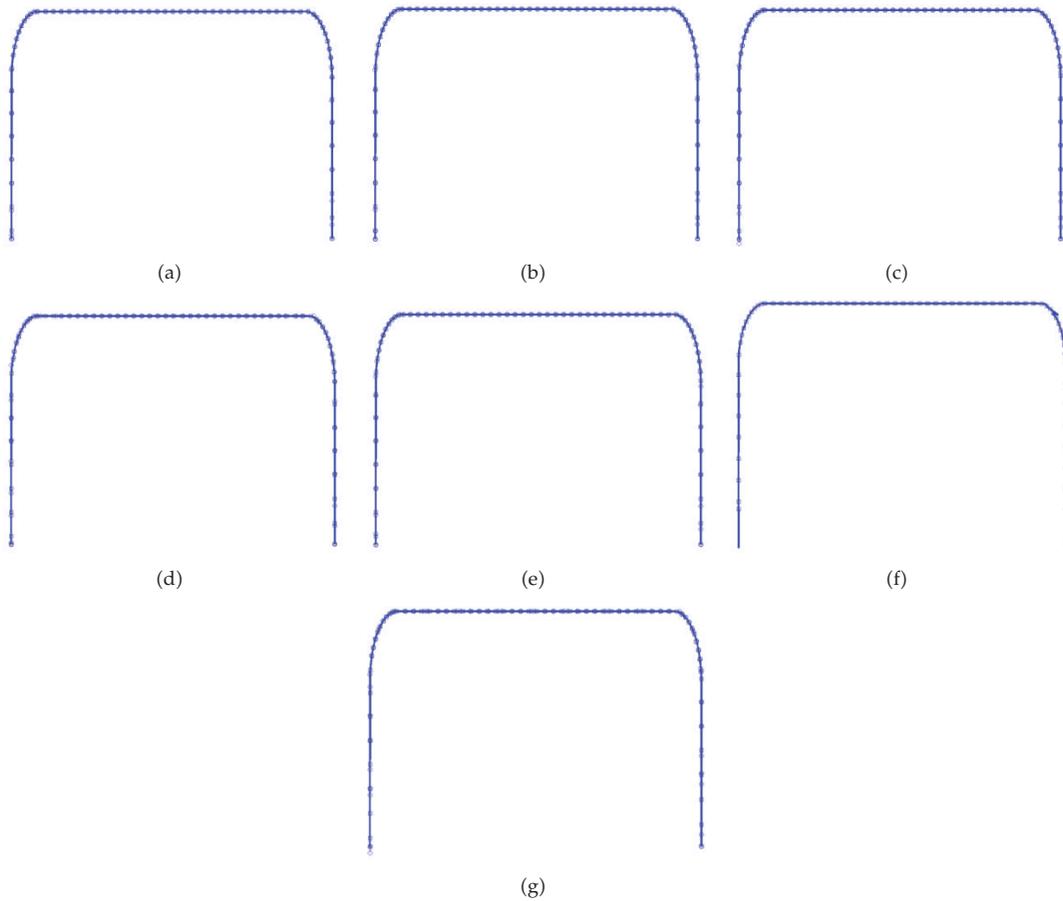


Figure 21: Curve interpolation for dataset 6 using (a) centripetal, (b) exponential, (c) uniform, (d) proposed, (e) chord length, (f) hybrid, and (g) universal parameterization method.

produced based on Datasets 3 and 5. Maybe it is due to the type of dataset. This method could be better in producing the results if using other dataset.

The main contribution of this research is the development of new parameterization method which can handle collinear, two adjacent, and a long distance of two consecutive data points. As shown in the experiment result, Figure 17(d), the proposed method has bulky result compared to the chord length method. Therefore, the reparameterization can be conducted to the proposed method to improve the result. The reparameterization process can be done through optimization method.

Appendix

See Table 2.

Acknowledgments

This study is a joint research between Universiti Teknologi Malaysia and Al-Imam Muhammad Bin Saud Islamic University. The authors honorably appreciate both parties for the support in making this project successful.

References

- [1] K. Hormann and G. Greiner, "MIPS: an efficient global parametrization method, curve and surface design," in *Innovations in Applied Mathematics*, P.-J. Laurent, P. Sablonnière, and L. L. Schumaker, Eds., pp. 153–162, Vanderbilt University Press, Nashville, Tenn, USA, 2000.
- [2] E. B. Kuznetsov and A. Y. Yakimovich, "The best parameterization for parametric interpolation," *Journal of Computational and Applied Mathematics*, vol. 191, no. 2, pp. 239–245, 2006.
- [3] S. M. Shamsuddin, M. A. Ahmed, and Y. Samian, "NURBS skinning surface for ship hull design based on new parameterization method," *International Journal of Advanced Manufacturing Technology*, vol. 28, no. 9, pp. 936–941, 2006.
- [4] S. Y. Chen, J. Zhang, Q. Guan, and S. Liu, "Detection and amendment of shape distortions based on moment invariants for active shape models," *IET Image Processing*, vol. 5, no. 3, pp. 273–285, 2011.
- [5] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 480–487, 2011.
- [6] C. de Boor, *A Practical Guide to Splines*, vol. 27, Springer, New York, NY, USA, 2001.
- [7] E. T. Y. Lee, "Choosing nodes in parametric curve interpolation," *Computer-Aided Design*, vol. 21, no. 6, pp. 363–370, 1989.
- [8] C.-G. Lim, "A universal parametrization in B-spline curve and surface interpolation," *Computer Aided Geometric Design*, vol. 16, no. 5, pp. 407–422, 1999.
- [9] S. M. H. Shamsuddin and M. A. Ahmed, "A hybrid parameterization method for NURBS," in *Proceedings of the International Conference on Computer Graphics, Imaging and Visualization (CGIV '04)*, pp. 15–20, July 2004.
- [10] W. Lü, "Curves with chord length parameterization," *Computer Aided Geometric Design*, vol. 26, no. 3, pp. 342–350, 2009.
- [11] J. Hoschek and D. Lasser, *Fundamentals of Computer Aided Geometric Design*, A. K. Peters, Wellesley, Mass, USA, 1993.
- [12] H. Park, "Choosing nodes and knots in closed B-spline curve interpolation to point data," *Computer Aided Design*, vol. 33, no. 13, pp. 967–974, 2001.
- [13] S. S. Yuhaniz, Lecture 07: B-Spline and NURBS, Computer Aided Design and Manufacturing, Power Point Slides, 2010.
- [14] S. S. Yuhaniz, Lecture 6: Geometric Modelling—Curves, Computer Aided Design and Manufacturing, Power Point Slides, 2010.

- [15] D. F. Rogers, "An Introduction to NURBS with Historical Perspective," Morgan Kaufmann Publishers, 2001.
- [16] H. Zhang and J. Feng, B-Spline Curves and Surfaces (2), Power Point Slides, 2006.
- [17] Spline, Bezier, B-Spline.
- [18] G. S. Kumar, P. K. Kalra, and S. G. Dhande, "Parameter optimization for B-spline curve fitting using genetic algorithms," in *Proceedings of the Congress on Evolutionary Computation (CEC '03)*, 2003.
- [19] S. Chen, J. Zhang, H. Zhang et al., "Myocardial motion analysis for determination of tei-index of human heart," *Sensors*, vol. 10, no. 12, pp. 11428–11439, 2010.
- [20] K. Höllig, U. Reif, and J. Wipper, B-Spline Approximation Of Neumann Problems, Mathematics Subject Classification, 65N30, 65N12, 41A63, 41A15, 1991.
- [21] Y. Liu, H. Yang, and W. Wang, "Reconstructing B-spline curves from point clouds—a tangential flow approach using least squares minimization," in *Proceedings of the International Conference on Shape Modeling and Applications (SMI '05)*, pp. 4–12, June 2005.
- [22] L. A. Piegl and W. Tiller, "Surface approximation to scanned data," *Visual Computer*, vol. 16, no. 7, pp. 386–395, 2000.
- [23] C. Yuksel, S. Schaefer, and J. Keyser, "Parameterization and applications of Catmull-Rom curves," *Computer Aided Design*, vol. 43, no. 7, pp. 747–755, 2011.
- [24] M.S. Floater and E. Zagar, "On general approach to optimal parameterization of polynomial curves," in *Proceedings of the Fusion Simulation Project Workshop*, Styria, Austria, November 2006.
- [25] C.-K. Shene, CS3621 Introduction to Computing with Geometry Note, 2003, <http://www.cs.mtu.edu/~shene/COURSES/cs3621/NOTES/notes.html>.
- [26] W.-K. Wang, H. Zhang, H. Park, J.-H. Yong, J.-C. Paul, and J.-G. Sun, "Reducing control points in lofted B-spline surface interpolation using common knot vector determination," *Computer Aided Design*, vol. 40, no. 10-11, pp. 999–1008, 2008.

Research Article

Entropy and Multifractality for the Myeloma Multiple TET 2 Gene

Carlo Cattani,¹ Gaetano Pierro,² and Giuseppe Altieri³

¹ *DipMat, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

² *System Biology, PhD School, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

³ *SSFO, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

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

Received 27 June 2011; Accepted 8 September 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Carlo Cattani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The nucleotide and amino-acid distributions are studied for two variants of mRNA of gene that codes for a protein which is involved in multiple myeloid. Some patches and symmetries are singled out, thus, showing some distinctions between the two variants. Fractal dimensions and entropy are discussed as well.

1. Introduction

In some recent papers, the concepts of fractality [1–19] and entropy [19–21] have been considered as fundamental parameters to investigate the existence of correlations [22–36] and simple rules [37] in the DNA sequences. In particular, it has been observed that the increasing fractal dimension [7–13] can be related to a degeneration in sequences, having as a consequence pathological evolution of related diseases. A fundamental role is played by the concept of information entropy [20, 21] so that a change in the nucleotide distribution in DNA implies a corresponding change in the information content and, as a consequence, a variation in the entropy. Since the cell activity is functionally dependent on the nucleotide distribution our task is to understand better about this distribution and/or about the existence of large scale structure [1–6, 15, 22–37]. So that we could relate the functional activity of cells to some epitomizing patches in the nucleotide distribution. We will propose, in the following, also to take into account the information content in the amino acid distribution. In particular, we will see that the amino-acid distribution shows a higher level structure, and some patchiness which are undetectable in the nucleotide distribution. Our statistical approach is based on the transformation of the symbolic string into a numerical string by the Voss indicator function

[4, 5] which is a discrete binary function. On this function, the indicator matrix is defined and on this matrix the fractal dimension and entropy can be simply computed. We will compare the fractal dimension and complexity of two mRNA variants of TET 2 (ten-eleven translocation 2) gene downloaded from gene bank [38] (similar data are also available from [39–41]), by showing that these parameters can be used to classify the two variants. Multiple myeloma is a pathology which involves plasma cell, but it can move and spread into the whole body. Some aspects are still unclear; however, it is known that this pathology is characterized by the activation of abnormal genes through chromosomal translocations and other genetic anomalies. One of the genes involved in the birth and progression of multiple myeloma is the TET 2. In fact it is present in some myelodysplastic syndromes, and it seems to play a key role when it is subject to mutation. The gene TET 2 is related with myelopoiesis; in fact it encodes protein that we can find significantly expressed in hematopoietic cells and granulocytes.

2. Multiple Myeloma and the Oncogene TET 2

Multiple myeloma (MM) is a blood cancer of the plasma cell. Myeloma originates in a specific type of cell, the plasma cell, but it can move, so that it spreads by the blood to the whole body. Like other cancers, multiple myeloma will develop in steps. Myeloma begins when the normal plasma cell becomes abnormal. The abnormal cell divides, and the new cells divide again and again, thus proliferating the number of abnormal cells. Myeloma cells collect in the bone marrow and in the solid part of the bone. These malignant plasma cells produce a para protein, an inactive antibody known also as M-protein or Bence Jones protein, that attack bone marrow, bones, blood, and kidneys. As a consequence, there happens extensive destruction within the skeleton involving multiple bones, and resulting in widespread bone pain and multiple fractures; for this reason, such a disease is also called multiple myeloma. Some genetic factors are also involved in this pathology. In absence of other symptoms and clinical signs, this condition is more properly called benign monoclonal gammopathy of uncertain significance (MGUS). In fact, the uncertainty about the future progression, it shows that also benign diseases might evolve into MM. It is likely that the evolution of MGUS into MM depends on many mutations of the MGUS clone. Initially, MM has a low progression, but afterward it becomes more aggressive. The signs that characterize onset of multiple myeloma are mostly high concentration of calcium ions with damages in the kidneys, the weakening of the immune system with abnormal production of immune globulin, and some other signs such as an evident osteoporosis. Both MGUS and MM diseases are characterized by the presence of alterations in gene expression [42–46]. The chromosomes that are more involved are 1,11,13,14, respectively. The alteration at chromosome 1 is found in half of cases of MM patients [47–54]. The same aberrations chromosome seem to be evident both in MM and in MGUS, thus supporting the thesis that these two diseases are closely related [53].

Gene TET 2 is located on the chromosome 4 exactly in 4q24. More precisely, the TET2 gene is located from base pair 106,067,942 to base pair 106,200,957 on chromosome 4, position as shown in Figure 1 [38].

The gene TET 2 plays a key role in the conversion of methylcytosine (5mC) to 5-hydroxymethylcytosine (hmC) moreover is related to myelopoiesis. For the hmC many roles were noted like for example (1) remodeling of chromatin structure (2) recruitment of some factors (3) demethylation of cytosine [55, 56]. The gene TET2 encodes a protein that we find significantly expressed in hematopoietic cells and granulocytes. In almost all patients with

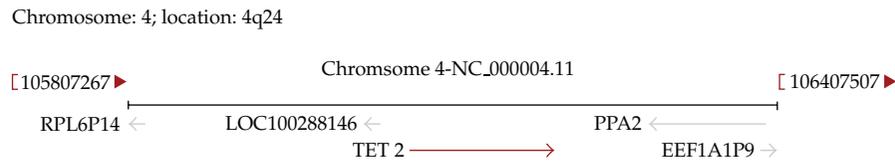


Figure 1: Location of the TET 2 on chromosome 4.

myelodysplastic syndromes, the protein is decreased in peripheral blood granulocytes. TET 2 gene is usually mutated in myeloproliferative disorders (MPDs). The MPD is part of a larger group of disorders called myeloproliferative neoplasms (MPNs). The mutation of TET 2 characterizes some disorder known as systemic mast cell disease, but TET 2 is above all mutated in myelodysplastic syndromes [57].

We will see that, by using some parameters defined on the indicator function, we can single out some patches which characterize abnormal functional activity [1–3, 35, 58, 59].

3. DNA Representation

The DNA, as well as the mRNA, of each organism of a given species is a sequence of a specific number of base pairs defined on the 4 elements alphabet of nucleotides:

$$A = \text{adenine}, C = \text{cytosine}, G = \text{guanine}, T = \text{thymine}. \quad (3.1)$$

Since the base pairs are distributed along a double helix, when straightened, the helix appears as a complementary double-strand system. The two sequences on opposite strands are complementary in the sense that opposite nucleotides must fulfil the ligand rules (A with T and G with C) of base pairs, between purines A and G and pyrimidines T and C. In a DNA sequence, there are some subsequences, which can be roughly subdivided into coding and noncoding regions, having special meaning. In particular, genes (belonging to coding regions) are characteristic sequences of base pairs, and the genes in turn are made by some alternating subsequences of exons and introns (except Procaryotes where the introns are missing). Each exon region is made of triplets of adjacent bases called codons. There are 64 possible codons, inasmuch the number of combination of the 4 nucleotides into 3 length classes. There are only 20 amino acids, therefore, the correspondence codons to amino acids are many to one. The 20 elements alphabet of amino acids is in Table 1. In the following, we will analyze two mRNA sequences: (H1) and (H2), downloaded from the National Center for Biotechnology Information [38], which represent respectively.

(H1) homo sapiens tet oncogene, family member 2 (TET2), transcript variant 1, mRNA, locus NM_001127208 (9796 bp mRNA linear). The accession number is NM_001127208, version NM_001127208.2 GI: 325197189.

(H2) homo sapiens tet oncogene family member 2 (TET2), transcript variant 2, mRNA, locus NM.017628 (9236 bp mRNA linear). The accession number is NM.017628, version NM.017628.4 GI: 325197183.

Some differences between two variants are the following: (H2) is different from (H1) in 5'UTR (untranslate region) and in 3'UTR (untranslate region); furthermore, (H2) variant,

Table 1

	Letter	Name	Codons
1	M	Methionine	ATG
2	E	Glutamic acid	GAA, GAG
3	Q	Glutamine	CAA, CAG
4	D	Aspartic acid	GAT, GAC
5	R	Arginine	CGT, CGC, CGA, CGG, AGA, AGG
6	T	Threonine	ACT, ACC, ACA, ACG
7	N	Asparagines	AAT, AAC
8	H	Histidine	CAT, CAC
9	V	Valine	GTT, GTC, GTA, GTG
10	G	Glycine	GGT, GGC, GGA, GGG
11	L	Leucine	TTA, TTG, CTT, CTC, CTA, CTG
12	S	Serine	TCT, TCC, TCA, TCG, AGT, AGC
13	P	Proline	CCT, CCC, CCA, CCG
14	F	Phenylalanine	TTT, TTC
15	I	Isoleucine	ATT, ATC, ATA
16	C	Cysteine	TGT, TGC
17	A	Alanine	GCT, GCC, GCA, GCG
18	K	Lysine	AAA, AAG
19	Y	Thyroxine	TAT, TAC
20	W	Tryptophan	TGG
		Stop	TAA, TAG, TGA

compared with (H1), is shown to have c-terminal to be distinct and even shorter of (H1), which is also represented by a longer transcript [37].

4. Dot Plot on the Indicator Matrix

In this section, we will define the indicator matrix [4, 5] on which the computation of multifractality and entropy are based.

4.1. Indicator Function for the 4-Symbol Alphabet

Let

$$\mathfrak{K}_4 = \{A, C, G, T\} \quad (4.1)$$

be the finite set (alphabet) of nucleotides and $x \in \mathfrak{K}_4$ any member of the 4 symbols alphabet.

A DNA sequence is the finite symbolic sequence

$$\mathfrak{D}(N) = \mathbb{N} \times \mathfrak{K}_4 \quad (4.2)$$

so that

$$\mathfrak{D}(N) = \{x_h\}_{h=1,\dots,N}, \quad N < \infty \tag{4.3}$$

being

$$x_h = (h, x) = x(h), \quad (h = 1, 2, \dots, N; x \in \mathfrak{K}_4) \tag{4.4}$$

the acid nucleic x at the position h .

Let $\mathfrak{D}_1(N), \mathfrak{D}_2(N)$ be two DNA sequences; the indicator function [4, 5] is the map

$$u : \mathfrak{D}_1(N) \times \mathfrak{D}_2(N) \longrightarrow \{0, 1\} \tag{4.5}$$

such that

$$u_{hk} = u(x_h, x_k) = \begin{cases} 1 & \text{if } x_h = x_k \\ 0 & \text{if } x_h \neq x_k \end{cases} \quad (x_h \in \mathfrak{D}_1(N), x_k \in \mathfrak{D}_2(N)). \tag{4.6}$$

When $\mathfrak{D}_1(N) \equiv \mathfrak{D}_2(N)$ the indicator function, it shows the existence of autocorrelation on the same sequence. According to (4.5), the indicator map of the N -length sequence can be easily represented by the $N \times N$ sparse matrix of binary values $\{0, 1\}$, and this matrix can be visualized by the following (autocorrelation) dot-plot:

\vdots												
G	0	1	0	0	0	0	0	0	0	1	...	
C	0	0	0	1	0	0	0	0	0	1	0	...
A	1	0	0	0	1	0	1	1	0	0	...	
A	1	0	0	0	1	0	1	1	0	0	...	
T	0	0	1	0	0	1	0	0	0	0	...	
A	1	0	0	0	1	0	0	1	0	0	...	
C	0	0	0	1	0	0	0	0	1	0	...	
T	0	0	1	0	0	1	0	0	0	0	...	
G	0	1	0	0	0	0	0	0	0	1	...	
A	1	0	0	0	1	0	0	1	0	0	...	
u_{hk}	A	G	T	C	A	T	A	A	C	G	...	

4.2. Indicator Function for the 20-Symbols Alphabet of Amino Acids

As a generalization of the 4-symbols alphabet of nucleotides, we can define the 20-symbols alphabet of amino acids as follows:

$$\mathfrak{K}_{20} = \{M, E, Q, D, R, T, N, H, V, G, L, S, P, F, I, C, A, K, Y, W\}. \tag{4.8}$$

A protein sequence is the finite symbolic sequence

$$\mathcal{L}(N) = \mathbb{N} \times \mathfrak{A}_{20} \quad (4.9)$$

so that

$$\mathcal{L}(N) = \{x_h\}_{h=1,\dots,N}, \quad N < \infty \quad (4.10)$$

being

$$x_h = (h, x) = x(h), \quad (h = 1, 2, \dots, N; x \in \mathfrak{A}_{20}) \quad (4.11)$$

the amino acid x at the position h .

The indicator function [4, 5] can be extended also to protein sequences as the map

$$u : \mathcal{L}_1(N) \times \mathcal{L}_2(N) \longrightarrow \{0, 1\} \quad (4.12)$$

such that

$$u_{hk} = u(x_h, x_k) = \begin{cases} 1 & \text{if } x_h = x_k \\ 0 & \text{if } x_h \neq x_k \end{cases} \quad (x_h \in \mathcal{L}_1(N), x_k \in \mathcal{L}_2(N)). \quad (4.13)$$

After a transduction of the two DNA sequences (H1) and (H2) into their amino acids components, we can see that the corresponding dot plots can show some (higher-level) structure on the distribution of nucleotides (see Figure 2). In particular, (H2) shows a special pattern which is more evident in the amino acids dot plot.

5. Probability Distribution

5.1. Frequency Distribution

The probability distribution of nucleotides can be defined by the frequency

$$p_X(n) = \frac{1}{n} \sum_{i=1}^n u_{Xi}, \quad (X \in \mathfrak{A}_4, x_i \in \mathcal{D}(N); 1 \leq n \leq N) \quad (5.1)$$

that the acid nucleic X can be found at the position n . This value can be approximated by the frequency count (on the indicator matrix) of the nucleotide distribution before n . So that, for the transcript variant, we have the probability density distribution of Figure 3 which, however, tends to assume some different constant values thus showing that nucleotides are heterogeneously distributed.

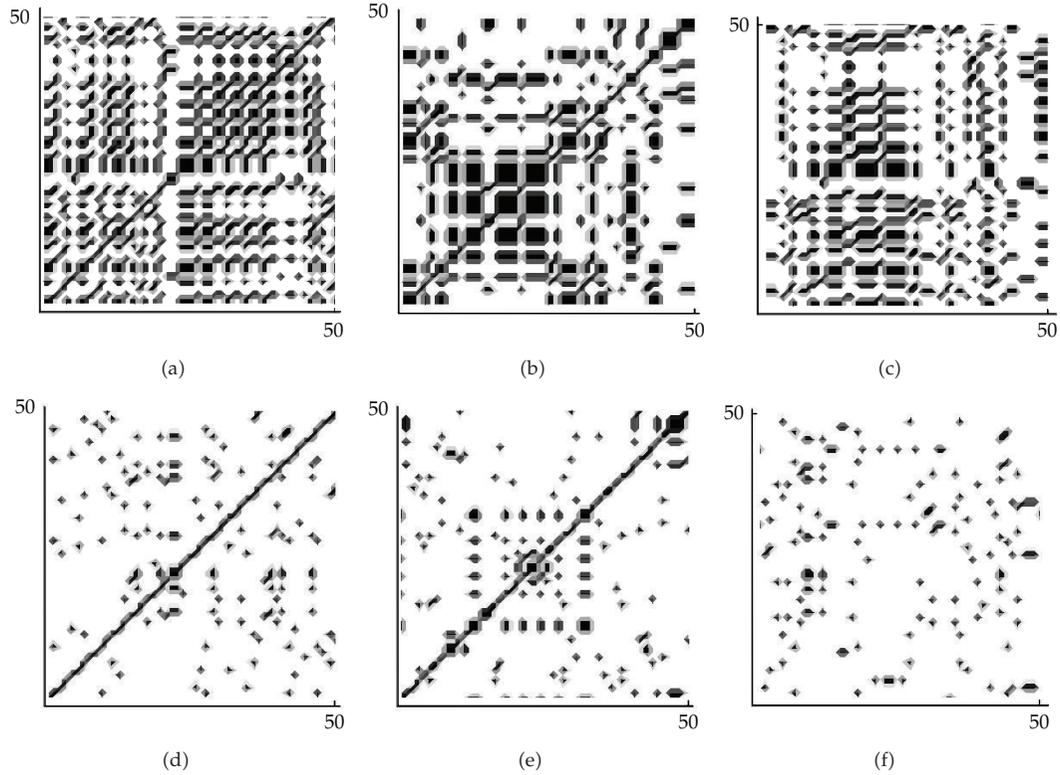


Figure 2: Dot plot for the first 50 nucleotides distribution in the (H1)-(H1), (H2)-(H2) DNA sequences (a,b) and corresponding amino acids (d) and (e). In (c) and (f), the cross correlations (H1)-(H2) are given.

5.2. Distribution of the Essential Amino Acids

Analogously to the nucleotides frequency distribution, we can compute also the amino-acid distribution

$$p_X(n) = \frac{1}{n} \sum_{i=1}^n u_{Xi}, \quad (X \in \mathfrak{S}_{20}, x_i \in \mathcal{L}(N); 1 \leq n \leq N) \quad (5.2)$$

that the amino-acid X can be found at the position n .

In particular, we have noticed that even if the nucleotides distribution is nearly the same in both sequences (H1) and (H2), the amino acid shows different distributions for the same amino-acid in each sequence. In other words, the “second”-level distribution seems to be organized according to a different distribution law (see Figures 4 and 5).

6. Fractal Dimension and Entropy

6.1. Fractal Dimension

The frequency distribution implies a corresponding frequency of correlation in the correlation matrix. By using the indicator matrix, it is possible to give a simple formula which enables

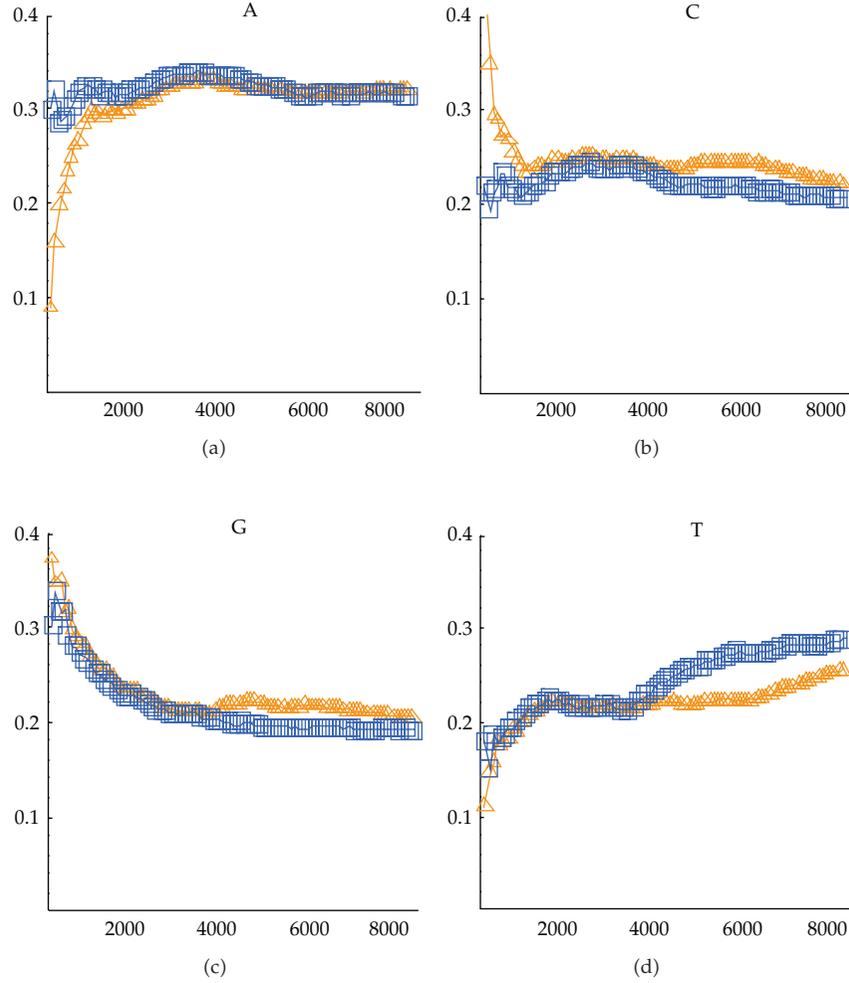


Figure 3: Probability density distribution of nucleotides along TET2 oncogene variants (H1) (brown) and (H2) (blue).

us to estimate the fractal dimension as the average of the number $p(n)$ of 1 in the randomly taken $n \times n$ minors of the $N \times N$ correlation matrix u_{hk}

$$D = \frac{1}{2} \frac{1}{N} \sum_{n=2}^N \frac{\log p(n)}{\log n}. \quad (6.1)$$

If we compare the fractal dimensions of the two mRNA sequences (H1) and (H2), we can see (Figure 6) that the fractal dimension of nucleotide distribution tends, for both variants, to the value 1.26.

It is interesting to notice that the corresponding amino acids of the two sequences have (more or less) the same fractal dimension which tends for both (Figure 7) to the value 1.29.

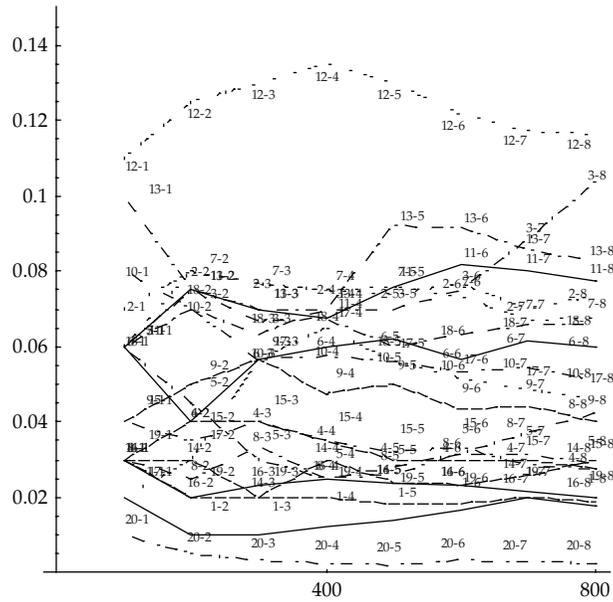


Figure 4: Distribution of amino acids along (H1), being 1 methionine, 2 glutamic acid, 3 glutamine, 4 aspartic acid, 5 arginine, 6 threonine, 7 asparagine, 8 histidine, 9 valine, 10 glycine, 11 leucine, 12 serine, 13 proline, 14 phenylalanine, 15 isoleucine, 16 cysteine, 17 alanine, 18 lysine, 19 thyroxine, and 20 tryptophan.

6.2. Entropy Estimate

As a measure of the information distribution, we consider the normalized Shannon entropy, which is defined, for a distribution over the alphabet \mathfrak{A}_ℓ , as

$$H(n) = -\frac{1}{\log \ell} \sum_{i=1}^{\ell} p_{X_i}(n) \times \begin{cases} \log_2 p_{X_i}(n) & \text{if } p_{X_i}(n) \neq 0, \\ 0 & \text{if } p_{X_i}(n) = 0, \end{cases} \quad (6.2)$$

where $p_{X_i}(n)$ is given by (5.1) for nucleotides and (5.2) for amino acids.

Since $\sum_{i=1}^{\ell} p_{X_i}(n) = 1$, for all n , the main values of this function are the following.

- (1) If $p_{X_i}(n) = 1, p_{X_j}(n) = 0 (j \neq i)$, then $H(n) = 0$. This happens when the information is concentrated in only one symbol.
- (2) If $p_{X_i}(n) = p_{X_j}(n) = 1/\ell, i \neq j$ then $H(n) = 1$. In this case, the information is equally distributed over all symbols.
- (3) Equation $0 \leq H(n) \leq 1$. In general, the information content is distributed over the range $[0, 1]$.

Therefore, the entropy is a positive function ranging in the interval $[0, 1]$, the minimum value is obtained when the distribution is concentrated on a single symbol, while the maximum value is obtained when all symbols are equally distributed.

In particular for higher values of n , according to the frequency definition of probability, the entropy tends to the constant value 1 (see Figures 8 and 9) both for nucleotides and amino

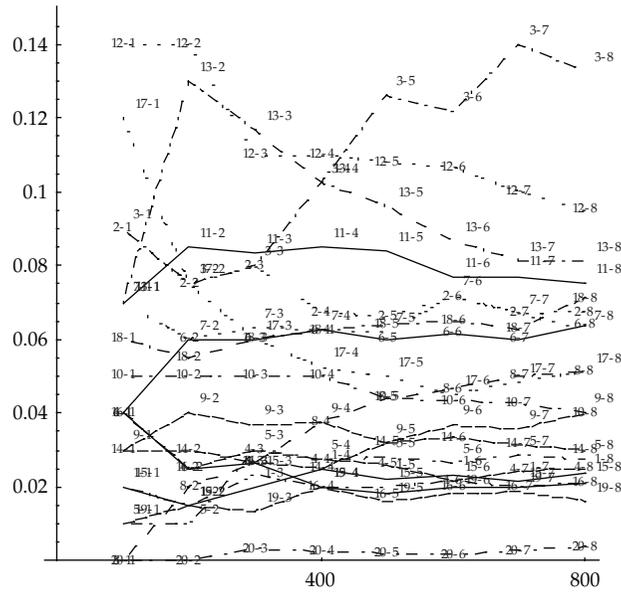


Figure 5: Distribution of amino acids along (H2), being 1 methionine, 2 glutamic acid, 3 glutamine, 4 aspartic acid, 5 arginine, 6 threonine, 7 asparagine, 8 histidine, 9 valine, 10 glycine, 11 leucine, 12 serine, 13 proline, 14 phenylalanine, 15 isoleucine, 16 cysteine, 17 alanine, 18 lysine, 19 thyroxine, and 20 tryptophan.

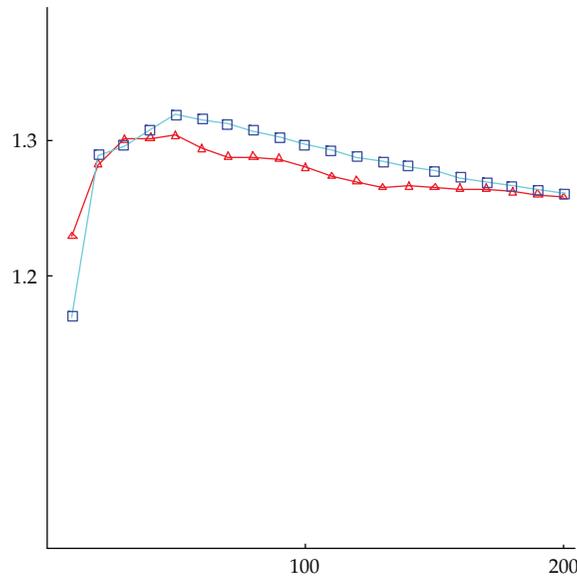


Figure 6: Fractal dimension as function of the length, $i = 10, \dots, 200$ for (H1) (red) and (H2) (blue).

acids. However, in the first case (Figure 8), the entropy of (H1) is lower than (H2), while on the contrary, for the corresponding amino acids, the entropy of (H1) is greater than (H2).

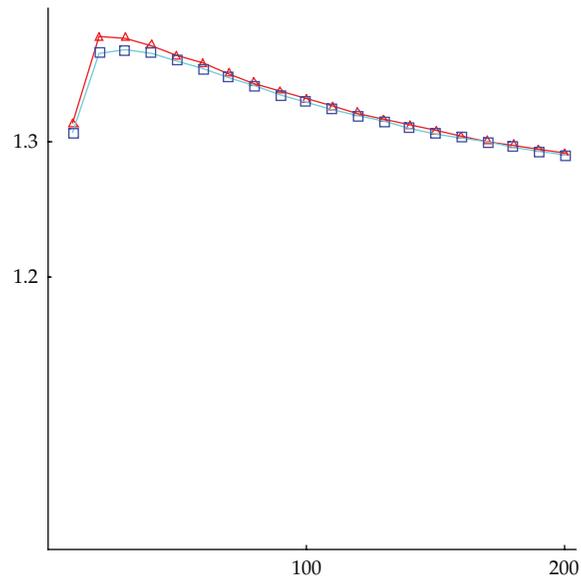


Figure 7: Fractal dimension as function of the length, $i = 10, \dots, 200$ for the amino acids of (H1) (red) and (H2) (blue).

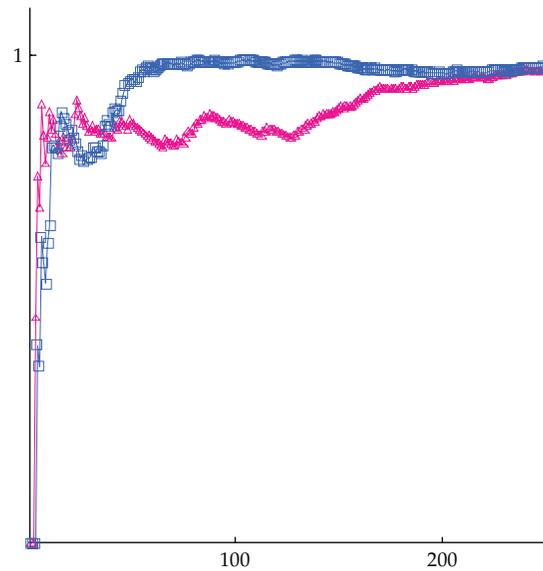


Figure 8: Entropy for the first 300 nucleotides of the sequence (H1) (red) and (H2) (blue).

7. Conclusions

In this paper, two variants of mRNA of isoforms TET2 gene have been analyzed through their nucleotide and amino acids distribution. By using the indicator function (and matrix), the fractal dimension and the entropy have been easily computed. We have noticed that, at

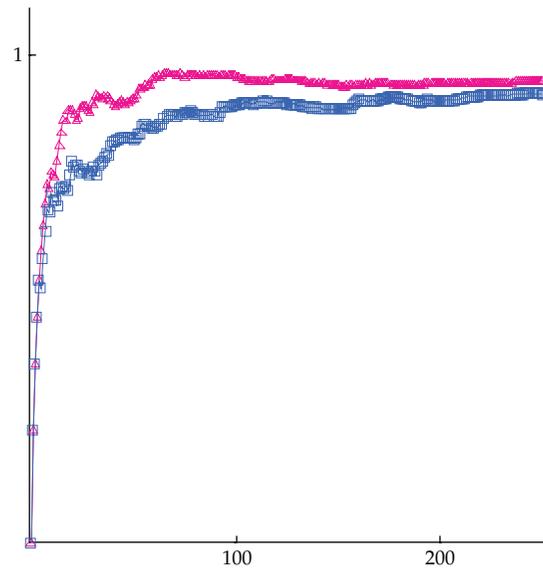


Figure 9: Entropy for the first 300 amino acids of the sequence (H1) (red) and (H2) (blue).

the amino acid level, some patches can be easily singled out. Moreover, the second variant (H2) of TET 2 shows some more randomness than (H1).

References

- [1] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.
- [2] C. Cattani, "Wavelet algorithms for DNA analysis," in *Computational Molecular Biology: Techniques, Approaches and Applications*, M. Elloumi and A. Y. Zomaya, Eds., Wiley Series in Bioinformatics, chapter 35, John Wiley & Sons, New York, NY, USA, 2010.
- [3] C. Cattani and G. Pierro, "Complexity on acute myeloid leukemia mRNA transcript variant," *Mathematical Problems in Engineering*, vol. 2011, Article ID 379873, 16 pages, 2011.
- [4] R. F. Voss, "Evolution of long-range fractal correlations and $1/f$ noise in DNA base sequences," *Physical Review Letters*, vol. 68, no. 25, pp. 3805–3808, 1992.
- [5] R. F. Voss, "Long-Range Fractal Correlations in DNA introns and exons," *Fractals*, vol. 2, pp. 1–6, 1992.
- [6] S. V. Buldyrev, A. L. Goldberger, S. Havlin et al., "Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis," *Physical Review E*, vol. 51, no. 5, pp. 5084–5091, 1995.
- [7] K. Metze, "Fractal dimension of chromatin and cancer prognosis," *Epigenomics*, vol. 2, no. 5, pp. 601–604, 2010.
- [8] R. L. Adam, R. C. Silva, F. G. Pereira, N. J. Leite, I. Lorand-Metze, and K. Metze, "The fractal dimension of nuclear chromatin as a prognostic factor in acute precursor B lymphoblastic leukemia," *Cellular Oncology*, vol. 28, no. 1-2, pp. 55–59, 2006.
- [9] K. Metze, I. Lorand-Metze, N. J. Leite, and R. L. Adam, "Goodness-of-fit of the fractal dimension as a prognostic factor," *Cellular Oncology*, vol. 31, no. 6, pp. 503–504, 2009.
- [10] L. Goutzanis, N. Papadogeorgakis, P. M. Pavlopoulos et al., "Nuclear fractal dimension as a prognostic factor in oral squamous cell carcinoma," *Oral Oncology*, vol. 44, no. 4, pp. 345–353, 2008.
- [11] A. Mashiah, O. Wolach, J. Sandbank, O. Uziel, P. Raanani, and M. Lahav, "Lymphoma and leukemia cells possess fractal dimensions that correlate with their biological features," *Acta Haematologica*, vol. 119, no. 3, pp. 142–150, 2008.
- [12] K. Metze, D. P. Ferro, M. A. Falconi et al., "Fractal characteristics of nuclear chromatin in routinely stained cytology are independent prognostic factors in patients with multiple myeloma," *Virchows Archiv*, vol. 445, supplement 1, pp. 7–21, 2009.

- [13] V. Bedin, R. L. Adam, B. C. S. de Sá, G. Landman, and K. Metze, "Fractal dimension of chromatin is an independent prognostic factor for survival in melanoma," *BMC Cancer*, vol. 10, article 260, 2010.
- [14] D. V. Lebedev, M. V. Filatov, A. I. Kuklin et al., "Fractal nature of chromatin organization in interphase chicken erythrocyte nuclei: DNA structure exhibits biphasic fractal properties," *FEBS Letters*, vol. 579, no. 6, pp. 1465–1468, 2005.
- [15] J. G. McNally and D. Mazza, "Fractal geometry in the nucleus," *The EMBO Journal*, vol. 29, no. 1, pp. 2–3, 2010.
- [16] M. Takahashi, "A fractal model of chromosomes and chromosomal DNA replication," *Journal of Theoretical Biology*, vol. 141, no. 1, pp. 117–136, 1989.
- [17] A. Delides, I. Panayiotides, A. Alegakis et al., "Fractal dimension as a prognostic factor for laryngeal carcinoma," *Anticancer Research*, vol. 25, no. 3, pp. 2141–2144, 2005.
- [18] R. C. Ferreira, P. S. de Matos, R. L. Adam, N. J. Leite, and K. Metze, "Application of the Minkowski-Bouligand fractal dimension for the differential diagnosis of thyroid follicular neoplasias," *Cellular Oncology*, vol. 28, no. 5-6, pp. 331–333, 2006.
- [19] J. M. Adams, A. W. Harris, and C. A. Pinkert, "The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice," *Nature*, vol. 318, no. 6046, pp. 533–538, 1985.
- [20] L. Pontrjagin and L. Schnirelmann, "Sur une propriété métrique de la dimension," *Annals of Mathematics*, vol. 33, pp. 156–162, 1932.
- [21] A. N. Kolmogorov and V. M. Tihomiroff, " ϵ_λ -Entropy and ϵ -capacity of sets in functional spaces," *Uspekhi Mat. Nauk*, vol. 14, no. 2, pp. 3–86, 1949.
- [22] J. Patrick Fitch and B. Sokhansanj, "Genomic engineering: moving beyond DNA sequence to function," *Proceedings of the IEEE*, vol. 88, no. 12, pp. 1949–1971, 2000.
- [23] H. Gee, "A journey into the genome: what's there," *Nature*, 2001.
- [24] P. D. Cristea, "Large scale features in DNA genomic signals," *Signal Processing*, vol. 83, no. 4, pp. 871–888, 2003.
- [25] H. Herzel, E. N. Trifonov, O. Weiss, and I. Große, "Interpreting correlations in biosequences," *Physica A*, vol. 249, no. 1–4, pp. 449–459, 1998.
- [26] W. Li, "The study of correlation structures of DNA sequences: a critical review," *Computers and Chemistry*, vol. 21, no. 4, pp. 257–271, 1997.
- [27] W. Li and K. Kaneko, "Long-range correlations and partial spectrum in a noncoding DNA sequence," *Europhysics Letters*, vol. 17, pp. 655–660, 1992.
- [28] C. K. Peng, S. V. Buldyrev, A. L. Goldberger et al., "Long-range correlations in nucleotide sequences," *Nature*, vol. 356, no. 6365, pp. 168–170, 1992.
- [29] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," *Physical Review E*, vol. 49, no. 2, pp. 1685–1689, 1994.
- [30] O. Weiss and H. Herzel, "Correlations in protein sequences and property codes," *Journal of Theoretical Biology*, vol. 190, no. 4, pp. 341–353, 1998.
- [31] Z. G. Yu, V. V. Anh, and B. Wang, "Correlation property of length sequences based on global structure of the complete genome," *Physical Review E*, vol. 63, no. 1, Article ID 011903, 2001.
- [32] P. P. Vaidyanathan and B. J. Yoon, "The role of signal-processing concepts in genomics and proteomics," *Journal of the Franklin Institute*, vol. 341, no. 1-2, pp. 111–135, 2004.
- [33] P. Bernaola-Galván, R. Román-Roldán, and J. L. Oliver, "Compositional segmentation and long-range fractal correlations in DNA sequences," *Physical Review E*, vol. 53, no. 5, pp. 5181–5189, 1996.
- [34] W. Li and J. Bentley, "The complexity of DNA: the measure of compositional heterogeneity in DNA sequence and measures of complexity," *Complexity*, vol. 3, pp. 33–37, 1997.
- [35] S. Karlin and V. Brendel, "Patchiness and correlations in DNA sequences," *Science*, vol. 259, no. 5095, pp. 677–680, 1993.
- [36] J. M. Bennett, M. L. Young, J. W. Andersen et al., "Long-term survival in acute myeloid leukemia: the Eastern Cooperative Oncology Group experience," *Cancer*, vol. 80, no. 11, pp. 2205–2209, 1997.
- [37] E. Schrödinger, *What is Life? Physical Aspects of Living Cell*, Cambridge University Press, Cambridge, UK, 1948.
- [38] National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/genbank/>.
- [39] Universal Protein Resource, <http://www.uniprot.org/help/about>.
- [40] Gene Location, Weizmann Institute of Science, <http://genecards.weizmann.ac.il/geneloc/>.
- [41] e!Ensemble, Ensembl project, EMBL-EBI & Wellcome Trust Sanger Institute, <http://www.ensemble.org>.

- [42] H. Avet-Loiseau, T. Facon, A. Daviet et al., "14q32 translocations and monosomy 13 observed in monoclonal gammopathy of undetermined significance delineate a multistep process for the oncogenesis of multiple myeloma," *Cancer Research*, vol. 59, no. 18, pp. 4546–4550, 1999.
- [43] J. Drach, J. Schuster, H. Nowotny et al., "Multiple myeloma: high incidence of chromosomal aneuploidy as detected by interphase fluorescence in situ hybridization," *Cancer Research*, vol. 55, no. 17, pp. 3854–3859, 1995.
- [44] M. Flactif, M. Zandecki, J. L. Laï et al., "Interphase fluorescence in situ hybridization (FISH) as a powerful tool for the detection of aneuploidy in multiple myeloma," *Leukemia*, vol. 9, no. 12, pp. 2109–2114, 1995.
- [45] R. Fonseca, G. J. Ahmann, S. M. Jalal et al., "Chromosomal abnormalities in systemic amyloidosis," *British Journal of Haematology*, vol. 103, no. 3, pp. 704–710, 1998.
- [46] M. Zandecki, J.-L. Laï, F. Geneviève et al., "Several cytogenetic subclones may be identified within plasma cells from patients with monoclonal gammopathy of undetermined significance, both at diagnosis and during the indolent course of this condition," *Blood*, vol. 90, no. 9, pp. 3682–3690, 1997.
- [47] G. W. Dewald, R. A. Kyle, G. A. Hicks, and P. R. Greipp, "The clinical significance of cytogenetic studies in 100 patients with multiple myeloma, plasma cell leukemia, or amyloidosis," *Blood*, vol. 66, no. 2, pp. 380–390, 1985.
- [48] B. Barlogie, J. Epstein, P. Selvanayagam, and R. Alexanian, "Plasma cell myeloma—New biological insights and advances in therapy," *Blood*, vol. 73, no. 4, pp. 865–879, 1989.
- [49] W. Liang, J. E. Hopper, and J. D. Rowley, "Karyotypic abnormalities and clinical aspects of patients with multiple myeloma and related paraproteinemic disorders," *Cancer*, vol. 44, no. 2, pp. 630–644, 1979.
- [50] G. Gahrton, L. Zech, and K. Nilsson, "2 Translocations, t(11;14) and t(1;6), in a patient with plasma cell leukaemia and 2 populations of plasma cells," *Scandinavian Journal of Haematology*, vol. 24, no. 1, pp. 42–46, 1980.
- [51] G. J. Morgan, F. E. Davies, and M. Linet, "Myeloma aetiology and epidemiology," *Biomedicine and Pharmacotherapy*, vol. 56, no. 5, pp. 223–234, 2002.
- [52] H. Avet-Loiseau, M. Attal, P. Moreau et al., "Genetic abnormalities and survival in multiple myeloma: the experience of the Intergroupe Francophone du Myélome," *Blood*, vol. 109, no. 8, pp. 3489–3495, 2007.
- [53] J. R. Sawyer, J. A. Waldron, S. Jagannath, and B. Barlogie, "Cytogenetic findings in 200 patients with multiple myeloma," *Cancer Genetics and Cytogenetics*, vol. 82, no. 1, pp. 41–49, 1995.
- [54] O. Landgren, R. A. Kyle, R. M. Pfeiffer et al., "Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study," *Blood*, vol. 113, no. 22, pp. 5412–5417, 2009.
- [55] S. M. C. Langemeijer, R. P. Kuiper, M. Berends et al., "Acquired mutations in TET2 are common in myelodysplastic syndromes," *Nature Genetics*, vol. 41, no. 7, pp. 838–842, 2009.
- [56] M. Ko, Y. Huang, A. M. Jankowska et al., "Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2," *Nature*, vol. 468, no. 7325, pp. 839–843, 2010.
- [57] R. B. Lersback, J. Moore, S. Mathew, S. C. Raimondi, S. T. Mukatira, and J. R. Downing, "TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;23) [3]," *Leukemia*, vol. 17, no. 3, pp. 637–641, 2003.
- [58] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, part 1, pp. 480–487, 2011.
- [59] S. Chen, J. Zhang, H. Zhang et al., "Myocardial motion analysis for determination of tei-index of human heart," *Sensors*, vol. 10, no. 12, pp. 11428–11439, 2010.

Research Article

A Simple Framework for Face Photo-Sketch Synthesis

Xuwei Li and Xiaochun Cao

School of Computer Science and Technology, Tianjin University, Tianjin 300072, China

Correspondence should be addressed to Xiaochun Cao, xcao@tju.edu.cn

Received 30 June 2011; Accepted 24 August 2011

Academic Editor: Shengyong Chen

Copyright © 2012 X. Li and X. Cao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a simple framework for face photo-sketch synthesis. We first describe the shadow details on faces and extract the prominent facial feature by two-scale decomposition using bilateral filtering. Then, we enhance the hair and some unapparent facial feature regions by combining the edge map and hair color similarity map. Finally, we obtain the face photo sketch by adding the results of the two processes. Compared with current methods, the proposed framework demands non feature localization, training or iteration process, creating vivid hair in sketch synthesis, and process arbitrary lighting conditions of input images, especially for complex self-shadows. And more importantly, it can be easily expanded to natural scene. The effectiveness of the presented framework is evaluated on a variety of databases.

1. Introduction

Face sketching is a simple yet expressive representation of faces. It depicts a concise sketch of a face that captures the most essential perceptual information with a number of strokes [1]. It has useful applications for both digital entertainment and law enforcement.

In recent years, two kinds of representative methods for computer-based face sketching have been presented: (1) line drawing [1–5] and (2) Eigen transformation based [6, 7]. Line drawing-based methods are expressive to convey 3D shading information, at the cost of losing sketch texture. The performances of these approaches largely depend on the shape extraction and facial feature analysis algorithms, such as active appearance model [5]. Other line drawing methods use compositional and-or graph representation [8, 9] or the direct combined model [10] to generate face photo cartoon. Eigen-transformation-based approaches use complex mathematical models to synthesize face sketches, such as PCA, LDA, E-HMM, and MRF. Gao et al. [6] use an embedded hidden Markov model and a selective ensemble strategy to synthesize sketches from photos. However, the hair region is excluded in PCA/LDA/E-HMM-based methods [11]. Wang and Tang [2] use a multiscale Markov random fields (MRFs) model to synthesize face photo-sketch and recognize it. The face

region is first divided into overlapping patches for learning, and the size of the patches decides the scale of local face structures to be learned. From a training set, then the joint photo-sketch model is learned at multiple scales using a multiscale MRF model. This method requires modeling both face shape and texture and can provide more texture information. However, the current approaches have three disadvantages: (1) both line drawing and Eigen-transformation-based methods require complex computing; (2) human face sketch is unnecessarily exaggerated, which depicts the facial feature with distortion; (3) most of the existing methods can only sketch the human face but fail to apply to natural scene images.

In this paper, we present a novel and simple face photo-sketch synthesis framework. The hair is synthesized using a two-scale decomposition and a color similarity map. The proposed framework is very simple, without any iteration, or facial feature extraction. Specially, the proposed method can easily be applied to natural scene for sketching.

A schematic overview of our framework is shown in Figure 1. Firstly, for an input face image, a two-scale image decomposition by bilateral filtering is used to describe the shading texture and the prominent feature shape, while the color similarity map-based hair creating can generate the hair texture and the unapparent facial feature. Then, the edge map is computed by edge detector from skin color similarity map. As a result, the hair, eyes, and thus mouth regions are enhanced by multiplying the edge map and hair color similarity map. Finally, the face photo sketch is synthesized by combining the results of former two processes using addition operation.

2. Bilateral Filter

2.1. Bilateral Filter

The bilateral filter is an edge-preserving filter developed by Tomasi and Manduchi [12]. It is a normalized convolution in which the weighting for each pixel p is determined by the spatial distance from the center pixel q , as well as its relative difference in intensity. The spatial and intensity weighting functions f and g are typically Gaussian [13, 14]. The spatial kernel increases the weight of pixels that are spatially close, and the weight in the intensity domain decreases the weight of pixels with large intensity differences. Therefore, bilateral filter effectively blurs an image while keeping sharp edges intact. For input image I , output image J , and a window Ω neighboring to q , the bilateral filtering is defined as follows:

$$J = \frac{\sum_{p \in \Omega} f(p - q) g(I_p - I_q) I_p}{\sum_{p \in \Omega} f(p - q) g(I_p - I_q)}, \quad (2.1)$$

where

$$f(p - q) = \exp \left[-\frac{(\|p - q\|^2)}{2\sigma_s^2} \right], \quad (2.2)$$

$$g(I_p - I_q) = \exp \left[-\frac{(I_p - I_q)^2}{2\sigma_r^2} \right].$$

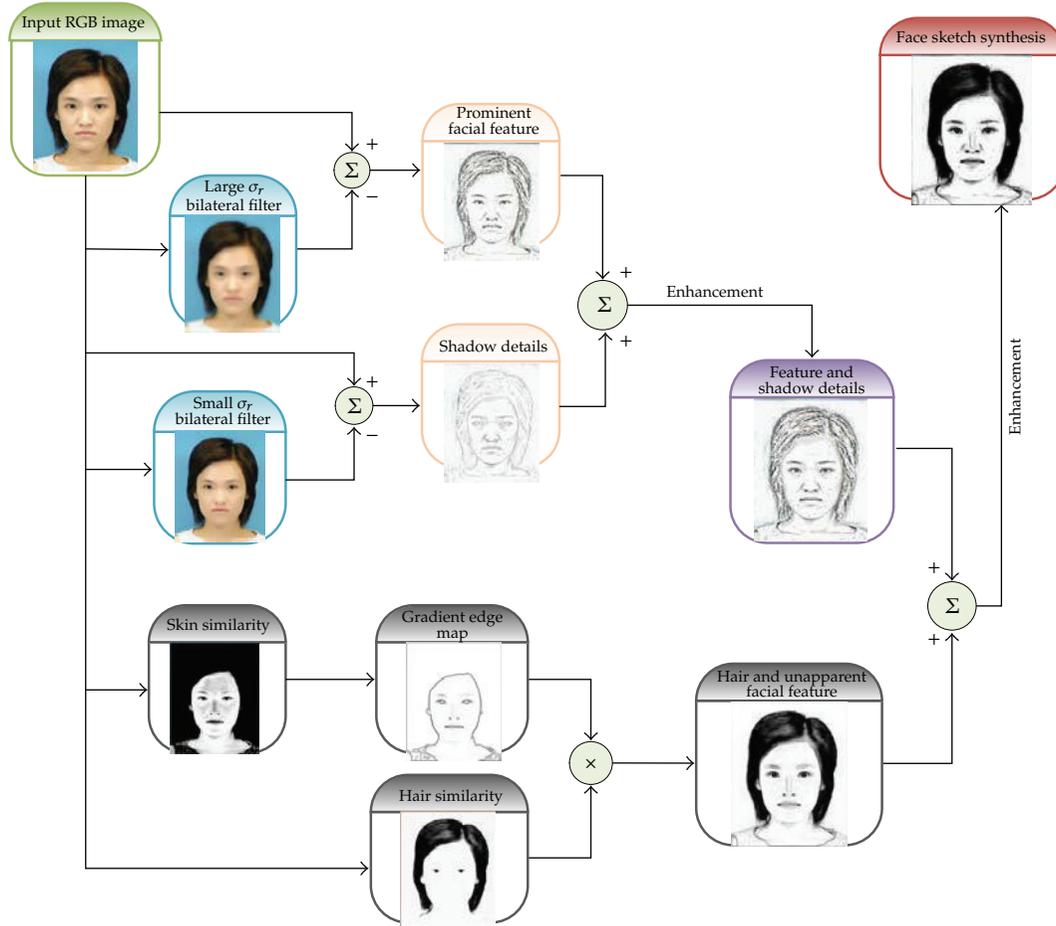


Figure 1: The system framework.

σ_s and σ_r are the size of spatial kernel and the range kernel, corresponding to the Gaussian functions f and g . When σ_s increases, the larger features in image will be smoothed; when σ_r increases, the bilateral filter will become closer to Gaussian blur.

2.2. Facial Feature Detail Detection by Two-Scale Image Decomposition

Multiscale image decomposition (or multiscale retinex, MSR) is developed by Jobson et al. [15–17], in an attempt to bridge the gap between images and the human observation of scenes. It is widely used in HDR image rendering for color reproduction and contrast reduction [18, 19], color enhancement, and color constancy processing. In HDR image rendition, the smallest scale is strong on detail and dynamic range compression but weak on tonal and color rendition. The reverse is true for the largest spatial scale. Multiscale retinex combines the strengths of each scale and mitigates the weaknesses of each [15].

Durand and Dorsey [19] only use two-scale decomposition to decompose the input image into a “base” and a “detail” image. The base layer has its contrast reduced and contains only large-scale intensity variations, which is obtained by bilateral filtering. The detail layer

is the division of the input intensity by the base layer, while the magnitude is unchanged, thus preserving detail.

Then, we will describe how two-scale decomposition can be used for lighting conditions and facial feature detail detection in face photo-sketch synthesis. According to image decomposition, we can get the detail image by subtracting the base layer from the input image. And the detail image can preserve the important details of the input image, such as edges, texture, and shadows, depending on the smoothing degree. In theory, the input image is smoothed more heavily; the details will be preserved more. Figures 2 and 3 illustrate the phenomenon.

On the other hand, facial feature and shadow details are very important to face photo sketch. The difference between sketches and photos mainly lies in two aspects: texture and shape, which are often exaggerated by the artist in sketch. The texture contains hair texture and shadow texture [2]. In this paper, we perform good shading effects near the interest features from detail images by two-scale image decomposition, proposed by Durand and Dorsey [19]. And the shape of obvious facial features is also obtained from the detail images.

The two-scale decomposition is performed on the logs of pixel intensities using piecewise-linear bilateral filtering and subsampling. On the one hand, the use of logs of intensities is because image can be considered as a product of reflectance and illumination component. So the decomposition can be viewed as an image separated into intrinsic layers of reflectance and illumination [20–22], while base layer corresponds to illumination component and detail corresponds to reflectance [23]. Therefore, we can obtain the facial feature from the detail layer because of the distinct reflectance difference in facial feature and skin region. In fact, human vision is mostly sensitive to the reflectance rather than to the illumination conditions. Even more, the logarithm function deals with the low intensity far better than those high-intensity pixels because of its function character. On the other hand, the piecewise bilateral-linear filtering in the intensity domain and a subsampling in the spatial domain can efficiently accelerate bilateral filtering.

As is described in (2.1), when the scale σ_s of the spatial kernel or/and the scale σ_r of intensity domain increases, the input image will be smoothed more. Although the scale σ_s of the spatial kernel has little influence on the result, it plays an important role in facial feature detection. Several conclusions can be observed. (1) For the same small scale σ_r of intensity domain, the increase of scale σ_s of the spatial domain will smooth more pixels near the edges of the facial feature, while it results in the heavier shadow in detail layer without change in edges. The results are shown in Figure 2 by using different spatial scales. (2) For the same scale σ_s of spatial domain, the increase of scale σ_r of the intensity domain will directly smooth the input image more than (1), because the larger intensity scale will smooth more pixels on the edges [24, 25]. Figure 3 illustrates the different results with the intensity-scale changes.

Further research demonstrates that facial feature, such as hair, eyebrows, eyes, mouth, and nose, which have low intensity than skin, can be obtained in detail image by two-scale decomposition using piecewise bilateral filtering. With the small scale σ_r in intensity domain and the larger scale σ_s in spatial domain, we can get more reflectance component responses of the low contrast area, such as the shadow near the nose and the chin, as shown in Figure 2. While using the small scale σ_s in spatial domain and larger scale σ_r of intensity domain, lower intensity pixel in small area regions can appear in detail image, including eyes and eyebrows, as shown in Figure 3. When the scales in both intensity and spatial domain are set to be largest values, we can get the lowest intensity region, such as black hair, as shown in Figure 4. However, it is noticeable that this works only for the dark hair with fuscous color. To create good hair for arbitrary color, we propose a new method in Section 3. In addition,

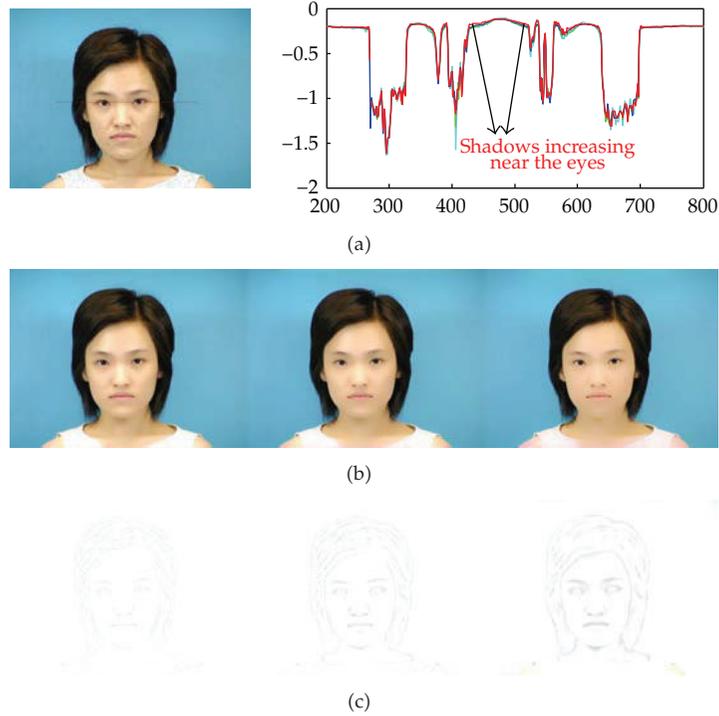


Figure 2: The influence of fixed intensity scale and varied spatial scale in decomposition. Row 1 is input image with scanline, and the scanplots of the log intensity along a row of input images (cyan) were represented using a floating point. Green, blue, and red are the corresponding scanplots of log intensity along the second row. Rows 2 and 3 are the decompositions of input images using different spatial scale, while row 2 is base images and row 3 is the detail images. From left to right, $\sigma_s = 1\%$, 2% , and 5% the size of the input image, respectively, while $\sigma_r = 0.08$.

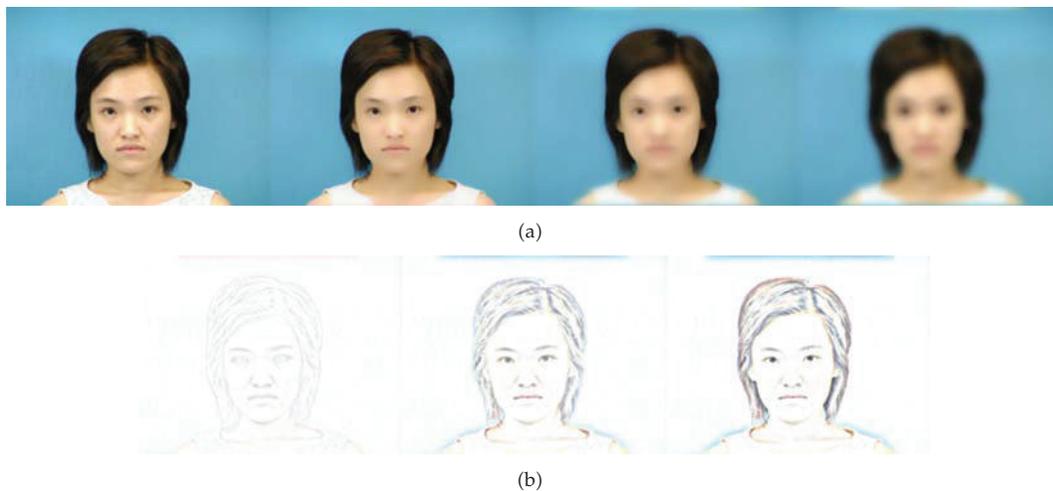


Figure 3: The influence of fixed spatial scale and varied intensity scale in decomposition. Row 1 is the base image obtained using piecewise bilateral filtering, and row 2 is the detail image. From left to right, $\sigma_r = 0.08$, 0.35 , and 0.95 , respectively, while σ_s equals 2% the size of the input image.



Figure 4: Two-scale decomposition with large spatial scale and large intensity scale. The input image is the same with Figure 3. From left to right, $\sigma_s = 10\%$, while $\sigma_r = 0.60$ and 0.95 , respectively.

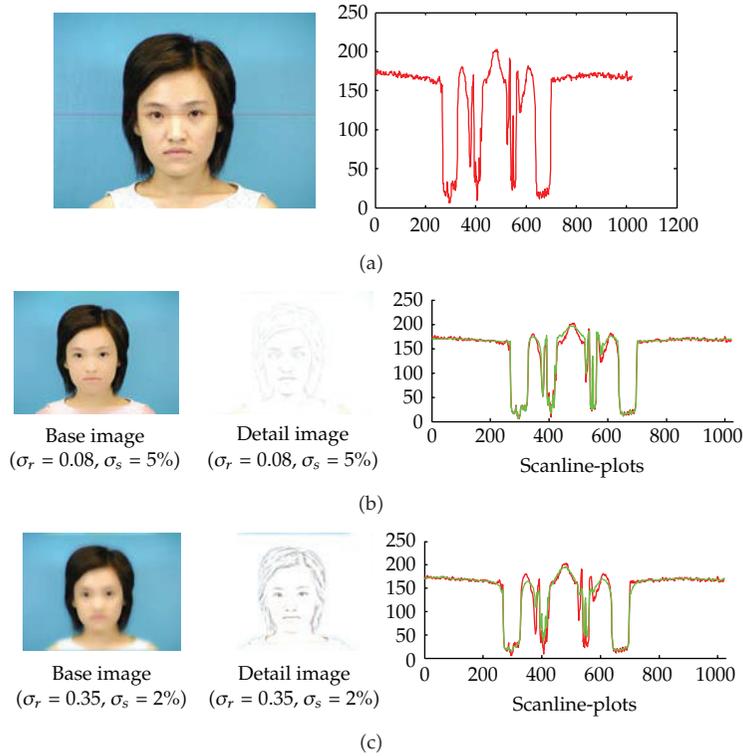


Figure 5: The input intensities (red) along a row and its decomposition into base (green) and detail by bilateral filtering. Row 1 is the input image and its scanline plots of intensity. Row 2 and row 3 are the base image and detail image using different scales of intensity and spatial domain. Row 2 shows that weak edges and reflecting details are preserved in detail image at large spatial scale and small intensity scale. Row 3 shows that the sharp edge of facial features is preserved in detail image at small spatial scale and large intensity scale.

because of the highlight on some human mouth, especially when its color and intensity are similar to skin on human face, the mouth will not be extracted at this step, and unapparent mouth extraction is dealt with in Section 3.

In this paper, we define the two-scale decomposition of input image as follows [16]:

$$R_{MSR_i}(x, y) = \sum_{n=1}^N w_n \cdot \{(\log(I_i(x, y))) - \log[F_n(x, y) * I_i(x, y)]\}. \quad (2.3)$$

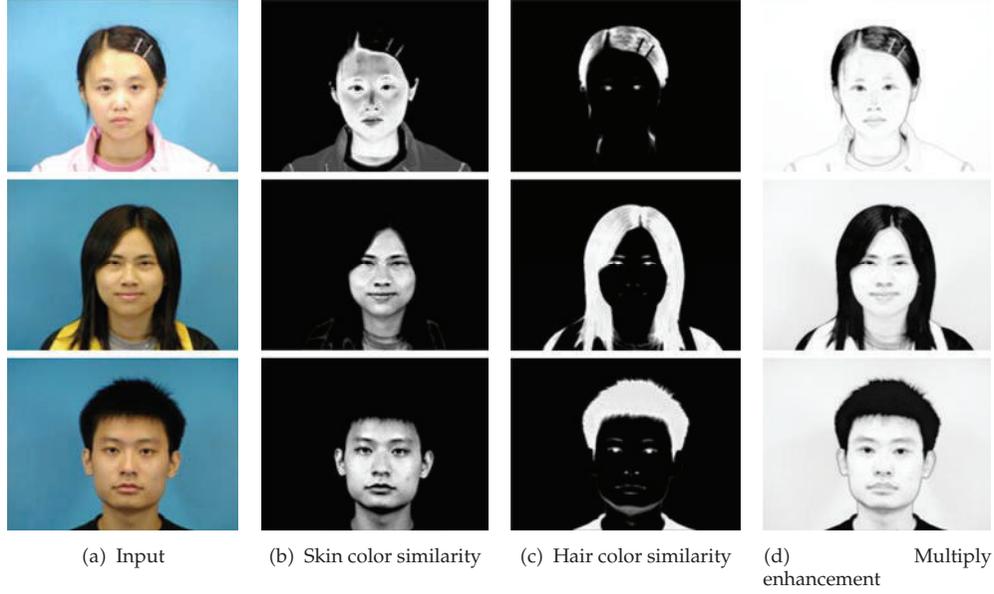


Figure 6: Hair and unapparent facial feature creating.

R_{MSR_i} is the i th color component of the MSR output, $i \in (R, G, B)$, N is the number of scales, w_n is the weighting factor for the n th scale, $I_i(x, y)$ is the image distribution in the i th color band, “*” denotes the convolution operation, and $F_n(x, y)$ is the weighting function in the n th bilateral filtering, that is, the $F_n(x, y)$ is given by,

$$F_n(x, y) = f(p - q)g(I_p - I_q) = \exp\left[\frac{-\left(\|p - q\|^2\right)}{2\sigma_{s_n}^2}\right] \cdot \exp\left[-\frac{(I_p - I_q)^2}{2\sigma_{r_n}^2}\right]. \quad (2.4)$$

So the base image is the output of bilateral filtering, and the detail image D is

$$D(x, y) = \min\{R_{MSR_i}(x, y)\}. \quad (2.5)$$

As a conclusion, to ensure the speed and effect, we set the relative scale in two-scale decomposition as follows. (1) For creating the shadows near the facial features, the small intensity scale σ_r is set to constant value of 0.05~0.08, and the associated spatial scale σ_s is set to constant value of 5% the image column size. (2) For creating the clear facial features, including eyes, eyebrows, and sometimes mouth, the large intensity scale σ_r is set to a constant value of 0.35, and the spatial scale σ_s is set to constant to a value of 2% the image column size. Experimental results demonstrate that the above fixed-scale values perform consistently well for all our face images. The results and analysis process are shown in Figure 5.

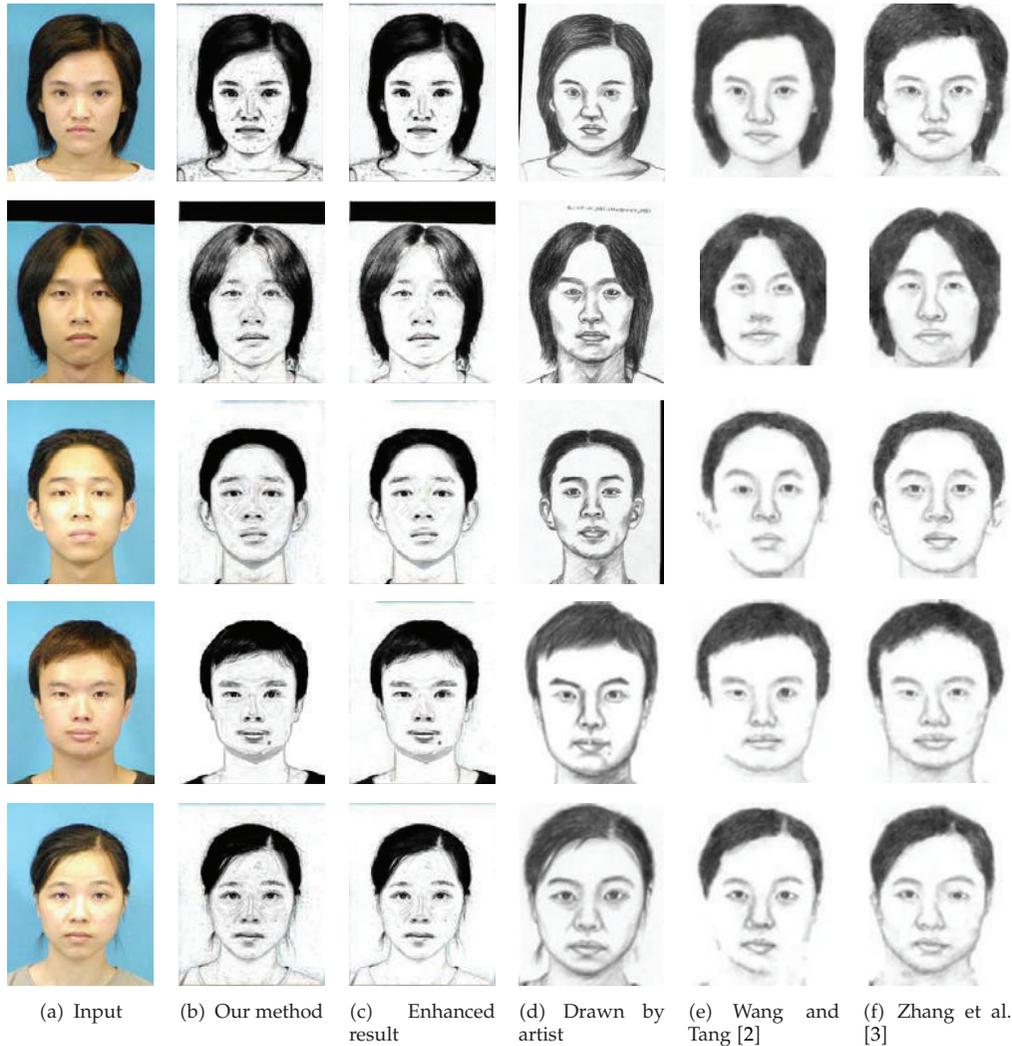


Figure 7: Face photo-sketch synthesis results: (a) the background of the cropped images in CUHK database is simple; (b) our method illustrates the hair and facial feature more accurately, especially the profile of the face; (c) in order to get the better effect, we smoothed (b) by simple enhancing method; (d) the artist draws the face sketches with some exaggerations, such as nose bridge and mouth edge; (e) in the synthesis results of Tang's method, obviously, face profile is defective, and some of the important marks, such as the mole in the fourth man's face, are lost; (f) Zhang's method draws the facial features, eyes, and mouth and so forth, with some distortion in size and shape.

3. Hair and Facial Feature Creating

3.1. Color Similarity Map

In this section, we discuss two problems: (1) computing color similarity map for input image, which is used to select the skin region and hair region separately and (2) hair and unapparent facial feature creating.

To detect the skin color region, we propose a skin/hair classification method based on color similarity. A Gaussian similarity measuring function is defined to compare

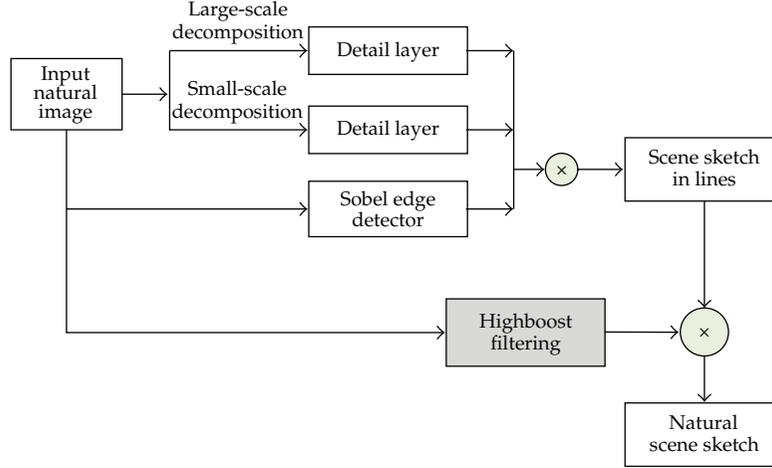


Figure 8: Flowchart of natural scene sketch.

the similarity between two colors. Gaussian convolution performs directly on every pixel of the image in the same way, which achieves the similarity of all the pixels to specific skin color by the concept of color difference to compute it. If the difference between current pixel's color with the specific color is larger, the probability of this pixel of skin/hair is lower. Let $G(x)$ denote the Gaussian masks, and E_b is the specified color (known color), color similarity function can be defined as

$$g(E_p) = e^{-\|E_p - E_b\|^2 / 2\sigma_c^2}, \quad (3.1)$$

E_p is the color of pixel p in CIE Lab color space, and $\|E_p - E_b\|^2$ is the color difference between the specified color and the pixel p in input image, whose value is determined by CIELAB color difference equation. σ_c is the threshold of color difference, which determines whether the current pixel belongs to the same kind as the known color. Generally, we keep σ_c constant to a value of 30.

To exactly compute the similarity of any color to the specified color, the first step is to confirm a benchmark color. It can be set by two methods: specified by user interaction or the program produces an average color automatically as benchmark while using a known color to select a certain percent of the most similar pixels. The time complexity of this method on confirming a skin color is lower than any face detection-based algorithm. Consequently, hair similarity map and skin similarity map can be achieved by different benchmark colors.

3.2. Hair and Unapparent Facial Feature Creating

After the skin color similarity computing, we can get the color similarity map, as shown in Figure 6. Because the color and intensity of skin distribute uniformly, the gradient is very small in skin region, while the gradient is large in hair region due to the irregular hair colors. In hair color similarity map, the hair region's value is the largest. Then the product image of gradient skin color similarity map and the negative of the hair color similarity map will strengthen the hair region by hair value minimization.



Figure 9: Comparisons of our line extraction based on initial scene sketch with existing methods: difference of Gaussian, Kang's line drawing [26, 27].

The hair region enhancement is realized by multiplying the gradient edge map of skin color similarity and the hair color similarity map, which is proposed in Section 3.1. Since facial feature regions, such as nose and mouth regions, have obvious color difference, they will be also enhanced by the multiplication. On the other hand, the color of eyes and eyebrows either is similar to hair color or is distinct from skin color. Both of the cases will be enhanced by the above operation. Figure 6 shows the process of hair and facial feature creating. The proposed method can extract good hair texture under different lighting conditions. It is noticeable that the hair creating method can be applied to other scenes, for example, image abstraction, cartoon making, and wig wearing, in which hair region is necessary.

4. Experimental Results

We tested the two-scale decomposition-based face sketch synthesis framework on CUHK face photo-sketch database, which contains 606 faces totally. All the input images' size is 1024×768 pixels. On average for the uncropped CUHK student face image, one decomposition of an input image for RGB component took about 15.5 seconds (24s with large spatial scale and 6.9s with small spatial scale) using two-scale image decomposition based on piecewise bilateral filtering. And the time for hair and unapparent feature creating based on color similarity computing took about 3 seconds. So the total synthesis time is

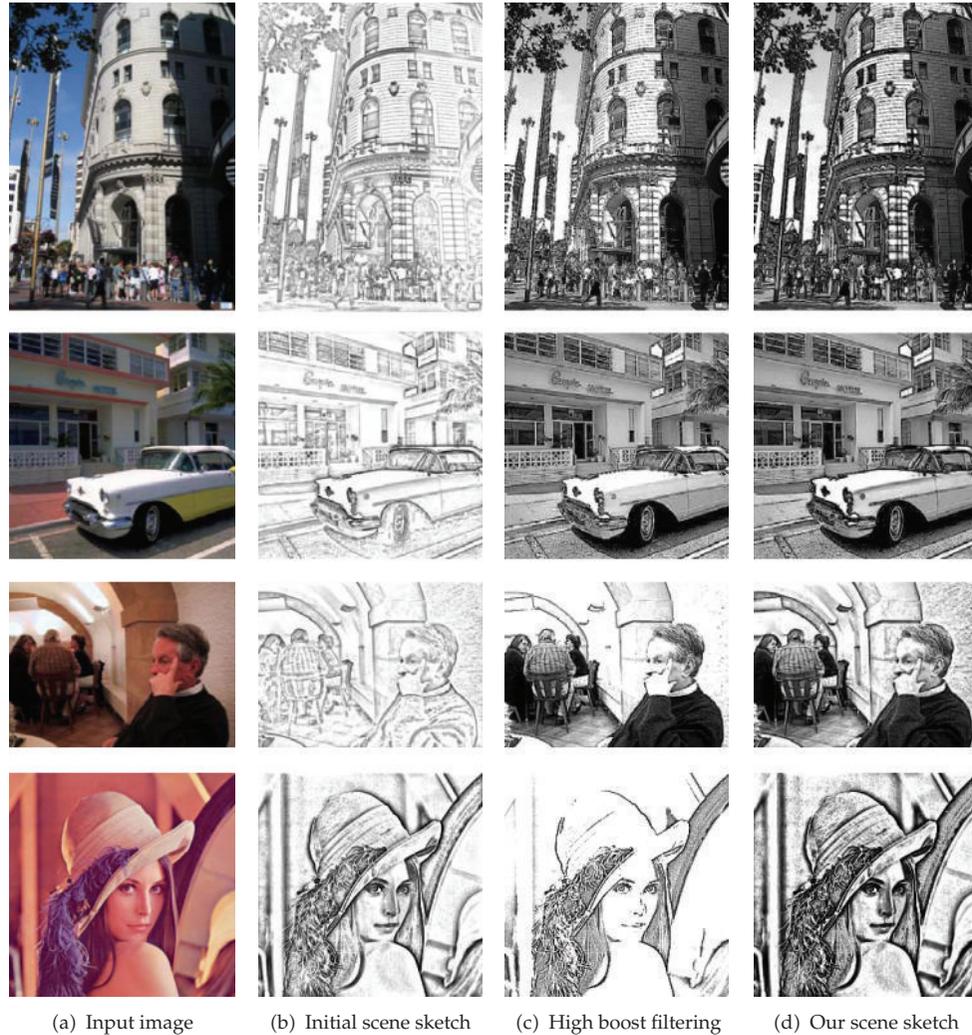


Figure 10: Arbitrary scene sketch results: our method sketches the images vividly and fast. And the test images in the second and third rows were cited in D&Carlo's paper [28].

Table 1: The computing speed of the proposed method and others.

Method	Wang and Tang [2]	Zhang et al. [3]	The proposed
Speed	>3 mins	>2 mins	<35 s

less than 35 seconds. In fact, if the decomposition is for gray image, the corresponding time consumption will be reduced to about 30 percent. All the above experiments were implemented on a 2.66G PC. If realized on the cropped image of CUHK (URL: <http://mmlab.ie.cuhk.edu.hk/facesketch.html>), whose size is 250 * 200, the time cost is 2.4 seconds for two-scale decomposition and 0.45 seconds for hair and unapparent feature creating. The speed comparison is shown in Table 1. Figure 7 shows the experimental results on CUHK cropped images.

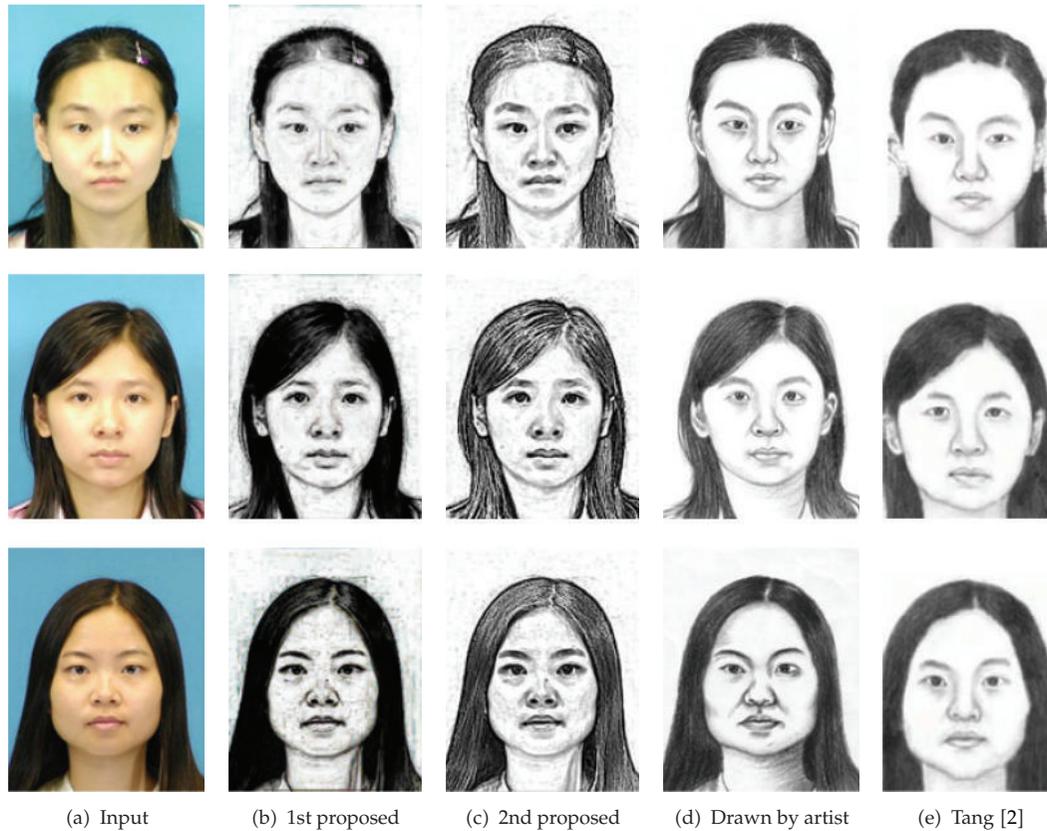


Figure 11: Comparisons of our human face sketch results with the artist and Tang's approaches. The first proposed method is introduced in Figure 1, while the second method is in Figure 8. Both the presented methods in this paper successfully depict the facial feature without distortion.

5. Extended Applications

5.1. Natural Scene Sketch and Line Extraction

The method we proposed can be applied in natural scene sketch and line extraction, with a little modification. The modified framework is shown in Figure 8. While illustrating the detail by two-scale bilateral filtering, which is the same with proposed framework, high boost filtering [29] in Figure 8 is used to enhance highlights and shadows in image. Sobel detector can give prominence to the distinct edges. Both the computation cost of high boost filtering and Sobel detector are small, so the natural scene sketch is very fast.

When multiplying the Sobel edge by the two detail layers, we can get the initial natural scene sketch, which has a line look, as shown in Figure 9(b). Then, we can extract the lines using Difference of Gaussian (DoG). Without human skin similarity computing and human hair extraction, the line drawing process is speeded up. The results are shown in Figure 9. Our method performs well on thick subtle edges in input images. Although Kang's method depicts the edges with smooth and coherent lines [26, 27], its speed is very low because of LIC. In addition, it is hard to detect the dense edges in some region, such as the edges in the building in the second image. On the other hand, DoG operator on the input image fails to deal with the edges in shadows and details.



Figure 12: Image abstraction based on our line extraction

Table 2: The computing speed of the proposed line extraction method and other.

Image size	Our line extraction method	Kang et al. [26]
512*512	3.8 s	13.4 s
256*256	0.64 s	3.3 s

Our line extraction approach takes less than 0.7 second for a $256 * 256$ image to synthesize a sketch. We implement our method using MATLAB and run the codes on a computer with 2.20 GHz CPU. The speed comparison is shown in Table 2.

If we want to get the better sketch of any input images, high boost filtering is preferred. Our approach operates well on any kind of images, such as outdoor natural scenes, animals, plants, building, and human faces. Some of the sketch results are shown in Figure 10. Figure 11 shows the human face sketch results of the CUHK database, which is introduced in Section 4. And all the parameters in bilateral filtering are the same with Section 2.2. The performance of our approach is well when it is used in human face. It can deal with gray images and color images of different races and different lighting conditions. More results on other kinds of images are shown in the appendix.

5.2. Image Stylization

Based on the initial line sketch, we can easily extract the edges by DoG operator. In combination with base layer in two-scale decomposition after color quantization, we can get the simple image abstraction as proposed in [27, 28]. The results are shown in Figure 12.



Figure 13: The 1st row is input human face. The 2nd row is initial face sketch results in lines by the proposed approach in Section 5. The 3rd row is the final sketches by the proposed approach in Section 5.



Figure 14: Some sketch results of Olivetti-Att-ORL gray-scale database, which is built for face identification. The first row and the second row are input images and the corresponding sketch results. The ORL database can be downloaded from <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.



Figure 15: Some sketch results of MUCT color face database, which contains human faces of different races and different lighting conditions. The 1st row and the 2nd row are input images, and sketch results, respectively. And the database can be downloaded from <http://code.google.com/p/muct/downloads/list>. The input image size is 640*480.

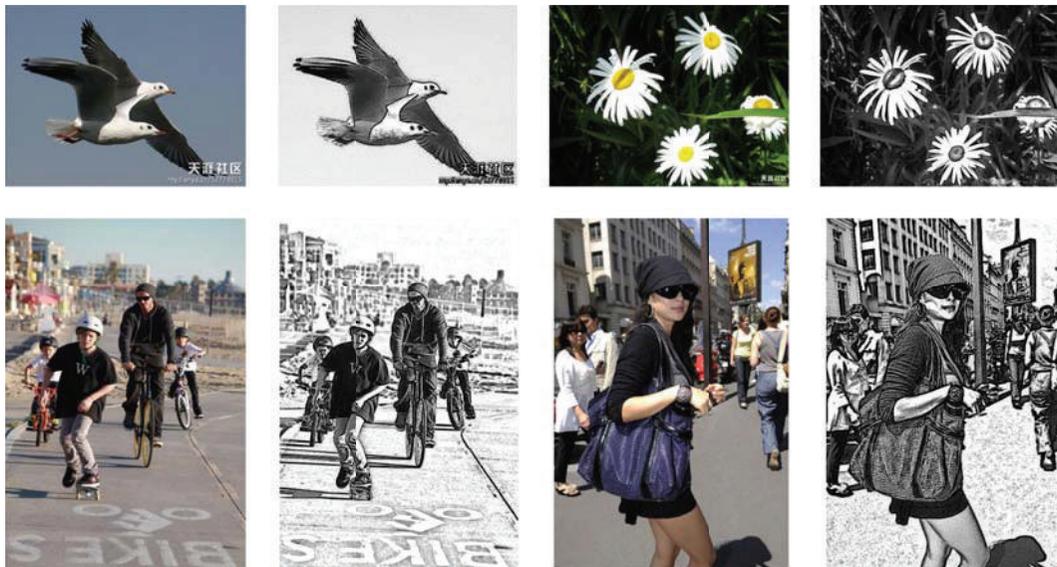


Figure 16: Some sketch results of images under varying lighting conditions downloaded from Internet. The first and the third columns are input images and the other columns are sketch results, respectively.

6. Discussion and Conclusion

We have presented a novel framework for human face photo-sketch synthesis, which is very simple, fast and requires no parameter setting. Firstly, by combing the two-time two-scale image decomposition results, the detail reflecting conditions of facial features and the prominent facial features are obtained. Secondly, based on the color similarity map, we extracted the unapparent facial features and created the vivid hair region. Finally, we

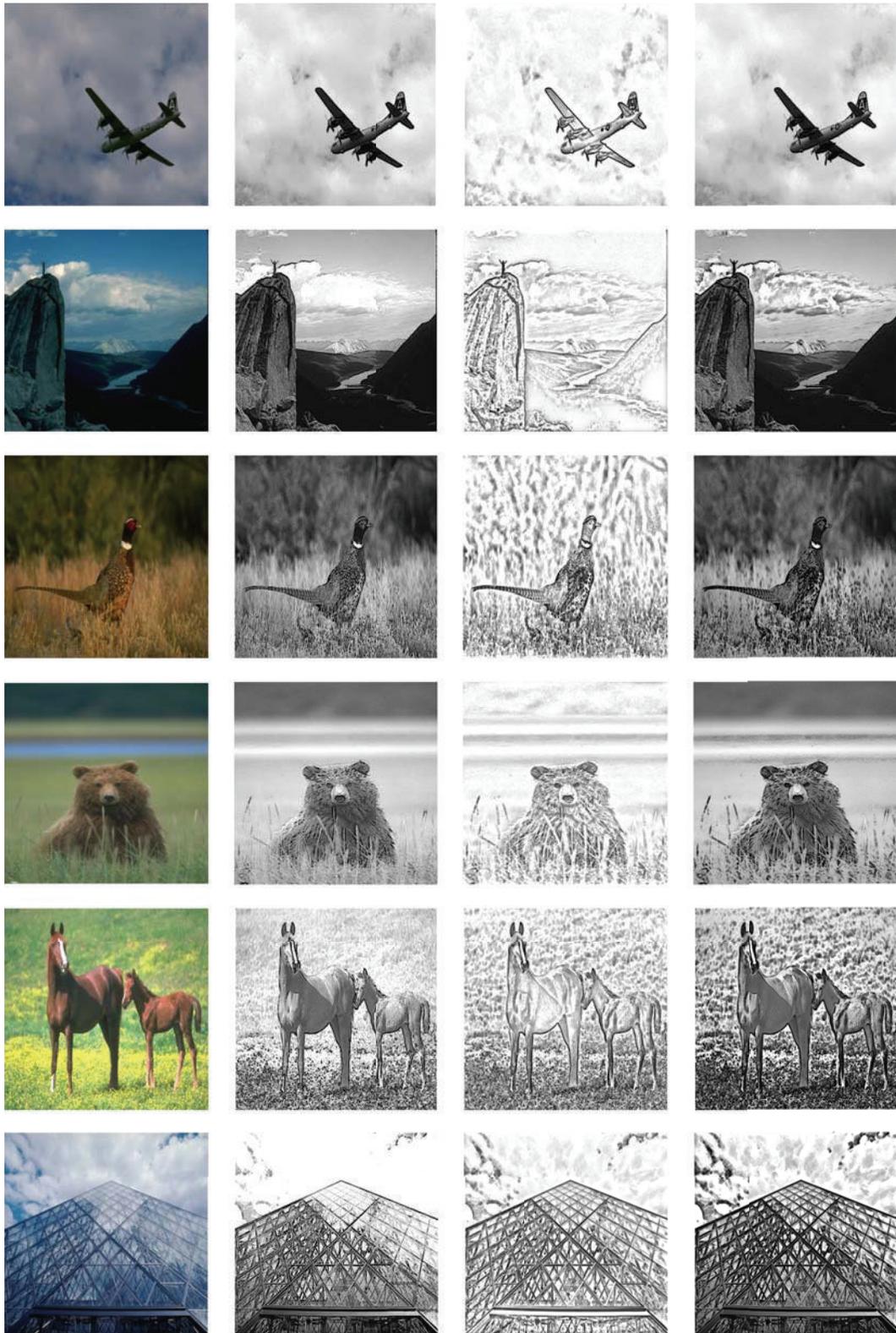


Figure 17: Continued.

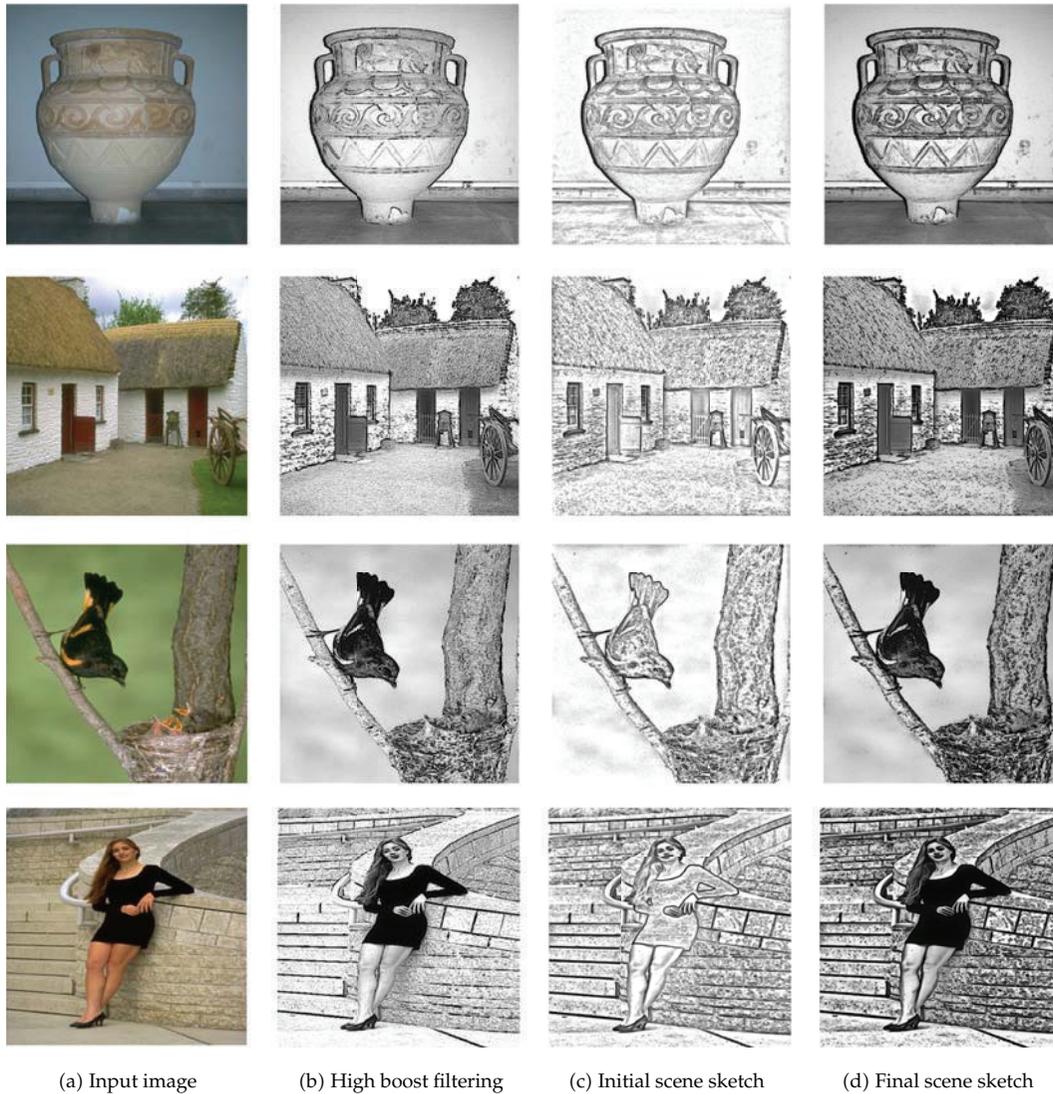


Figure 17: Other natural scene sketches of images downloaded from the Berkeley image segmentation database [30].

exploited the framework by the simple addition operation of the former results, which can be applied to other races, such as the white race. Moreover, the framework can also be expanded to other applications, such as natural scene sketch and line extraction. In conclusion, the proposed framework is very simple in that no feature localization algorithm, complex mathematic model, or iteration is needed. In addition, the sketch synthesis result is more vivid than other methods, especially in hair texture creating. Most importantly, the method is easy to be used for other applications, such as line extraction, natural scene sketches, and image abstraction. Sketch recognition is more and more widely used in Sketch-based user interfaces [31]. Since the predominant forces in line extraction of our sketch method, we will try to implement recognition of face photo sketch or other image sketch.

Appendix

See Figures 13, 14, 15, 16, and 17.

Acknowledgment

This work was supported by the National Science Foundation of China under Grant no. 61003200 and by Program for New Century Excellent Talents in University, Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology, SKL of PMTI, SKL of CG&CAD, and Open Projects Program of NLPR.

References

- [1] Uhl R. G. and N. D. V. Lobo, "A Framework for recognizing a facial image from a police sketch," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 586–593, 1996.
- [2] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009.
- [3] W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," in *Proceedings of the European Conference on Computer Vision (ECCV '10)*, pp. 420–433, 2010.
- [4] H. Chen, Z. Liu, C. Rose, Y. Xu, H.-Y. Shum, and D. Salesin, "Example-based composite sketching of human portraits," in *Proceedings of the 3rd International ACM Symposium on Non-Photorealistic Animation and Rendering (NPAR '04)*, pp. 95–153, June 2004.
- [5] Z. J. Xu, H. Chen, and S. C. Zhu, "A high resolution grammatical model for face representation and sketching," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition 2 (CVPR 2 '05)*, vol. 2, pp. 470–477, 2005.
- [6] X. Gao, J. Zhong, J. Li, and C. Tian, "Face sketch synthesis algorithm based on E-HMM and selective ensemble," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 487–496, 2008.
- [7] X. Tang and X. Wang, "Face sketch synthesis and recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 1, pp. 687–694, 2003.
- [8] F. Min, J. L. Suo, S. C. Zhu, and N. Sang, "An automatic portrait system based on and-or graph representation," in *Proceedings of the International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR '07)*, pp. 184–197, August 2007.
- [9] Z. Xu, H. Chen, S. C. Zhu, and J. Luo, "A hierarchical compositional model for face representation and sketching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 955–969, 2008.
- [10] C.-T. Tu and J.-J. Lien, "Automatic location of facial feature points and synthesis of facial sketches using direct combined model," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 40, no. 4, pp. 1158–1169, 2010.
- [11] S. Y. Chen, H. Tong, C. Cattani et al., "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
- [12] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 839–846, January 1998.
- [13] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," in *Proceedings of the European Conference on Computer Vision (ECCV '06)*, pp. 568–580, 2006.
- [14] B. Weiss, "Fast median and bilateral filtering," in *Proceedings of the ACM Siggraph*, pp. 519–526, July 2006.
- [15] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.
- [16] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 1003–1006, 1996.
- [17] K. Barnard and B. Funt, "Investigations into multi-scale retinex," in *Proceedings of the Colour Imaging in Multimedia*, pp. 9–17, Derby, UK, March 1998.

- [18] J. Kuang and D. F. Mark, "iCAM06, HDR, and image appearance," in *Proceedings of the IS&T/SID 15th Color Imaging Conference*, Albuquerque, NM, USA, 2007.
- [19] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high dynamic-range image," in *Proceedings of the ACM Transactions on Graphics, ACM Siggraph (ACM '02)*, pp. 257–266, 2002.
- [20] W. Zhang, X. Cao, J. Zhang et al., "Detecting photographic composites using shadows," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '09)*, pp. 1042–104, 2009.
- [21] B. M. Oh, M. Chen, J. Dorsey, and F. Dorsey, "Image-based modeling and photo editing," in *Proceedings of the ACM Siggraph*, pp. 433–442, 2001.
- [22] Q. Liu, X. Cao, C. Deng, and X. Guo, "Identifying image composites through shadow matte consistency," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, 2011.
- [23] J. Tumblin and G. Turk, "LCIS: a boundary hierarchy for detail-preserving contrast reduction," in *Proceedings of the Computer Graphics Proceedings (SIGGRAPH '99)*, A. Rockwood, Ed., Annual Conference Series, pp. 83–90, Addison Wesley Longman, Los Angeles, Calif, USA, 1999.
- [24] 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.
- [25] 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.
- [26] H. Kang, S. Lee, and C. Chui, "Coherent line drawing," in *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering (NPAR '07)*, pp. 43–50, August 2007.
- [27] H. Kang, S. Lee, and C. K. Chui, "Flow-based image abstraction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 1, pp. 62–76, 2009.
- [28] D. DeCarlo and A. Santella, "Stylization and abstraction of photographs," in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques (ACM SIGGRAPH '02)*, pp. 769–776, 2002.
- [29] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing*, Publishing House of Electronics Industry, Beijing, China, 2nd edition, 2002.
- [30] <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/> .
- [31] R. Arandjelovic and T. M. Sezgin, "Sketch recognition by fusion of temporal and image-based features," *Pattern Recognition*, vol. 44, no. 6, pp. 1225–1234, 2011.

Research Article

Key Issues in Modeling of Complex 3D Structures from Video Sequences

Shengyong Chen,¹ Yuehui Wang,² and Carlo Cattani³

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

³ Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy

Correspondence should be addressed to Shengyong Chen, sy@ieee.org

Received 1 July 2011; Accepted 22 August 2011

Academic Editor: Gani Aldashev

Copyright © 2012 Shengyong Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Construction of three-dimensional structures from video sequences has wide applications for intelligent video analysis. This paper summarizes the key issues of the theory and surveys the recent advances in the state of the art. Reconstruction of a scene object from video sequences often takes the basic principle of structure from motion with an uncalibrated camera. This paper lists the typical strategies and summarizes the typical solutions or algorithms for modeling of complex three-dimensional structures. Open difficult problems are also suggested for further study.

1. Introduction

Over the past two decades, many researchers seek to reconstruct the model of a three-dimensional (3D) scene structure and camera motion from video sequences taken with an uncalibrated camera or unordered photo collections from the Internet. Most traditionally, depth measurement and 3D metric reconstruction can be done from two uncalibrated stereo images [1]. Nowadays, reconstructing a 3D scene from a moving camera is one of the most important issues in the field of computer vision. This is a very challenging task because of its computational efficiency, generality, complexity, and exactitude. In this paper, we aim to show the development and current status of the 3D reconstruction algorithms on this topic.

The basic concept and knowledge of the problem can be found from the fundamentals of the multiview geometry through the books and thesis such as *Multiple View Geometry in Computer Vision* [2], *The Geometry of Multiple Images* [3], *Triangulation* [4], and some typical publications [5–8], which are independent for implementing an entire system. Multiple-view geometry is most fundamental in computer vision, and the algorithms of structure from

motion are based on the perspective geometry, affine geometry, and the Euclidean geometry. For simultaneous computation of 3D points and camera positions, this is a linear algorithm framework for the Euclidean structure recovery utilizing a scaled orthographic view and perspective views based on having a reference plane visible in all views [9]. There is an affine framework for perspective views that are captured by a single extremely simple equation based on a viewer-centered invariant, called relative affine structure [10]. A comprehensive method is used for estimating scene structure and camera motion from an image sequence taken by affine cameras which can incorporate all point, line, and conic features in a unified manner [11]. The other approach tries to calculate the cameras along with the 3D points, only relying on established correspondences between the observed images. These systems and improvements are covered in many publications [2, 6, 12–15]. The literature gives a compact yet accessible overview covering a complete reconstruction system.

For multiview modeling of a rigid scene, an approach is presented in [16], which merges traditional approaches to reconstructing image-extractable features, and modeling via user-provided geometry includes steps to obtain features for a first guess of the structure and motion, fit geometric primitives, correct the structure so that reconstructed features would lie exactly on geometric primitives, and optimize both structure and motion in a bundle adjustment manner. A nonlinear least square algorithm is presented in [17] for recovering 3D shape and motion from image streams.

Sparse 3D measurements of real scenes are readily estimated from N-view image sequences using structure-from-motion techniques. There is a fast algorithm for rigid structure from image sequences in [18]. Hilton presents a geometric theory for reconstruction of surface models from sparse 3D data captured from N camera views [19] for 3D shape reconstruction by using vanishing points [20]. Relative affine structure is given for canonical model for 3D from 2D geometry and applications [10].

The paper describes the progress in automatic recovering 3D scene structures together with 3D camera positions from a sequence of images acquired by an unknown camera undergoing unknown movement [12]. The main departure from previous structure from motion strategies is that the processing is not sequential. Instead, a hierarchical approach is employed for building from image triplets and associated trifocal tensors. A method is presented for dealing with hundreds of images without precise calibration knowledge [21]. Optimizing just over the motion unknowns is fast, and given the recovered motion, one can recover the optimal structure algebraically for two images [4].

In fact, reconstruction of nonrigid scenes is very important in structure from motion. The recovery of 3D structure and camera motion for nonrigid scenes from single-camera video footages is a key problem in computer vision. For an implicit imaging model of nonrigid scenes, there is an approach that gives a nonrigid structure-from-motion algorithm based on computing matching tensors over subsequences, and each nonrigid matching tensor is computed, along with the rank of the subsequence, using a robust estimator incorporating a model selection criterion that detects erroneous image points [22]. Uncalibrated motion captures exploiting articulated structure constraints [23] such as humans. The technique shows promise as a means of creating 3D animations of dynamic activities such as sports events. For the problem of 3D reconstruction of nonrigid objects from uncalibrated image sequences, under the assumption of an affine camera and that the nonrigid object is composed of a rigid part and a deformation part, a stratification approach can be used to recover the structure of nonrigid objects by first reconstructing the structure in affine space and then upgrading it to the Euclidean space [24]. In addition, a general framework of locally rigid motion for solving the M-point and N-view structure-from-motion problem for

unknown bodies deforming under orthography is presented in [25]. An incremental approach is presented in [26], where a new framework for nonrigid structure from motion simultaneously addresses three significant challenges: severe occlusion, perspective camera projection, and large non-linear deformation.

With the development of structure-from-motion algorithms, geometry constraint and optimization are necessary for reconstructing a good 3D model of the object or scene. Many researchers give us some useful approaches. For example, a technique is proposed in [27] for estimating piecewise planar models of objects from their images and geometric constraints and 3D structure from a single calibrated view using distance constraints [28]. Marques and Costeira present an approach to estimating 3D shape from degenerated sequences with missing data [29]. Beyond the epipolar constraint, it improves the effect of structure from motion [30].

3D affine measurements may be computed from a single perspective view of a scene given only minimal geometric information determined from the image. This minimal information is typically the vanishing line of a reference plane and a vanishing point for a direction not parallel to the plane. Without camera parameters, Criminisi et al. [31] show how to (i) compute the distance between planes parallel to the reference plane; (ii) compute area and length ratios on any plane parallel to the reference plane; (iii) determine the camera' location. Direct estimation is the fundamental estimation of scene structure and camera motion from a sequence of images. No computation of optical flow or feature correspondences is required [32]. A good critique on structure-from-motion algorithms can be found in [33] by Oliensis.

The remainder of this paper is organized as follows. Section 2 briefly gives some typical applications of structure from video sequences. Section 3 introduces the general reconstruction principle of structure from video sequences and unstructured photo collections. Section 4 outlines the methods for structure and motion estimation. Section 5 discusses the relevant available algorithms for every step to obtain a better result. We offer our impressions of current and future trends in the topic and conclude the development in Sections 6 and 7.

2. Typical Applications

2.1. Modeling and Reconstruction of 3D Buildings or Landmarks

For 3D reconstruction of an object or building, Pollefeys et al. typically present a complete system to build visual model with a hand-held camera [6]. There is a system for photorealistic 3D reconstruction from hand-held cameras [34]. Sinha et al. [35] present an algorithm for interactive 3D architectural models from unordered photo collections. There is a fully automated 3D reconstruction and visualization system for architectural scenes including its interiors and exteriors [36]. The system utilizes structure-from-motion, multiview stereo and a stereo algorithm.

The 3D models of historical relics and buildings, for example, the Emperor Qin's Terracotta Warriors and Piazza San Marco, have very significant meanings for archeologists. A system that can match and reconstruct 3D scenes from extremely large collections of photographs has been developed by Agarwal et al. [37]. A method for enabling existing multiview stereo algorithms to operate on extremely large unstructured photograph collections has been contrived by Furukawa et al. [38]. This approach is to decompose the collection into a set

of overlapping sets of photos that can be processed in parallel and to merge the resulting reconstructions [38]. People want to sightsee the famous buildings or landscapes from the Internet; they could tour the world via building a web-scale landmark recognition engine [39].

Modeling and recognizing landmarks at world scale is a useful yet challenging task. There exists no readily available list of worldwide landmarks. Obtaining reliable visual models for each landmark can also pose problems, and efficiency is another challenge for such a large-scale system. Zheng et al. leverage the vast amount of multimedia data on the web, the availability of an Internet image search engine, and advances in object recognition and clustering techniques, to address these issues [39].

2.2. Urban Reconstruction

Modeling the world and reconstructing a city present many challenges for a visualization system in computer vision. It can use some products such as Google Earth Google Map. For instance, Pollefeys et al. [40] present a system for automatic, georegistered, real-time multiview stereo 3D reconstruction from long image sequences of urban scenes. The system collects video streams, as well as GPS and inertia measurements in order to obtain the georegistered coordinates of the 3D models [40]. Faugeras et al. [41] address the problem of recovery of a realistic textured model of a scene from a sequence of images, without any prior knowledge either about the parameters of the cameras or about their motion.

2.3. Navigation

If the world's model or the city's reconstruction is exhaustively completed, we can obtain relative location of the buildings and find related views for navigation for robots or other vision systems. Photo Tourism can enable full 3D navigation and exploration of the set of images and world geometry, along with auxiliary information such as overhead maps [14]. It gives several modes for navigation, including free-flight navigation, moving between related views, object-based navigation, and creating stabilized slideshows. The system by Pollefeys et al. also contains the navigation function [40]. Supplying realistically textured 3D city models at ground level promises to be useful for previsualizing upcoming traffic situations in car navigation systems [42].

2.4. Visual Servoing

In the literature, there are applications that can employ SfM algorithms successfully in practical engineering. For instance, based on structure from controlled motion or on robust statistics, a visual servoing system is presented in [43]. A general-purpose image understanding system via a control structure is designed by Marengoni et al. [44] and 3D video compression via topology matching [45]. More applications are being developed by researchers and engineers in the community.

2.5. Scene Recognition and Understanding

3D reconstruction is an important application to face recognition, facial expression analysis, and so on. Fidaleo and Medioni [46] design a model-assisted system for reconstruction of 3D

faces from a single-consumer quality camera using a structure-from-motion approach. Park and Jain [47] present an algorithm for 3D-model-based face recognition in video.

Reconstruction of 3D scene geometry is an important element for scene understanding, autonomous vehicle and robot navigation, image retrieval, and 3D television [48]. Nedovic et al. propose accounting for the inherent structure of the visual world when trying to solve the scene reconstruction problem [48].

3. Information Organization

The goal of structure-form-motion is automatic recovery of camera motion and scene structure from two or more images. The problem of using pixel correspondences or track points to determine camera and point geometry in this manner is known as structure from motion. It is a self-calibration technique and called automatic camera tracking or match moving. We must consider several questions like

- (1) Correspondence (feature extracting and tracking or matching): given a point in one image, how does it constrain the position of the corresponding point in other images?
- (2) Scene geometry (structure): given point matches in two or more images, where are the corresponding points in 3D?
- (3) Camera geometry (motion): given a set of corresponding points in two or more images, what are the camera matrices for these views?

Based on these questions, we can give the 3D reconstruction pipeline as in Figure 1. The goal of correspondence is to build a set of matching 2D coordinates of pixels across the video sequences. It is a significant step in the flow of the structure from motion. Correspondence is always a challenging task in computer vision. So far, many researchers have developed some practical and robust algorithms. Given a video sequence of scene, how can we find matching points?

Firstly, there are some well-known algorithms for image sequences or videos; one popular is the KLT tracker [49–51]. It gives us an integrated system that can automatically detect the KLT feature points and track them. However, it cannot apply to the situations with wide baseline, illustration changing, variant scale, duplicate and similar structure, occlusion, noise, image distortion, and so on. Generally speaking, for video sequences, the KLT tracker can perform a good effect. Figures 2 and 3 show examples of the feature points of the KLT detector output with example images from <http://www.ces.clemson.edu/~stb/klt/>.

In the KLT tracker [49–51], if the time interval between two frames of video is sufficiently short, we can suppose that the positions of feature points move, but their intensities do not change; that is,

$$I(\mathbf{x}, t) = I(\delta(\mathbf{x}), t + \Delta t), \quad (3.1)$$

where \mathbf{x} is the position of a feature point and $\delta(\mathbf{x})$ is a transformation function.

In the papers of Lucas and Kanade [49], Tomasi and Kanade [50], and Shi and Tomasi [51], the authors made an important hypothesis that for high enough frame rates, $\delta(\mathbf{x})$ can be approximated with a displacement vector \mathbf{d} :

$$I(\mathbf{x}, t) = I(\mathbf{x} + \mathbf{d}, t + \Delta t). \quad (3.2)$$

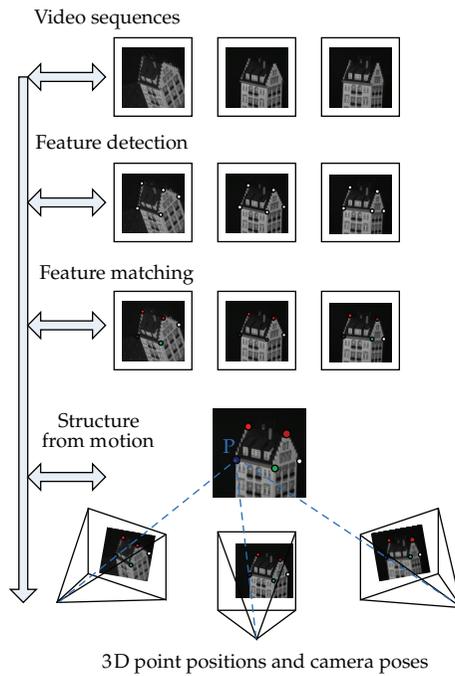


Figure 1: 3D reconstruction pipeline.



Figure 2: Example set of detected KLT features.

Then symmetric definition for the dissimilarity between two windows, one in image $I(x, t)$ and one in image $I(x + \mathbf{d}, t + \Delta t)$, is as follows:

$$\varepsilon = \iint_W [I(x + \mathbf{d}, t + \Delta t) - I(x, t)]^2 \omega(x) dx, \quad (3.3)$$

where $\omega(x)$ is the weighting function, usually set to the constant 1. The algorithm is calculating the vector \mathbf{d} which minimizes. Now, utilizing the first-order Taylor expansion of



Figure 3: Tracking trajectory of KLT tracker through a video sequence.

$I(\mathbf{x} + \mathbf{d}, t + \Delta t)$ to truncate to the linear term and setting the derivative of ϵ with respect to \mathbf{d} to $\mathbf{0}$, obtaining the linear equation:

$$Z\mathbf{d} = \mathbf{e}, \quad (3.4)$$

where Z is the following 2×2 matrix:

$$Z = \iint_W g(\mathbf{x})g^T(\mathbf{x})\omega(\mathbf{x})d\mathbf{x} \quad (3.5)$$

and \mathbf{e} is the following 2×1 vector:

$$\mathbf{e} = \iint_W [I(\mathbf{x} + \mathbf{d}, t + \Delta t) - I(\mathbf{x}, t)]g(\mathbf{x})\omega(\mathbf{x})d\mathbf{x}, \quad (3.6)$$

where $g(\mathbf{x}) = \partial I / \partial \mathbf{x}$.

On the other hand, for a completely unorganized set of images, the tracker becomes invalid. There is another popular algorithm in computer vision area, named scale-invariant feature transform (SIFT) [52]. It is effective to feature detection and matching in a wide class of image transformation, including rotations, scales, and changes in brightness or contrast, and to recognize panoramas [53]. Figures 4 and 5 show examples of the feature points of the SIFT output with example images from <http://www.cs.ubc.ca/~lowe/keypoints/>.

4. Structure and Motion Estimation

Assume that we have obtained a set of correspondences between images or video sequence, and then we use the set to reconstruct the 3D structure of each point in the set of correspondences and recover the motion of a camera. This task is called structure from motion. The problem has been an active research topic in computer vision since the development of the Longuet-Higgins eight-point algorithm [54] that focused on reconstructing geometry

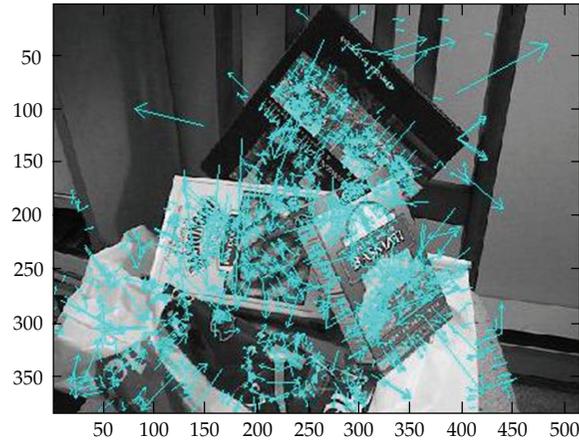


Figure 4: Example set of detected SIFT features.

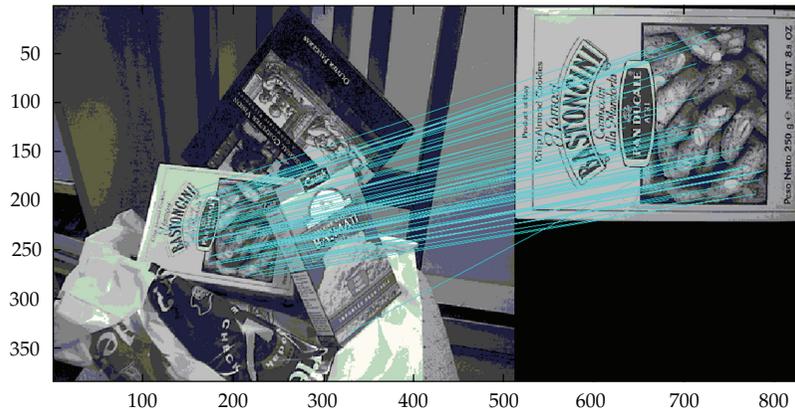


Figure 5: SIFT feature matches between images.

from two views. In the literature [2], several different approaches to solve the structure-from-motion problem are given.

4.1. Factorization

There is a popular factorization algorithm for image streams under orthography, using many images and tracking many feature points to obtain highly redundant feature position information, which was firstly developed by Tomasi and Kanade [55] in the 1990s. The main idea of this algorithm is to factorize the tracking matrix into structure and motion matrices simultaneously via singular value decomposition (SVD) method with low-rank approximation, taking advantage of the linear algebraic properties of orthographic projection.

However, an orthographic formulation limits the range of motions the method can accommodate. Perspective projection is a projection model that closely approximates perspective projection by modeling several effects not modeled under orthographic projection, while retaining linear algebraic properties [56, 57]. Poelman and Kanade [56] have developed

a paraperspective factorization method that can be applied to a much wider range of motion scenarios, including image sequences containing motion toward the camera and aerial image sequences of terrain taken from a low-altitude airplane.

With the development of factorization method, a factorization- based algorithm for multi-image projective structure and motion is developed by Sturm and Triggs [57]. This technique is a practical approach for recovery of scaled feature points, using fundamental matrix and epipoles estimated from the image sequences.

Because matrix factorization is a key component for solving several computer vision problems, Tardif et al. have proposed batch algorithms for matrix factorization [58] that are based on closure and basis constraints, which handle the presence of missing or erroneous data, which often arise in structure from motion.

In mathematical expression of the factorization algorithm, assume that the tracked points are $\{(x_i^j, y_i^j) \mid i = 1, \dots, n; j = 1, \dots, m\}$. The algorithm defines the measurement matrix $\mathbf{W} : \mathbf{W} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}$. The rows of \mathbf{U} and \mathbf{V} are then registered by subtracting from each entry the mean of the entries in that row:

$$\begin{aligned}\bar{x}_i^j &= x_i^j - \frac{1}{n} \sum x_i^j, \\ \bar{y}_i^j &= y_i^j - \frac{1}{n} \sum y_i^j.\end{aligned}\tag{4.1}$$

The goal of the Tomasi-Kanade algorithm [55] is to factorize $\bar{\mathbf{W}}$ into two matrices as follows:

$$\bar{\mathbf{W}} = \mathbf{M}\mathbf{X},\tag{4.2}$$

where \mathbf{M} , named motion matrix, is a $2m \times 3$ matrix which represents the camera rotation in each frame and \mathbf{X} , named structure matrix, is a $3 \times n$ matrix which denotes the positions of the feature points in object space. So in the absence of the Gauss noise, $\text{rank}(\bar{\mathbf{W}}) \leq 3$.

Then we can compute SVD decomposition of $\bar{\mathbf{W}}$ to obtain $\mathbf{U}\mathbf{D}\mathbf{V}^T$:

$$\bar{\mathbf{W}} = \mathbf{U}\mathbf{D}\mathbf{V}^T,\tag{4.3}$$

where if the singular value of $\bar{\mathbf{W}}$ is $[\sigma_1, \sigma_2, \sigma_3]$, we can get the matrix $\mathbf{M} = [\sigma_1\mathbf{u}_1, \sigma_2\mathbf{u}_2, \sigma_3\mathbf{u}_3]$ and $\mathbf{X} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$.

The method can also handle and obtain a full solution from a partially filled-in measurement matrix, which occurs when features appear and disappear in the video due to occlusions or tracking failures [55]. This method gives accurate results and does not introduce smoothing in structure and motion. Using the above method, the problem can be solved for the video of general scene such as building and sculpture (Figure 6).

4.2. Bundle Adjustment

Bundle adjustment is a significant component of most structure from motion systems. It is the joint nonlinear refinement of camera and point parameters, so it can consume a large amount of time for large problems. Unfortunately, the optimization underlying structure from motion

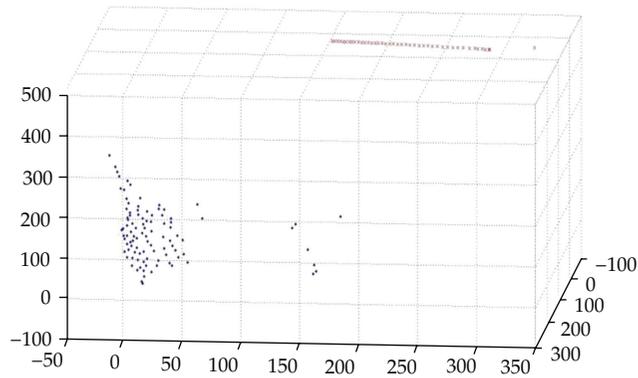


Figure 6: Example of recovering structure and motion.

involves a complex, nonlinear objective function with no closed-form solution, due to nonlinearities in perspective geometry. Most modern approaches use nonlinear least squares algorithms [17] to minimize this objective function, a process known as bundle adjustment; [53] that is, basic mathematics of the bundle adjustment problem is well understood [59]. Generally speaking, bundle adjustment is a global algorithm, but it consumes much time and cannot achieve real time to solve the minimize restriction. Mouragnon et al. [60] propose an approach for generic and real-time structure from motion using local bundle adjustment. It allows 3D points and camera poses to be refined simultaneously through the image sequence. Zhang et al. [61] apply bundle optimization to further improve the results of consistent depth maps from a video sequence.

4.3. Self-Calibration

To upgrade the projective and affine reconstruction to a metric reconstruction (i.e., determined up to an arbitrary Euclidean transformation and a scale factor), calibration techniques, to which we follow the approach described in [2, 6, 9, 15, 62], can deal with this problem. It can be done by imposing some constraints on the intrinsic camera parameters. This approach that is called self-calibration has received a lot of attention in recent years. The ambiguity on the reconstruction is restricted from projective to metric through self-calibration [6]. Mostly self-calibration algorithms are concerned with unknown but constant intrinsic camera parameters [2, 4, 12]. The paper presented the problem of 3D Euclidean reconstruction of structured scenes from uncalibrated images based on the property of vanishing points [63]. They propose a multistage linear approach, with structure from motion technique based on point and vanishing point matches in images [64].

4.4. Correlative Improvement

Traditional SFM algorithms using just two images often produce inaccurate 3D reconstructions, mainly due to incorrect estimation of the camera's motion. Thomas and Oliensis [65] present a practical algorithm that can deal with noise in multiframe structure from motion. It describes a new incremental algorithm for reconstructing structure from multi-image sequences which estimates and corrects for the error in computing the camera motion.

The research of structure from motion has shown great progress throughout several decades, but the algorithms on structure from motion still exhibit some faults and shortages. The result of Structure from Motion cannot satisfy people in many situations. However, many researchers present a lot of improving approaches, such as dual computation of projective shape and camera positions from multiple images [66].

For incremental algorithms that solve progressively larger bundle adjustment problems, Crandall et al. present an alternative formulation for structure from motion based on finding a coarse initial solution using a hybrid discrete-continuous optimization and then improve the solution using bundle adjustment. The initial optimization step uses a discrete Markov random field (MRF) formulation, coupled with a continuous Levenberg-Marquardt refinement [67].

For time efficiency, Havlena et al. present a method of efficient structure from motion by graph optimization [68]. Gherardi et al. improve the algorithm of efficiency with hierarchical structure and motion [69].

For duplicate or similar structure, Roberts et al. couple an expectation maximization (EM) algorithm for structure from motion for scenes with large duplicate structures [70]. A hierarchical framework that resamples 3D reconstructed points to reduce computation cost on time and memory for very-large-scale structure from motion [71]. Savarese and Bao propose a formulation called semantic structure from motion (SSFm), where SSFM takes advantages of both semantic and geometrical properties associated with objects in the scene [72].

5. Relevant Algorithms

5.1. Features

(1) Line

For the problem of camera motion and 3D structure reconstruction from line correspondences across multiple views, there is a triangulation algorithm that outperforms standard linear and bias-corrected quasi-linear algorithms, and that bundle adjustment using our orthonormal representation yields results similar to the standard maximum likelihood trifocal tensor algorithm, while being usable for any number of views [73]. Spetsakis and Aloimonos [74] present a system for structure from motion using line correspondences. The recovery algorithm is formulated in terms of an objective function which measures the total squared distance in the image plane between the observed edge segments and the projections of the reconstructed lines [75]. A linear method is developed for reconstruction using lines and points simultaneously [76].

(2) Curve

Tubic et al. [77] present an approach for reconstructing a surface from a set of arbitrary, unorganized, and intersecting curves. There is an approach for reconstructing open surfaces from image data [78]. Kaminski and Shashua [79] introduce a number of new results in the context of multiview geometry from general algebraic curves, which start with the recovery of camera geometry from matching curves. Berthilsson et al. present a method for reconstruction of general curves, using factorization and bundle adjustment [80].

(3) *Silhouette*

Liang and Wong [81] develop an approach that produces relatively complete 3D models similar to volumetric approaches, with the topology conforming to what is observed from the silhouettes. In addition, the method neither assumes nor depends on the spatial order of viewpoints. Hartley and Kahl give us critical configurations for projective reconstruction from multiple views in [82]. Joshi et al. design an algorithm for structure and motion estimation from dynamic silhouettes under perspective projection [83]. Liu et al. present a method that is shaped from silhouette outlines using an adaptive dandelion model [84]. Yemez and Wetherilt develop a volumetric fusion technique for surface reconstruction from silhouettes and range data [85].

5.2. Other Aspects

(1) *Multiview Stereo*

Multiview stereo (MVS) techniques take as input a set of images with known camera parameters (i.e., position and orientation of the camera, focal length, image distortion parameters) [38, 53, 86]. We can refer to [87] for a classification and evaluation of recent MVS techniques.

(2) *Clustering*

There are clustering techniques to partition the image set into groups of related images, based on the visual structure represented in the image connectivity graph for the collection [88, 89].

6. Existing Problems and Future Trends

While algorithms of structure from motion have been developed for 3D reconstruction in many applications, some problems of reconstructing geometry from video sequences still exist in computer vision and photography. Until recently, however, there have been no good computer vision techniques for recovering this kind of structure from motion. Many researchers are still making efforts to improve the methods mainly in the following aspects.

6.1. Feature Tracking and Matching

Zhang et al. give a robust and efficient algorithm on efficient nonconsecutive feature tracking for structure from motion via two main steps, that is, consecutive point tracking and nonconsecutive track matching [90]. They improve the KTL tracker by the invariant feature points and a two-pass matching strategy to significantly extend the track lifetime and reduce the sensitivity of feature points to variant scale, duplicate and similar structure, and noise and image distortion. The results can be found at <http://www.cad.zju.edu.cn/home/gfzhang/>.

6.2. Active Vision

The method is based on the structure from controlled motion that constrains camera motions to obtain an optimal estimation of the 3D structure of a geometrical primitive [91]. Stereo

geometry is acquired from 3D egomotion streams [92]. Wide-area egomotion estimation is acquired from known 3D structure [93]. A work on estimating surface reflectance properties of a complex scene under captured natural illumination can be found in [94]. Other algorithms are also attempted on selective attention of human eyes.

6.3. Unorganized Images

To solve the resulting large-scale nonlinear optimization, we reconstruct the scene incrementally, starting from a single pair of images, then adding new images and points in rounds, and running a global nonlinear optimization after each round [53]. Structure from motion could be applied to photos found in the wild, reconstructing scenes from several large Internet photo collections [14]. The large redundancy in online photo collections means that a small fraction of images may be sufficient to produce high-quality reconstructions. An investigation has begun to explore by extracting image “skeletons” from large collections [95]. Perhaps the most important challenge is to find ways to effectively parallelize all the steps of the reconstruction pipeline to take advantage of multicore architectures and cloud computing [37, 38, 53, 89].

7. Conclusion

This paper has summarized the recent development of structure from motion algorithm that is able to metrically reconstruct complex scenes and objects. The wide applications have been addressed in computer vision area. Typical contributions are introduced for feature point detection, tracking, matching, factorization, bundle adjustment, multiview stereo, self-calibration, line detection and matching, modeling, and so forth. Representative works are listed for readers to have a general overview of the state of the art. Finally, a summary of existing problems and future trends of structure modeling is addressed.

Acknowledgments

This work was supported by the National Natural Science Foundation of China and Microsoft Research Asia (NSFC nos. 61173096, 60870002, and 60802087), Zhejiang Provincial S&T Department (2010R10006, 2010C33095), and Zhejiang Provincial Natural Science Foundation (R1110679).

References

- [1] H. Kuffar and K. Takaya, “Depth measurement and 3D metric reconstruction from two uncalibrated stereo images,” in *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECD '07)*, pp. 1460–1463, April 2007.
- [2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd edition, 2004.
- [3] O. Faugeras and Q. T. Luong, *The Geometry of Multiple Images*, MIT Press, Cambridge, Mass, USA, 2001.
- [4] R. I. Hartley and P. Sturm, “Triangulation,” in *Proceedings of the American Image Understanding Workshop*, pp. 957–966, 1994.
- [5] P. F. McLauchlan and D. W. Murray, “Unifying framework for structure and motion recovery from image sequences,” in *Proceedings of the 5th International Conference on Computer Vision (ICCV '95)*, pp. 314–320, June 1995.

- [6] M. Pollefeys, L. Van Gool, M. Vergauwen et al., "Visual modeling with a hand-held camera," *International Journal of Computer Vision*, vol. 59, no. 3, pp. 207–232, 2004.
- [7] S. Avidan and A. Shashua, "Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 4, pp. 348–357, 2000.
- [8] N. Molton and M. Brady, "Practical structure and motion from stereo when motion is unconstrained," *International Journal of Computer Vision*, vol. 39, no. 1, pp. 5–23, 2000.
- [9] A. Marugame, J. Katto, and M. Ohta, "Structure recovery with multiple cameras from scaled orthographic and perspective views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 7, pp. 628–633, 1999.
- [10] A. Shashua and N. Navab, "Relative affine structure: canonical model for 3D from 2D geometry and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 873–883, 1996.
- [11] F. Kahl and A. Heyden, "Affine structure and motion from points, lines and conics," *International Journal of Computer Vision*, vol. 33, no. 3, pp. 163–180, 1999.
- [12] A. W. Fitzgibbon and A. Zisserman, "Automatic camera recovery for closed or open image sequences," in *Proceedings of the ECCV*, pp. 311–326, 1998.
- [13] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?,"" in *Proceedings of the IEEE Conference on Computer Vision*, vol. 1, pp. 414–431, 2002.
- [14] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3D," in *Proceedings of the International Conference on Computer Graphics and Interactive Technologies*, pp. 835–846, 2006.
- [15] R. Hartley, "Euclidean reconstruction from uncalibrated views," in *Applications of Invariance in Computer Vision*, J. L. Mundy, A. Zisserman, and D. Forsyth, Eds., Lecture Notes in Computer, 1994.
- [16] A. Bartoli and P. Sturm, "Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene," *International Journal of Computer Vision*, vol. 52, no. 1, pp. 45–64, 2003.
- [17] R. Szeliski and S. B. Kang, "Recovering 3D Shape and Motion from Image Streams Using Nonlinear Least Squares," *Journal of Visual Communication and Image Representation*, vol. 5, no. 1, pp. 10–28, 1994.
- [18] P. M. Q. Aguiar and J. M. F. Moura, "A fast algorithm for rigid structure from image sequences," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '99)*, vol. 3, pp. 125–129, Kobe, Japan, 1999.
- [19] A. Hilton, "Scene modelling from sparse 3D data," *Image and Vision Computing*, vol. 23, no. 10, pp. 900–920, 2005.
- [20] P. Parodi and G. Piccioli, "3D shape reconstruction by using vanishing points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 211–217, 1996.
- [21] M. Lhuillier, "Toward flexible 3D modeling using a catadioptric camera," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1560–1567, June 2007.
- [22] A. Bartoli and S. I. Olsen, "A batch algorithm for implicit non-rigid shape and motion recovery," in *Proceedings of the International Conference on Dynamical Vision*, 2006.
- [23] D. Liebowitz and S. Carlsson, "Uncalibrated motion capture exploiting articulated structure constraints," *International Journal of Computer Vision*, vol. 51, no. 3, pp. 171–187, 2003.
- [24] G. Wang and Q. M. J. Wu, "Stratification approach for 3-D euclidean reconstruction of nonrigid objects from uncalibrated image sequences," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 38, no. 1, pp. 90–101, 2008.
- [25] J. Taylor, A. D. Jepson, and K. N. Kutulakos, "Non-rigid structure from locally-rigid motion," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2761–2768, June 2010.
- [26] S. Zhu, L. Zhang, and B. M. Smith, "Model evolution: an incremental approach to non-rigid structure from motion," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 1165–1172, June 2010.
- [27] M. Farenzena and A. Fusiello, "Stabilizing 3D modeling with geometric constraints propagation," *Computer Vision and Image Understanding*, vol. 113, no. 11, pp. 1147–1157, 2009.
- [28] R. Gong and G. Xu, "3D structure from a single calibrated view using distance constraints," *IEICE Transactions on Information and Systems*, vol. 87, no. 6, pp. 1527–1536, 2004.
- [29] M. Marques and J. Costeira, "Estimating 3D shape from degenerate sequences with missing data," *Computer Vision and Image Understanding*, vol. 113, no. 2, pp. 261–272, 2009.

- [30] T. Brodsky, C. Fermuller, and Y. Aloimonos, "Structure from motion: beyond the epipolar constraint," *International Journal of Computer Vision*, vol. 37, no. 3, pp. 231–258, 2000.
- [31] A. Criminisi, I. Reid, and A. Zisserman, "Single view metrology," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 123–148, 2000.
- [32] H. Joachim, "Direct estimation of structure and motion from multiple zframes," MIT AI Lab. Memo 1190, Massachusetts Institute of Technology, Mass, USA, 190.
- [33] J. Oliensis, "Critique of structure-from-motion algorithms," *Computer Vision and Image Understanding*, vol. 80, no. 2, pp. 172–214, 2000.
- [34] T. Rodriguez, P. Sturm, P. Gargallo et al., "Photorealistic 3D reconstruction from handheld cameras," *Machine Vision and Applications*, vol. 16, no. 4, pp. 246–257, 2005.
- [35] S. N. Sinha, D. Steedly, R. Szeliski, M. Agrawala, and M. Pollefeys, "Interactive 3D architectural modeling from unordered photo collections," *ACM Transactions on Graphics*, vol. 27, no. 5, article 159, 2008.
- [36] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, "Reconstructing building interiors from images," in *Proceedings of the International Conference on Computer Vision*, pp. 80–87, 2009.
- [37] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, "Building Rome in a day," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 72–79, Kyoto, Japan, October 2009.
- [38] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, "Towards internet-scale multi-view stereo," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 1434–1441, San Francisco, Calif, USA, June 2010.
- [39] Y. T. Zheng, M. Zhao, Y. Song et al., "Tour the World: building a web-scale landmark recognition engine," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 1085–1092, June 2009.
- [40] M. Pollefeys et al., "Detailed real-time urban 3D reconstruction from video," *International Journal of Computer Vision*, vol. 78, no. 2-3, pp. 143–167, 2008.
- [41] O. Faugeras, L. Robert, S. Laveau et al., "3-D reconstruction of urban scenes from image sequences," *Computer Vision and Image Understanding*, vol. 69, no. 3, pp. 292–309, 1998.
- [42] N. Cornelis, B. Leibe, K. Cornelis, and L. Van Gool, "3D urban scene modeling integrating recognition and reconstruction," *International Journal of Computer Vision*, vol. 78, no. 2-3, pp. 121–141, 2008.
- [43] C. Collewet and F. Chaumette, "Visual servoing based on structure from controlled motion or on robust statistics," *IEEE Transactions on Robotics*, vol. 24, no. 2, pp. 318–330, 2008.
- [44] M. Marengoni, A. Hanson, S. Zilberstein, and E. Riseman, "Decision making and uncertainty management in a 3D reconstruction system," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 852–858, 2003.
- [45] T. Tung, F. Schmitt, and T. Matsuyama, "Topology matching for 3D video compression," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 2719–2726, June 2007.
- [46] D. Fidaleo and G. Medioni, "Model-assisted 3D face reconstruction from video," in *Proceedings of the 3rd International Workshop on Analysis and Modeling of Faces and Gestures (AMFG '07)*, vol. 4778 of *Lecture Notes in Computer Science*, pp. 124–138, 2007.
- [47] U. Park and A. Jain, "3D model-based face recognition in video," in *Proceedings of the Proceedings International Conference on Advances in Biometrics (ICB '07)*, vol. 4642 of *Lecture Notes in Computer Science*, pp. 1085–1094, 2007.
- [48] V. Nedovic, A. W. M. Smeulders, A. Redert, and J. M. Geusebroek, "Stages as models of scene geometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1673–1687, 2010.
- [49] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.
- [50] C. Tomasi and T. Kanade, "Detection and tracking of point features," Tech. Rep. CMU-91-132, CMU, 1991.
- [51] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 593–600, June 1994.
- [52] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [53] N. Snavely, I. Simon, M. Goesele, R. Szeliski, and S. M. Seitz, "Scene reconstruction and visualization from community photo collections," *Proceedings of the IEEE*, vol. 98, no. 8, Article ID 5483186, pp. 1370–1390, 2010.

- [54] H. C. Longuet-higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, no. 5828, pp. 133–135, 1981.
- [55] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.
- [56] C. J. Poelman and T. Kanade, "A paraperspective factorization method for shape and motion recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 206–218, 1997.
- [57] P. Sturm and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion," in *Proceedings of the 4th European Conference on Computer Vision*, 1996.
- [58] J. P. Tardif, A. Bartoli, M. Trudeau, N. Guilbert, and S. Roy, "Algorithms for batch matrix factorization with application to structure-from-motion," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, June 2007.
- [59] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *Proceedings of the International Workshop on Vision Algorithms*, pp. 298–372, 1999.
- [60] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Generic and real-time structure from motion using local bundle adjustment," *Image and Vision Computing*, vol. 27, no. 8, pp. 1178–1193, 2009.
- [61] G. Zhang, J. Jia, T. T. Wong, and H. Bao, "Consistent depth maps recovery from a video sequence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 974–988, 2009.
- [62] T. Jebara, A. Azarbayejani, and A. Pentland, "3D structure from 2D motion," *IEEE Signal Processing Magazine*, vol. 16, no. 3, pp. 66–84, 1999.
- [63] G. Wang, H. T. Tsui, and Q. M. Jonathan Wu, "What can we learn about the scene structure from three orthogonal vanishing points in images," *Pattern Recognition Letters*, vol. 30, no. 3, pp. 192–202, 2009.
- [64] S. N. Sinha, D. Steedly, and R. Szeliski, "A multi-stage linear approach to structure from motion," in *Proceedings of the European Conference on Computer Vision (ECCV '10)*, 2010.
- [65] J. I. Thomas and J. Oliensis, "Dealing with noise in multiframe structure from motion," *Computer Vision and Image Understanding*, vol. 76, no. 2, pp. 109–124, 1999.
- [66] S. Carlsson and D. Weinshall, "Dual computation of projective shape and camera positions from multiple images," *International Journal of Computer Vision*, vol. 27, no. 3, pp. 227–241, 1998.
- [67] D. Crandall, A. Owens, N. Snavely, and D. Huttenlocher, "Discrete-continuous optimization for large-scale structure from motion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, 2011.
- [68] M. Havlena, A. Torii, and T. Pajdla, "Efficient structure from motion by graph optimization," in *Proceedings of the European Conference on Computer Vision (ECCV '10)*, 2010.
- [69] R. Gherardi, M. Farenzena, and A. Fusello, "Improving the efficiency of hierarchical structure-and-motion," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 1594–1600, June 2010.
- [70] R. Roberts, S. Sinha, R. Szeliski, and D. Steedly, "Structure from motion for scenes with large duplicate structures," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, 2011.
- [71] T. Fang and L. Quan, "Resampling structure from motion," in *Proceedings of the European Conference on Computer Vision (ECCV '10)*, 2010.
- [72] S. Savarese and S. Y. Z. Bao, "Semantic structure from motion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, 2011.
- [73] A. Bartoli and P. Sturm, "Structure-from-motion using lines: representation, triangulation, and bundle adjustment," *Computer Vision and Image Understanding*, vol. 100, no. 3, pp. 416–441, 2005.
- [74] M. E. Spetsakis and J. Aloimonos, "Structure from motion using line correspondences," *International Journal of Computer Vision*, vol. 4, no. 3, pp. 171–183, 1990.
- [75] C. J. Taylor and D. J. Kriegman, "Structure and motion from line segments in multiple images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 11, pp. 1021–1032, 1995.
- [76] R. I. Hartley, "Linear method for reconstruction from lines and points," in *Proceedings of the 5th International Conference on Computer Vision (ICCV '95)*, pp. 882–887, June 1995.
- [77] D. Tubic, P. Hebert, and D. Laurendeau, "3D surface modeling from curves," *Image and Vision Computing*, vol. 22, no. 9, pp. 719–734, 2004.
- [78] J. E. Solem and A. Heyden, "Reconstructing open surfaces from image data," *International Journal of Computer Vision*, vol. 69, no. 3, pp. 267–275, 2006.
- [79] J. Y. Kaminski and A. Shashua, "Multiple view geometry of general algebraic curves," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 195–219, 2004.

- [80] R. Berthilsson, K. Astrom, and A. Heyden, "Reconstruction of general curves, using factorization and bundle adjustment," *International Journal of Computer Vision*, vol. 41, no. 3, pp. 171–182, 2001.
- [81] C. Liang and K. Y. K. Wong, "3D reconstruction using silhouettes from unordered viewpoints," *Image and Vision Computing*, vol. 28, no. 4, pp. 579–589, 2010.
- [82] R. Hartley and F. Kahl, "Critical configurations for projective reconstruction from multiple views," *International Journal of Computer Vision*, vol. 71, no. 1, pp. 5–47, 2007.
- [83] T. Joshi, N. Ahuja, and J. Ponce, "Structure and motion estimation from dynamic silhouettes under perspective projection," *International Journal of Computer Vision*, vol. 31, no. 1, pp. 31–50, 1999.
- [84] X. Liu, H. Yao, and W. Gao, "Shape from silhouette outlines using an adaptive dandelion model," *Computer Vision and Image Understanding*, vol. 105, no. 2, pp. 121–130, 2007.
- [85] Y. Yemez and C. J. Wetherilt, "A volumetric fusion technique for surface reconstruction from silhouettes and range data," *Computer Vision and Image Understanding*, vol. 105, no. 1, pp. 30–41, 2007.
- [86] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from Internet photo collections," *International Journal of Computer Vision*, vol. 80, no. 2, pp. 189–210, 2008.
- [87] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 519–526, June 2006.
- [88] I. Simon, N. Snavely, and S. M. Seitz, "Scene summarization for online image collections," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, October 2007.
- [89] S. Agarwal, Y. Furukawa, N. Snavely, B. Curless, S. M. Seitz, and R. Szeliski, "Reconstructing Rome," *Computer*, vol. 43, no. 6, pp. 40–47, 2010.
- [90] G. Zhang, Z. Dong, J. Jia, T. T. Wong, and H. Bao, "Efficient non-consecutive feature tracking for structure-from-motion," in *Proceedings of the European Conference on Computer Vision (ECCV '10)*, 2010.
- [91] E. Marchand and F. Chaumette, "Active vision for complete scene reconstruction and exploration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 1, pp. 65–72, 1999.
- [92] F. Dornaika and C. K. R. Chung, "Stereo geometry from 3-D ego-motion streams," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 33, no. 2, pp. 308–323, 2003.
- [93] O. Koch and S. Teller, "Wide-area egomotion estimation from known 3D structure," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 437–444, June 2007.
- [94] P. Debevec, C. Tchou et al., "Estimating surface reflectance properties of a complex scene under captured natural illumination," Tech. Rep. ICT-TR-06.2004, University of Southern California Institute for Creative Technologies, Marina del Rey, Calif, USA, 2004.
- [95] N. Snavely, S. M. Seitz, and R. Szeliski, "Skeletal graphs for efficient structure from motion," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, June 2008.

Research Article

Advanced Signal Processing and Command Synthesis for Memory-Limited Complex Systems

Cristian Toma

Faculty of Applied Sciences, Politehnica University, Hagi-Ghita 81, 060032 Bucharest, Romania

Correspondence should be addressed to Cristian Toma, cgtoma@physics.pub.ro

Received 30 June 2011; Revised 3 September 2011; Accepted 3 September 2011

Academic Editor: Carlo Cattani

Copyright © 2012 Cristian Toma. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents advanced signal processing methods and command synthesis for memory-limited complex systems. For accurate measurements performed on limited time interval, some specific methods should be added. For signal processing, a robust filtering and sampling procedure performed on a specific working interval is required, so as the influence of low-amplitude and high-frequency fluctuations to be diminished. This study shows that such a signal processing method for the case of memory-limited complex systems requires the use of certain differentiation/integration procedures performed by oscillating systems, so as robust results suitable for efficient command synthesis to be available. A brief comparison with uncertainty aspects in modern physics (where quantum aspects can be considered as features of complex systems) is also presented.

1. Introduction

As it is known, an important aspect in observing and modeling dynamic environmental phenomena consists in measuring with higher accuracy some physical quantities corresponding to changes in the environment. Yet for accurate measurements performed on limited time interval, for memory-limited complex systems, some specific methods should be used. Sudden (sharp) changes in the environment require a pair of consecutive values for the measured quantity so as any difference to be detected as soon as possible. Moreover, any value taken into consideration by the complex system should be established using a robust filtering and sampling procedure performed on a specific working interval, so as the influence of low-amplitude and high-frequency fluctuations to be decreased in a significant manner. Being quite possible for sharp (sudden) changes in the environment to appear during such a working interval (on which filtering and sampling procedures are performed), it results that

specific signal processing methods based on the values achieved on a set of successive time intervals are necessary.

Filtering and sampling devices usually consist of asymptotically stable systems, sometimes an integration of the output over a certain time interval being added. Yet such structures are very sensitive at random variations of the integration period, being recommended for the signal which is integrated to be approximately equal to zero at the end of the integration period. For this reason, oscillating systems for filtering the received signal should be used, so as the filtered signal and its slope to be approximately zero at the end of a certain time interval (at the end of an oscillation). For avoiding instability of such oscillating systems on extended time intervals, certain electronic devices (gates) controlled by computer commands should be added, so as to restore the initial null conditions for the oscillating system before a new working cycle to start [1].

The filtering performances of asymptotically stable systems are determined by their transfer function. a Filtering and sampling devices consisting of low-pass filters of first or second order having the transfer function

$$H(s) = \frac{1}{T_0s + 1} \quad (1.1)$$

(for a first-order system) and

$$H(s) = \frac{1}{T_0^2s^2 + 2bT_0s + 1} \quad (1.2)$$

(for a second-order system) attenuate an alternating signal of angular frequency $\omega \gg \omega_0 = 1/T_0$ about ω/ω_0 times (for a first-order system) or about $(\omega/\omega_0)^2$ times (for a second-order system). The response time of such systems at a continuous useful signal is about $4 - 6T_0$ ($5T_0$ for the first-order system and $4T_0/b$ for the second-order system). If the signal given by the first- or second-order system is integrated over such a period, a supplementary attenuation for the alternating signal of about $4 - 6\omega/\omega_0$ can be obtained.

But such structures are very sensitive at the random variations of the integration period (for unity-step input, the signal which is integrated is equal to unity at the sampling moment of time), and the use of oscillators with a very high accuracy cannot solve the problem due to switching phenomena appearing at the end of the integration period (when an electric current charging a capacitor is interrupted).

These random variations cannot be avoided if we use asymptotically stable filters. For robustness, the signal processing structure based on an integration procedure should provide a null value for the integrating signal at the end of a certain working interval. This property is similar to wavelets aspects presented in [2, 3].

Mathematically, an ideal solution could consist in using an extended Dirac function for multiplying the received signal before the integration (see [1]) but is very hard to generate such extended Dirac functions (a kind of acausal pulses) using nonlinear differential equations for (i) symmetrical pulses (see [4]) or (ii) asymmetrical pulses (see [5] for more details).

A heuristic algorithm for generating practical test functions using MATLAB procedures was presented in [4]. First, it has been shown that ideal test functions cannot be generated by differential equations, being emphasized the fact that differential equations can

only generate functions similar to test functions (defined as practical test functions). Then a step-by-step algorithm for designing the most simple differential equation able to generate a practical test function was presented, based on the invariance properties of the differential equation and on standard MATLAB procedures. The result of the algorithm consists in an oscillating second-order system working at the stability limit from initial null conditions, on a limited working interval corresponding to the period of the generated oscillations.

It was shown that the simplest structure possessing such properties is represented by an oscillating second-order system having the transfer function

$$H_{\text{osc}} = \frac{1}{T_0^2 s^2 + 1}, \quad (1.3)$$

receiving a step input and working on the time interval $[0, 2\pi T_0]$. For initial conditions equal to zero, the response of the oscillating system at a step input with amplitude A will have the form

$$y(t) = A \left(1 - \cos\left(\frac{t}{T_0}\right) \right). \quad (1.4)$$

By integrating this result on the time interval $[0, 2\pi T_0]$, we obtain the result $2\pi A T_0$, and we can also notice that the quantity which is integrated and its slope are equal to zero at the end of the integration period. Thus, the influence of the random variations of the integration period (generated by the switching phenomena) is practically rejected.

This oscillating system attenuates about $(\omega/\omega_0)^2$ times such an input, and the influence of the integrator consists in a supplementary attenuation of about

$$\left[\frac{1}{(2\pi)} \left(\frac{\omega}{\omega_0} \right) \right] \quad (1.5)$$

times. The oscillations having the form

$$y_{\text{osc}} = a \sin(\omega_0 t) + b \cos(\omega_0 t) \quad (1.6)$$

generated by the input alternating component have a lower amplitude and give a null result after an integration over the time interval $[0, 2\pi T_0]$.

These results have shown that such a structure provides practically the same performances as a structure consisting of an asymptotically stable second-order system and an integrator (response time of about $6T_0$, an attenuation of about $(1/6)(\omega/\omega_0)^3$ times for an alternating component having frequency ω) moreover being less sensitive at the random variations of the integration period. For restoring the initial null conditions after the sampling procedure (at the end of the working period), some electronic devices must be added. Yet the previous analysis is valid for step inputs which are active on the whole working interval (the integration period).

In [6] has been performed the analysis of this structure by considering that the input is represented by a unity short-step pulse (instead of a unity step pulse) which differs to

zero on the time interval $[0, \tau]$. It was shown that certain free oscillations of the second-order oscillating system are generated for $t > \tau$ (when the action of the external short-step command u has ceased). These free oscillations have the angular velocity ω_0 , the amplitude

$$A = 2 \sin\left(\frac{\omega_0 \tau}{2}\right), \quad (1.7)$$

and the initial phase

$$\phi = -\frac{\omega_0 \tau}{2}. \quad (1.8)$$

Thus, the output $y(t)$ corresponding to the free oscillations of the system for $t > \tau$ (when the action of the external short-step command u has ceased) can be written as

$$y(t) = 2 \sin\left(\frac{\omega_0 \tau}{2}\right) \sin\left(\omega_0 t - \frac{\omega_0 \tau}{2}\right). \quad (1.9)$$

However, we must notice that, usually, such a filtering and sampling structure receives an electronic signal presenting possible step changes from an already measured value to a final unknown value. Since the previously measured value can be subtracted from the received signal during subsequent working intervals, the analysis of sudden (sharp) changes in the environment could start by considering that the input of the second-order oscillating system is represented by a null signal for $t \leq \tau$ (the first part of the working interval) and by a signal with amplitude A for $t > \tau$ (the second part of the working interval).

2. The Oscillating Signal Processing System for the Case of Short-Step Inputs

We will continue the analysis of this structure by considering that the input is represented by a short-step pulse which differs to zero on the time interval $[\tau, 2\pi T_0]$. This means that the input u can be represented under the form

$$\begin{aligned} u(t) &= 0, & \text{for } t < \tau, \\ u(t) &= A, & \text{for } t \in [\tau, 2\pi T_0], \end{aligned} \quad (2.1)$$

or using the Heaviside function

$$u(t) = Ah(t - \tau) \quad \text{for } t \in [0, \infty), \quad (2.2)$$

where $h(t)$ corresponds to the function $1/s$ if we apply the Laplace transformation.

The transfer function of the second-order oscillating system is

$$H(s) = \frac{1}{T_0^2 s^2 + 1}. \quad (2.3)$$

On the time interval $[0, \tau]$, the output of the second-order oscillating system equals zero

$$y(t) = 0. \quad (2.4)$$

On the time interval $[\tau, 2\pi T_0]$, the output of the second-order oscillating system is represented (using the Laplace transformation) as

$$y(s) = H(s)u(s) = \frac{1}{T_0^2 s^2 + 1} \frac{A}{s} \exp(-\tau s), \quad (2.5)$$

which corresponds to the output

$$y(t) = A \left(1 - \cos \left(\frac{t - \tau}{T_0} \right) \right), \quad (2.6)$$

which can be written as

$$y(t) = A[1 - \cos(\omega_0(t - \tau))], \quad (2.7)$$

where $\omega_0 = 1/T_0$.

By denoting with $z(t)$ the integral of $y(t)$ (considering as initial moment the zero moment of time), it results at the time moment $t_f = 2\pi T_0$ the set of values:

$$\begin{aligned} y(t_f) &= A[1 - \cos(\omega_0(2\pi T_0 - \tau))] = A[1 - \cos(\omega_0\tau)], \\ y'(t_f) &= A\omega_0 \sin(\omega_0(2\pi T_0 - \tau)) = -A\omega_0 \sin(\omega_0\tau), \\ z(t_f) &= \int_{\tau}^{2\pi T_0} A[1 - \cos(\omega_0(t - \tau))] dt = A \left(\frac{2\pi}{\omega_0} - \tau \right) + \frac{A}{\omega_0} \sin(\omega_0\tau). \end{aligned} \quad (2.8)$$

It can be easily noticed that

$$\omega_0 z(t_f) + \frac{1}{\omega_0} y'(t_f) = A(2\pi - \omega_0\tau), \quad (2.9)$$

which can be written also as

$$z(t_f) + \frac{1}{\omega_0^2} y'(t_f) = A(2\pi T_0 - \tau). \quad (2.10)$$

This result shows that the sampled values for $z(t)$ (the integral of $y(t)$) and for $y'(t)$ at the time moment $t_f = 2\pi T_0$ (the end of the working interval) can be used in a simple manner for obtaining the quantity

$$S(t_f) = A(2\pi T_0 - \tau) = At_A. \quad (2.11)$$

For this purpose, we can divide the sampled value for $y'(t)$ by $1/\omega_0^2$ and add this result to the sampled value for $z(t)$ (the integral of $y(t)$). All these operations can be performed electronically (using analog devices) in an accurate manner. It can be easily noticed that the quantity

$$t_A = 2\pi T_0 - \tau \quad (2.12)$$

represents the active time (on which the step input A acts upon the second-order oscillating system).

On the subsequent working interval, we can consider that the input of the second-order oscillating system equals A on the whole time interval $[0, 2\pi T_0]$. As a consequence, the integral of the generated output equals

$$z_{\text{next}} = 2\pi T_0 A, \quad (2.13)$$

which allows a robust estimation of the amplitude A of the step change for the input as

$$A = \frac{z_{\text{next}}}{2\pi T_0}, \quad (2.14)$$

where z_{next} is the sampled quantity for the integral over a period for the oscillating system output (starting to work from initial null conditions on the next working interval), and $2\pi T_0$ is a constant value. This operation can be also performed electronically in an easy manner.

However, this result is far of being useful for practical applications. Since the differential equation of the second-order oscillating system is

$$y(t) + \left(\frac{1}{\omega_0^2} \right) y''(t) = u(t), \quad (2.15)$$

it results that

$$S(t_f) = \int_0^{t_f} \left[y(t) + \left(\frac{1}{\omega_0^2} \right) y''(t) \right] dt = \int_0^{t_f} u(t) dt, \quad (2.16)$$

where $u(t)$ represents the amplitude of the received signal. So the algorithm previously presented performs the integral of the received signal on a period, without any filtering procedure. The influence of low-amplitude high-frequency alternating components of the received signal is not diminished in a significant manner, the advantages of a filtering procedure based on second-order systems being lost.

The previously presented algorithm could be accepted if the integral can be performed on an extended time interval. In this case, we can simply estimate the quantity $S(t_f)$ on a time period T_F several times greater than $2\pi T_0$ (the period of the oscillating system) so as requirements regarding filtering performances (for rejecting the influence of low-frequency high-amplitude components) to be fulfilled. Thus, the influence of an alternating component with time constant T and angular frequency $\omega = 2\pi/T$ is decreased about $\omega T_F = T_F/T$ times.

The sampling procedure is not robust any more (the signal which is integrated differs to zero at the end of the integration interval, in case of step changes) but for extended intervals, the relative error generated by switching phenomena can be neglected (the switching interval is very narrow as related to the integration interval, so the integral of stochastic switching phenomena represents a low value as related to the integral performed on the whole working interval).

3. Synthesis of Quick Command for Compensating Dynamic Environmental Changes

In case of electric drives, an extended integration interval for quantity $u(t)$ (as presented in the second section of this paper) could be allowed if it corresponds not to the controlled quantity (the angular frequency Ω of a shaft, for example) but to its derivative (supposing that the resistive torque can vary). As a consequence, $u(t)$ can be written as

$$u(t) = c \frac{d\Omega}{dt} \quad (3.1)$$

(c being a constant). It results that

$$S(t_f) = \int_0^{t_f} u(t) dt = \int_0^{t_f} c \frac{d\Omega}{dt} dt = c(\Omega(t_f) - \Omega(0)), \quad (3.2)$$

so it corresponds to the variation of the angular frequency Ω on the time interval T_F which should be adjusted by a supplementary active torque M_s represented by

$$M_s = M_a - M_r, \quad (3.3)$$

where M_a corresponds to the active torque and M_r corresponds to the resistive torque. The active and resistive torque should be equal in the stationary regime, so the supplementary torque should act on a limited time interval, so as to compensate the difference $\Omega(t_f) - \Omega(0)$ previously detected.

For this purpose, we could notice that quantity $S(t_f)$ is proportional to this difference, so a supplementary active torque can be transmitted in a limited time interval ($t_f, t_f + T_s$) as

$$M_s(t) = b \left(1 - \cos \left(2\pi \frac{t - t_f}{T_s} \right) \right) S(t_f) \quad (3.4)$$

(b being a constant). When $t = t_f + T_s$, both function $M_s(t)$ and its derivative $C'(t)$ have null values, so the action of the supplementary torque can be stopped avoiding the influence of any switching phenomena (which can generate errors).

Since

$$M_s = M_a - M_r = c \frac{d\Omega}{dt}, \quad (3.5)$$

it results that the action of this supplementary torque on this time interval T_s can be represented as

$$\Omega(t_f + T_s) - \Omega(t_f) = \frac{1}{c} \int_{t_f}^{t_f+T_s} M_s(t) dt = \frac{b}{c} \int_{t_f}^{t_f+T_s} \left(1 - \cos\left(2\pi \frac{t - t_f}{T_s}\right)\right) S(t_f) dt. \quad (3.6)$$

While $S(t_f)$ is proportional to the difference $\Omega(t_f) - \Omega(0)$ (as has been shown), it finally results that

$$\Omega(t_f + T_s) - \Omega(t_f) = bT_s(\Omega(t_f) - \Omega(0)). \quad (3.7)$$

By adjusting the relation between b and T_s , we can set quantity bT_s to unity, so as the action of the step change in the environment upon the angular velocity to be compensated in a subsequent finite time interval.

This intuitive model is also valid for any complex (biological) system which should maintain its position or the velocity of certain components at a specific value.

4. Efficient Signal Processing Methods Based on Two State Variables

The signal processing method presented in previous paragraph is based on sampled values for three successive working intervals of the second-order oscillating system. Considering that a step change for the input is detected on a certain working interval, the previous value for the input is determined on an initial working interval (on which the input equals a certain value A_{in}), and the final value for the input is determined on a final working interval (when the input equals a final value A_{next}). During the middle working interval, the quantity A_{in} is subtracted from the received signal, and the quantity

$$At_A = (A_{next} - A_{in})(2\pi T_0 - \tau) = z(t_f) + \frac{1}{\omega_0^2} y'(t_f) \quad (4.1)$$

is available at the end of the interval (using amplifying and sampling procedures). Finally, the step change $A = A_{next} - A_{in}$ and the time moment τ are determined (using subtracting and dividing procedures).

However, the linearity of this second-order oscillating system allows a more efficient and robust procedure to be used. The algorithm presented in the previous paragraph requires two identical oscillating second-order systems working at the same time: one for processing the input (so as to determine the estimated values for an input considered to be constant on that interval) and another for processing the difference between the received signal and the previously sampled value (so as to detect a possible step change during this interval by determining the quantity At_A , where A stands for the step change and t_A stands for the active time). Moreover, filtering aspects could require extended time intervals for processing the input signal (as was shown), and the absence of any control action during such an interval could allow significant changes for the output of the complex system from the desired value.

In previous paragraph, it was shown that filtering properties and robustness require an extended time interval for processing the input signal (received from transducers). For this purpose, we can use either a second-order oscillating system with a period equal to

the working interval (which means $2\pi T_0 = T_F$), or a second-order oscillating system with a period corresponding to a submultiple of the working interval (which means $2\pi T_0 = T_F/N$). The last choice is far more convenient, as it will be shown in this paragraph. Let us suppose that on the whole working interval the input $u(t)$ is represented just by the constant value A_{in} . In this case, the output $y(t)$ is represented by $A_{\text{in}}(1 - \cos(\omega_0 t))$. Supposing that quantity A_{in} was determined on previous working interval, we can determine in a very simple manner the expected values $y_{\text{ex}(k)}$ which would be sampled for $y(t)$ when phase ϕ equals $k\pi$ (this means $0, 2A_{\text{in}}, 0, -2A_{\text{in}}, \dots$, the effect of high-frequency low-amplitude fluctuations being neglected). At these time moments quantity $y'(t)$ would be zero, so the sampling procedure would be robust as related to the constant input A_{in} .

This suggests the possibility of sampling $y(t)$ at these time moments (this means when $\omega_0 t$ equals $k\pi$) and subtracting the expected values previously presented. Due to the linearity of the second-order oscillating system, if a step change with amplitude A is detected during the working interval, the sampled values $y_{(k)}$ at the moments of time $t_k = k\pi/\omega_0$ will be represented by a sum of values determined by step input A_{in} and by a step change A starting to act at moment τ . Thus, the result is

$$y_{A(k)} = y_{(k)} - y_{\text{ex}(k)}. \quad (4.2)$$

Thus, the successive values $y_{A(k)}$ of the subtracting procedure will correspond just to the influence of the step change A .

A quick analysis for first pair of values for $y_{A(1)}$ and $y_{A(2)}$ after the step change starts to act (when significant differences from expected values are detected) can be performed by comparing quantities

$$\begin{aligned} y_{A(1)} &= A[1 - \cos(\omega_0(t_1 - \tau))] = A[1 - \cos \phi_1] = 2A \sin^2 \frac{\phi_1}{2}, \\ y_{A(2)} &= A[1 - \cos(\omega_0(t_2 - \tau))] = A[1 - \cos \phi_2] = 2A \sin^2 \frac{\phi_2}{2}, \end{aligned} \quad (4.3)$$

where

$$\begin{aligned} \phi_1 &= \omega_0(t_1 - \tau), \\ \phi_2 &= \omega_0(t_2 - \tau). \end{aligned} \quad (4.4)$$

The phase difference between ϕ_1 and ϕ_2 equals π (as it was shown). It results that

$$y_{A(1)} = 2A \sin^2 \frac{\phi_1}{2}, \quad y_{A(2)} = 2A \sin^2 \frac{\phi_1 + \pi}{2} = 2A \cos^2 \frac{\phi_1}{2}. \quad (4.5)$$

It can be easily noticed that the ratio $y_{A(1)}/y_{A(2)}$ can be written as

$$\frac{y_{A(1)}}{y_{A(2)}} = \tan^2 \frac{\phi_1}{2}. \quad (4.6)$$

A quick comparison of $y_{A(1)}$ and $y_{A(2)}$ is useful for an approximation of ϕ_1 . For high values for $y_{A(1)}$ and low values for $y_{A(2)}$, $\phi_1/2$ can be approximated as $\pi/2$, and $2A$ can be approximated as $y_{A(1)}$. For low values for $y_{A(1)}$ and high values for $y_{A(2)}$, $\phi_1/2$ can be approximated as 0 and $2A$ can be approximated as $y_{A(2)}$. For similar values for $y_{A(1)}$ and $y_{A(2)}$, $\phi_1/2$ can be approximated as $\pi/4$, and $2A$ can be approximated as $y_{A(1)}/2 = y_{A(2)}/2$. Using ϕ_1 , τ can be determined as

$$\tau = \omega_0(t_1 - \phi_1), \quad (4.7)$$

where t_1 stands for the first time moment in the set determined by $\omega_0 t_k = k\pi, k \in Z$, for which significant differences from expected values $y_{ex(k)}$ are detected for sampled values y_k . This algorithm is efficient because it allows a preliminary command to be transmitted to the controlled system before the working interval (on which the received signal is integrated and sampled) to come to an end. It is more effective if the working interval includes several periods of the second-order oscillating system (thus, more moments for estimative sampling are established inside the working interval, and the preliminary command is transmitted faster).

5. Aspects Connected with the Uncertainty Principle

In the third section of this paper has been presented an algorithm for a preliminary estimation of a step change in the environment based on two state variables (two successive values for a $\sin^2\phi$ function sampled at two successive time moments) with the phase difference corresponding to these time moments determined as

$$\Delta\phi = \phi_2 - \phi_1 = \frac{\pi}{2}. \quad (5.1)$$

It can be noticed that measurements performed for sinusoidal functions at time moments when the phase difference equals $\pi/2$ imply some aspects regarding opposite requirements for sampling moments of time. Let us suppose that we are sampling just a sine function for phase

$$\begin{aligned} \phi_1 &= \psi, \\ \phi_2 &= \psi + \frac{\pi}{2}. \end{aligned} \quad (5.2)$$

It results that sampled values correspond to the pair

$$\sin \psi, \quad \cos \psi. \quad (5.3)$$

This pair can be considered also as a value for a sine function and a value for its derivative for the same phase ψ . By taking into account the fact that any sampling procedure requires a nonzero time interval, it results that it is desirable for these functions $\sin \psi, \cos \psi$ to be almost constant on a very short time interval necessary for this sampling procedure.

Mathematically, this would imply the necessity of both functions to have null values at sampling moments of time. However, it is well known that a maximum/minimum value for sine function (when the differential of sine function equals zero) corresponds to a null value for its derivative, the cosine function (at these time moments, the differential of the cosine function has the greatest value for its slope, considered as modulus). Quite similar, a maximum/minimum value for cosine function (when the differential of cosine function equals zero) corresponds to a null value for its derivative, the $-\sin$ function (at these time moments, the differential of the sine function has the greatest value for its slope, considered as modulus). So it is impossible to select a suitable phase ψ so as the measurement accuracy for both sine and cosine functions to be the best possible if nonzero sampling intervals are taken into account.

This aspect is similar to the uncertainty principle in physics, when measurements corresponding to a certain physical variable and to its conjugated variable (usually corresponding to a derivative of a function in respect to the previous variable) are involved. Moreover, it should be noticed that for a cosine function

$$f = A \cos \psi = \operatorname{Re} A \exp(i\psi), \quad (5.4)$$

the action of the operator

$$\operatorname{Cr} = 1 - i \frac{d}{d\psi} \quad (5.5)$$

would correspond to

$$\operatorname{Cr}(f) = \left(1 - i \frac{d}{d\psi}\right) f = \operatorname{Re} \left[A \exp(i\psi) - (i)^2 A \exp(i\psi) \right] = \operatorname{Re} [2A \exp(i\psi)]. \quad (5.6)$$

It can be noticed that another cosine function $\cos \psi$ has been generated.

Quite similar, the action of the operator

$$\operatorname{An} = 1 + i \frac{d}{d\psi} \quad (5.7)$$

would correspond to

$$\operatorname{An}(f) = \left(1 + i \frac{d}{d\psi}\right) f = \operatorname{Re} \left[A \exp(i\psi) + (i)^2 A \exp(i\psi) \right] = 0. \quad (5.8)$$

It can be noticed that the cosine function $\cos \psi$ has been annihilated. This aspect is similar to creation/annihilation of particles in advanced quantum mechanics, where such operators are derived using the decomposition of certain fields in plane waves. However, aspects connected with momentum and position operators are hard to be noticed at this stage of research.

As a conclusion for memory-limited complex systems, we can notice that the use of second-order oscillating systems allows of just robust sampling procedures on extended time

intervals for certain quantities corresponding to step changes in the environment, but also the use of just two state variables (corresponding to sampled values for sine and cosine function at certain moments of time) for a preliminary estimation of such step changes during the working interval. For an extended working interval which includes several alternances of the oscillations generated by the received signal, this implies the possibility of transmitting quick preliminary commands towards the actuators, a final adjustment being determined at the end of the whole working interval, based on the difference between required action and the already-performed action. Similar to aspects presented in previous section, these preliminary commands $\text{Com}(t)$ should be better transmitted as a pulse defined by

$$\text{Com}(t) = C[1 - \cos(\omega_0(t - t_2))], \quad (5.9)$$

for $t \in (t_2, t_2 + 2\pi T_0)$ (a period of the oscillating signal considered from the second moment of time when the preliminary sampling procedure for $y(t)$ was performed, and the pair of values sampled at t_1 and t_2 has been analyzed). Thus, subsequent expected values for $y(t)$ and $y'(t)$ at time moments determined by $\omega_0 t = k\pi$ could be computed in an easy manner by the signal processing device, and any differences could be analyzed (for a preliminary conclusion) in an easy manner on next periods of the oscillating system until the working interval comes to an end.

For this reason, the algorithm previously presented is suitable for memory-limited complex systems since it performs both a preliminary analysis of signal received from environment for detecting step changes (with preliminary commands transmitted towards actuators) and a final accurate estimation for the required action on next extended time intervals (computed as a difference between the whole action required by the step change and the action already performed by preliminary commands).

As in case of biological systems, this algorithm is based on values sampled at some successive moments of time. It generates a sequence of certain commands towards the environment as a sequence of pulses, analyzes the difference between expected values and real values for the signal received from the environment, and adjusts the command with higher accuracy after an extended time interval (a kind of multilevel control and command). Another important similarity between this algorithm and behaviour of biological systems should be noticed; the sampling moments of time (when the processed signal is recorded) differ to the time moments when the filtered (processed) signal has a great slope (considered as modulus) so as to allow a robust estimation using just two sampled (recorded) values. The fact that less memory is involved is essential for complex systems which have to survey a great number of parameters (motion parameters, for instance) in the environment and to check the effect of commands transmitted towards a great number of actuators, see the case of vision processing studied in [7, 8].

6. Conclusions

This study has presented advanced signal processing methods and command synthesis for memory-limited complex systems. It was shown that for observing, modeling, and controlling dynamic environmental phenomena in case of memory-limited complex systems, some specific methods based on accurate measurements performed on limited time intervals are required. Starting from the necessity of a set of consecutive measurements performed in a robust manner for detecting step changes in the environment, it was shown that an

extended time interval for processing the input signal is necessary. For this reason, the use of second-order oscillating systems was improved by adding a supplementary algorithm so as preliminary values for step changes in the environment to be available for control and command during the signal processing interval. A method for generating a robust command towards the control equipment on a limited time interval in order to compensate the step changes detected on previous working interval was also presented. Finally, similarities between measurements of a certain quantity and of its derivative for a sine function, by one side, and the uncertainty principle in physics (by the other side) were briefly mentioned.

References

- [1] C. Toma, "An extension of the notion of observability at filtering and sampling devices," in *Proceedings of the International Symposium on Signals, Circuits and Systems Iasi (SCS '01)*, p. 233, Iasi, Romania, July 2001.
- [2] C. Cattani and J. 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, Toh Tuck Link, Singapore, 2007.
- [3] C. Cattani, "Connection coefficients of Shannon wavelets," *Mathematical Modelling and Analysis*, vol. 11, no. 2, pp. 117–132, 2006.
- [4] G. Toma, "Practical test-functions generated by computer algorithms," *Lecture Notes Computer Science*, vol. 3482, pp. 576–584, 2005.
- [5] S. Pusca, "Invariance properties of practical test-functions used for generating asymmetrical pulses," *Lecture Notes Computer Science*, vol. 3980, pp. 763–770, 2006.
- [6] A. Toma and C. Morarescu, "Detection of short-step pulses using practical test-functions and resonance aspects," *Mathematical Problems in Engineering*, vol. 2008, Article ID 543457, 15 pages, 2008.
- [7] S. Y. Chen and Y. F. Li, "Vision sensor planning for 3-D model acquisition," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 35, no. 5, pp. 894–904, 2005.
- [8] S. Y. Chen, Y. F. Li, and J. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.

Research Article

A Method for Designing Assembly Tolerance Networks of Mechanical Assemblies

**Yi Zhang,^{1,2} Zongbin Li,^{1,2} Jianmin Gao,^{1,2} Jun Hong,^{1,2}
Francesco Vilecco,³ and Yunlong Li^{1,2}**

¹ School of Mechanical Engineering, Xi'an Jiaotong University, Shaanxi 710049, China

² State Key Laboratory for Manufacturing Systems Engineering of China, Shaanxi 710049, China

³ Department of Industrial Engineering, University of Salerno, Via Ponte Don Melillo,
84084 Fisciano, Italy

Correspondence should be addressed to Yi Zhang, zhangyi.jd@stu.xjtu.edu.cn

Received 16 May 2011; Accepted 16 July 2011

Academic Editor: Carlo Cattani

Copyright © 2012 Yi Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

When designing mechanical assemblies, assembly tolerance design is an important issue which must be seriously considered by designers. Assembly tolerances reflect functional requirements of assembling, which can be used to control assembling qualities and production costs. This paper proposes a new method for designing assembly tolerance networks of mechanical assemblies. The method establishes the assembly structure tree model of an assembly based on its product structure tree model. On this basis, assembly information model and assembly relation model are set up based on polychromatic sets (PS) theory. According to the two models, the systems of location relation equations and interference relation equations are established. Then, using methods of topologically related surfaces (TTRS) theory and variational geometric constraints (VGC) theory, three VGC reasoning matrices are constructed. According to corresponding relations between VGCs and assembly tolerance types, the reasoning matrices of tolerance types are also established by using contour matrices of PS. Finally, an exemplary product is used to construct its assembly tolerance networks and meanwhile to verify the feasibility and effectiveness of the proposed method.

1. Introduction

Assembly tolerance design is one of the research hotspots in the field of computer-aided tolerancing (CAT). Assembly constraints are essentially constraints between assembly feature surfaces of parts. Assembly tolerance is a vector constraint. Its orientation, type, and value are mainly determined by the assembly functional requirements of mechanical assemblies. Assembly tolerances are crucial for many activities in the product's life cycle, which not only affects assembly qualities but also determines manufacturing costs. Automatic generation of assembly tolerance networks can greatly reduce design complexity and improve design

quality. Meanwhile, optimization design of assembly tolerances can remarkably reduce manufacturing cost of mechanical assemblies. The design of assembly tolerance networks is a complex multiscale problem. It involves associations between the multiple scales, such as the assembly functional requirement, part positioning, datum reference frame, assembly sequence, assembly feature, and tolerance specification. The essences of many important issues are multiscale problems in medicine, physics, computers, chemistry, materials science, robotics, and other disciplines. Multiscale modeling and operations are widely used in these research fields. Kou et al. [1] used multiscale time schemes for simulating two-phase flow in fractured porous media. The method divides the total time interval into four temporal levels, which can take a large time-step size to achieve stable solutions. Picard et al. [2] introduced a new approach to model realistically sounding objects for animated real-time virtual environments. The number of hexahedral finite elements is only loosely determined by the geometry of the sounding object. Therefore, the approach can realize a multiscale solution. In [3], a multiscale time-frequency representation and the Stockwell transform were used to generalize the time-varying spectra and coherence, which makes the Stockwell transform an effective approach to investigate the characteristics of the spectrum and the interaction of locally stationary time series. The multiscale curvelet shrinkage and the total variation function are combined together, by which a new variational image model is established to compute an initial estimated image [4]. Bakhoun and Toma [5] established a higher-order differential equation to model multiscale phenomena and explain multiscale threshold transitions. Different delayed pulses and multiscale behaviour can be noticed when the order of the equation is higher. According to the multiscale continuous wavelet transform, Djilas et al. [6] presented an algorithm to separate activity of primary and secondary muscle spindle afferents. In [7], Chen and Beghdadi proposed a new algorithm for natural enhancement of color image by using the multiscale Retinex model, which can improve the luminance and chrominance contrast of image. Chen et al. [8] proposed a differential geometry-based multiscale framework to handle complex biological systems, which can deal with microscopic fluid dynamics, microscopic molecular dynamics, and surface dynamics in a unified framework. In [9], a multiscale linear system of algebraic equations is established to determine material motion and gradient velocity, which is adapted to the situation of tagging MRI data. M. N. Nounou and H. N. Nounou [10] developed a multiscale nonlinear modeling algorithm based on multiscale wavelet-based representation of data. The approach can enhance the estimation accuracy of the linear-in-the-parameters nonlinear model. By applying the same method in [10], Nounou and Nounou [11] presented a multiscale latent variable regression modeling algorithm to improve the prediction accuracy of some of the models, such as Principal Component Regression and Partial Least Squares. Lucia [12] developed a new multiscale methodology for cubic equations of state, which constraints the attraction parameter by using the Gibbs-Helmholtz equations.

In recent years, research on assembly tolerance design has made plentiful and substantial achievements. Based on TTRS theory, Clement et al. [13, 14] proposed a method to model and represent dimensional and geometric tolerances in CAT systems. Hoffmann [15] established a new model in three-dimensional space, in which geometric graph is regarded as a set of vector dots and tolerances are described by using tolerance functions in which point vectors are taken as parameters. Desrochers and Clément [16] provided a representation method of tolerance information based on TTRS theory, which is independent of modeling systems. In one-dimensional space, Wang and Ozsoy [17] established a model for generating tolerance chains according to mating relations between components of assemblies. Tolerance chains of an assembly can be obtained by searching its mating graph. Xue and Ji

[18] identified tolerance chains with a surface-chain model in parameterised tolerance chart. However, the method does not involve form and position tolerances. In [19], Zhou et al. proposed a method to generate dimension chains based on a joint chain model of assembly. Hu and Wu [20] classified geometric characteristics and variational geometric constraints and set up the corresponding rules between VGCs and tolerance types. On this basis, they set up VGC theory and constructed tolerance network, which effectively solves comprehensive design of dimensional and geometric tolerances.

Establishing assembly tolerance networks in CAD systems is a complex design problem, in which there are still many unresolved issues. Some commercial CAD systems have realized the automatic generation of dimensional tolerances. However, automatic design of assembly tolerance networks cannot completely be achieved. Current researches on tolerance networks do not meet the integration requirements of CAD/CAM systems. The aim of this paper is to propose a new mathematical method for establishing assembly tolerance networks based on PS theory [21] in computer systems. PS theory can use a unified mathematic expression to describe the associations between different scales, which provides a powerful system tool for resolving multiscale issues. By way of human-computer interaction, assembly information of a mechanical assembly is extracted from its three-dimensional model in CAD systems. The nonlinear assembly sequences of the assembly are generated according to related set rules. Based on this, a hierarchical assembly structure tree model of the mechanical assembly is established. According to VGC theory [20] and PS theory [21], four relation models and the systems of location relation equations and interference relation equations are established, by using which the generation of assembly tolerance types is realized. Analyzing the construction of assembly tolerance network, assembly tolerance chains between parent nodes and child nodes are set up in the assembly structure tree. Finally, the assembly tolerance networks of mechanical assemblies are constructed in the foundation of the hierarchical assembly structure tree model. On the premise of automatically extracting assembly information, the generation of assembly tolerance types is confined to subassemblies, which effectively reduces the scale of the problem and the complexity of computing and lays a good foundation for the automatic generation of assembly tolerance types of mechanical assemblies. Meanwhile, the establishment of assembly tolerance network, which is constructed based on VGC theory and PS theory, makes a useful exploration for the qualitative design of assembly tolerance types of mechanical assemblies in CAD systems.

The paper is organized as follows. Section 2 establishes an assembly structure tree model and four relation models based on PS theory and meanwhile sets up the systems of location relation equations and interference relation equations. Based on TTRS theory and contour matrices of PS theory, Section 3 builds the reasoning matrices of VGCs and corresponding assembly tolerance types. In Section 4, an exemplary mechanical assembly is used to demonstrate the constructing process of assembly tolerance networks and simultaneously to verify the feasibility and effectiveness of the proposed method.

2. Relation Models and Systems of Relation Equations

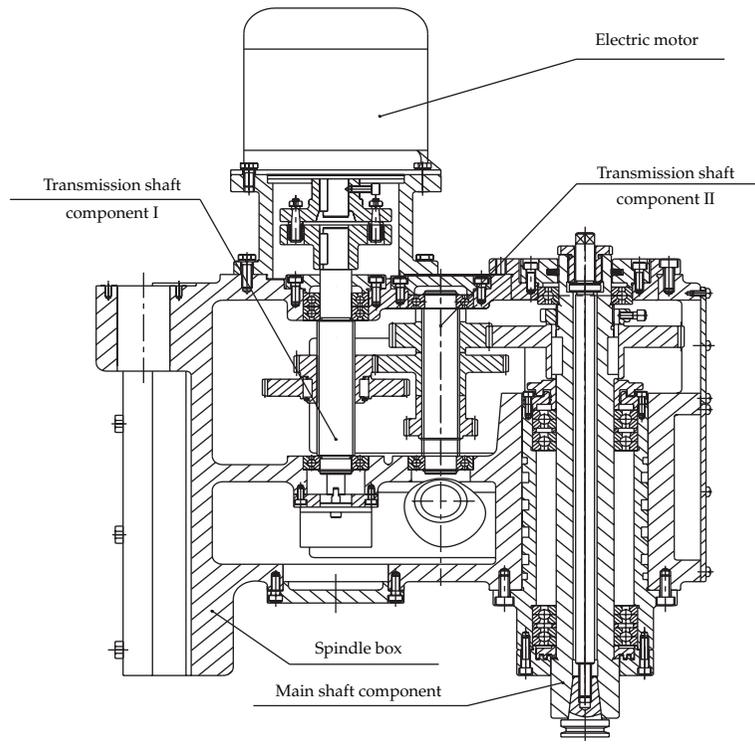
2.1. Assembly Structure Tree Model

For mechanical assemblies, their assembly processes are just like constructing objects with building blocks. Firstly, parts are assembled into components according to constraint relations

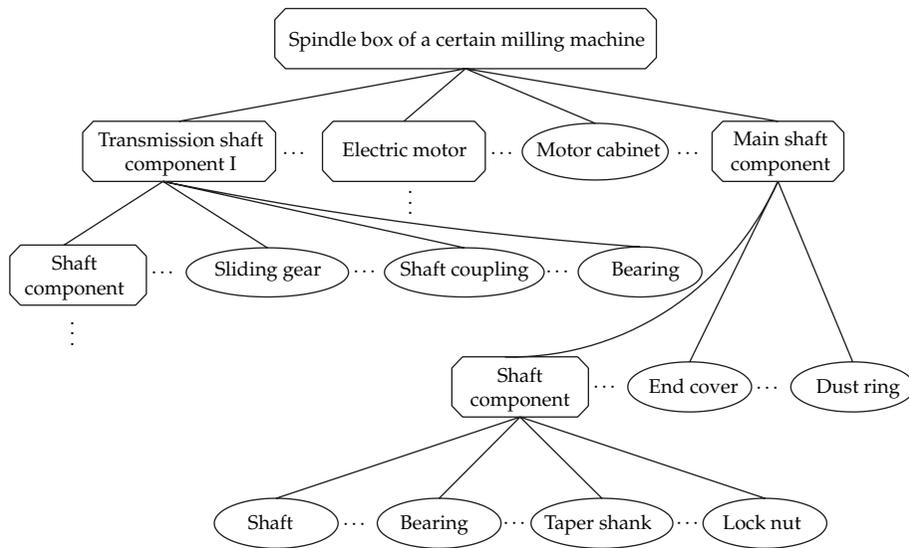
among them, and then these components and other parts are further assembled into high-level components. The above process is repeatedly carried out until the assembling of product is completed. For the majority of assemblies, their structures could be seen as follows: an assembly is composed of subassemblies and parts. Likewise, a subassembly is also composed of its subassemblies and parts directly under it. Therefore, an assembly could be disassembled into basic structural units, namely, parts. Different subassemblies and parts can be concurrently assembled. Obviously, the assembly process is a typical nonlinear process. From the above analysis, we can see that some mechanical assemblies have hierarchical structures; therefore, their structures can be expressed by product structure trees (PST). For example, the spindle box of a NC milling machine shown in Figure 1(a) just has such a structure. Its parts and subassemblies can be expressed in the different layers of a structure tree model, as shown in Figure 1(b). The assembly relations among the parts and subassemblies can also be clearly represented in the model. Hierarchical tree structure model not only expresses the structure of product but also implies its partial assembly information.

In this paper, a hierarchical structure tree model, namely, assembly structure tree (AST), is established to express assembly structure and assembly sequence of product. PST possesses a tree structure, which consists of all the parts of a product and reflects its functional relations and assembly relations. Based on the characteristics of PST and basic requirements of product assembly, we establish the following rules to reconstruct PST to obtain AST of product.

- (i) Rule 1: connection relations between parts can be divided into two types: contact and fitting. Contact includes physical contact and virtual contact. Fitting includes plane fitting, column fitting, conical fitting, spherical surface fitting, prism fitting, screw thread fitting, welding fitting, riveting fitting, and bonding fitting. Among them, connections between parts, formed by welding, riveting, and bonding, are not allowed to be dismantled; otherwise, it will result in the failure of connection form. Assembly tolerance mainly reflects assembly position precision between parts. Parts, which are connected together by bonding, riveting, or welding, should be considered as an independent part to be studied. This paper mainly considers plane fitting, column fitting, conical surface fitting, spherical surface fitting, prism fitting, and screw thread fitting.
- (ii) Rule 2: in AST, subassemblies or parts, which belong to a same parent node, cannot interfere in each other's assembling. If the interferences occur when reconstructing product structure tree, the interference components should further be broken into smaller ones at the same level, or all the subassemblies and parts in the same parent node are redivided and recombined until the assembling interferences disappear.
- (iii) Rule 3: assembling positions of parts can be determined by other parts or components' constraints in the same subassembly, and they can be assembled without interferences. Meantime, subassemblies, as a whole, can also be positioned and assembled through the contacting and fitting between them and other subassemblies or parts. They belong to the same parent node.
- (iv) Rule 4: in subassemblies, assembling positions of some parts are determined by other parts or components which do not belong to the subassemblies. These parts not only can determine the assembling positions of the subassemblies in the product but also are the assembling datums of the subassemblies. They are referred to as basic parts. In most cases, assembling process of subassembly should begin with basic parts. Under certain conditions, they can be concurrently assembled.



(a) Cutaway view of spindle box



(b) Product structure tree model

Figure 1: Spindle box of a NC milling machine.

Assembly sequence planning of subassemblies can begin with basic parts selected in the way of man-machine interaction.

- (v) Rule 5: in AST, an assembly subsequence exists among subassemblies belonging to the same parent node. Each subassembly as a whole is assembled with other subassemblies, and they all can independently realize functional requirements of product. The assembling of subassemblies in the same parent node has precedence relations.

2.2. Assembly Information Model

The assembly information model established in this paper includes geometry information and fitting information of parts. According to function, structure, and geometric shape of part, geometry information can be divided into four types: axle sleeve, wheel disc, fork, and shell. Fitting information consists of two parts: fitting type and fitting property. Based on Rule 1 mentioned above, fitting type includes plane fitting, column fitting, conical surface fitting, spherical surface fitting, prism fitting, and screw thread fitting. Fitting property can be classified into clearance fitting, transition fitting, and interference fitting.

Firstly, on the premise of automatically extracting assembly information, assembly information of parts is extracted from the three-dimensional model of product in CAD system. And then a mathematical model based on PS is established to describe the assembly information, in which the parts of assembly are used as the elements of PS and their geometric information and fitting information are used as the contour of PS. Finally, the assembly information can be formally expressed as follows:

$$\|c_{i(j)}\|_{A,F(A)} = [A \times F(A)] = \begin{bmatrix} F_1 & F_2 & \cdots & F_m \\ c_{1(1)} & c_{1(2)} & \cdots & c_{1(m)} \\ c_{2(1)} & c_{2(2)} & \cdots & c_{2(m)} \\ \vdots & \vdots & & \vdots \\ c_{n(1)} & c_{n(2)} & \cdots & c_{n(m)} \end{bmatrix} \begin{matrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{matrix}, \quad (2.1)$$

where the element set of PS is

$$A = \{a_i \mid i = 1, 2, \dots, n\}, \quad (2.2)$$

a_i represents parts or components of assembly. The contour of PS is the set

$$F(A) = \{F_j \mid j = 1, 2, \dots, m\}. \quad (2.3)$$

Its element F_j represents geometric and fitting features. In the matrix, if element a_i has corresponding relation with F_j , then $c_{i(j)} = 1$.

Based on the above analysis, a formal model of assembly information is built up as follows.

In Figure 2, F_1 – F_4 represent part types, which are, respectively, axle sleeve, wheel disc, fork, and shell. F_5 – F_{10} , respectively, represent plane fitting, column fitting, conical surface

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}
a_1	•				•	•					•		
a_2	•				•							•	
\vdots													
a_i			•		•	•				•	•	•	•
\vdots													
a_{n-1}		•				•							•
a_n	•						•				•		

Figure 2: Assembly information model.

fitting, spherical surface fitting, prism fitting, and screw thread fitting. F_{11} – F_{13} represent fitting properties of parts, namely, clearance fit, transition fit, and interference fit. The solid circle • represents that element a_i , ($i = 1, 2, \dots, n$) possesses feature F_j ($j = 1, 2, \dots, 13$).

The establishment of an assembly information model could provide basic information for subsequently setting up other models and equations.

2.3. Assembly Relation Model

If without being constrained by other parts, each part has six degrees of freedom (DOF) in free space, namely, three translational DOFs and three rotational DOFs, which translate and rotate, respectively, along the three mutually perpendicular coordinate axes. Except for the DOFs which are used to realize product functions, if all the other DOFs of a part are limited in assembly space, then its position is also determined. This implies that location requirement of the part is satisfied, which is realized by means of other related parts and components in the same subassembly. These parts and components have contact and location relations with the part and can limit its DOFs. When calculating feasible assembly sequences of a subassembly, we only need to consider assembly relations between parts which constitute the subassembly. Therefore, contact and location relations between parts in the directions of six DOFs are needed to be described in the assembly relation model.

In addition, whether a part can be assembled is also affected by other factors. For example, the space occupied by other parts probably interferes with the assembly path of the part, which makes it impossible to move the part to assembling position, interferences between a part and its assembling tools might occur if operational space is not large enough to use the assembly tools, assembling fixture interferes with assembling of parts because of its clamping method and the space occupies by it, and considering factors of man-machine engineering, there are interferences between man, part, assembly fixture, and machine. To simplify the analysis, this paper mainly considers interferences which exist between parts (or components). Consequently, assembly interference relations between parts are also needed to be described in the assembly relation model.

In an assembly, there are connection relations between parts. If not considering concrete forms of connection structures, location relations between parts can also reflect their connection relations. Thus, this paper no longer discusses connection relations of parts in detail.

By making use of hierarchical structure of the assembly structure tree, assembly information models of subassemblies in different layers could be set up. For subassembly, $S_A = \{a_i \mid i = 1, 2, \dots, n\}$ (a_i represents a part or a component belonging to the subassembly

and n is the total number of parts of the subassembly), its assembly process begins with two parts which contact each other. The two parts can be regarded as an assembly unit, and then other parts are added into it in turn under constraint conditions. Finally, the parts are assembled into a subassembly. With the same method, all the subassemblies in different levels of the assembly structure tree can be formed. This process is continuously carried out until a product is completely assembled. Based on PS theory, this paper extracts assembly information, respectively, from the assembly information model constructed above and three-dimensional model of product in CAD system to establish the assembly relation model of subassembly shown in Figure 3, in which the combination of any two parts is taken as the element of PS and the DOFs of part in assembly space are taken as the contour of PS.

Binary array (a_i, a_j) is the combination of any two parts or components in subassembly, which represents constraint relation imposed on element a_i by element a_j ; $F_{14}^L-F_{19}^L$ represent location relations between two parts, which are three translational DOFs and three rotational DOFs, respectively, along X-axis, Y-axis, and Z-axis. $F_{20}^I-F_{25}^I$ represent interference relations between two parts, respectively, along the positive and negative directions of X-axis, Y-axis, and Z-axis. The solid circle \bullet represents that there are assembly constraint relations between two parts.

2.4. Location Relation Model and System of Location Relation Equations

Positioning information of a part, which is extracted from an assembly relation model, can be used to establish its location relation model shown as Figure 4, in which $a_i, a_j \in S_A$ ($i, j \in \{1, 2, \dots, n\}$, $i \neq j$). In the location relation model, the location relation between Parts a_i and a_j is represented as function value of $\bigwedge_{k=14}^{19} F_k^L(a_i, a_j)$ where $1 \leq i, j \leq n$, $i \neq j$. If the assembly position of Part a_i is determined by Part a_j , then

$$\bigwedge_{k=14}^{19} F_k^L(a_i, a_j) = 1, \quad (2.4)$$

otherwise

$$\bigwedge_{k=14}^{19} F_k^L(a_i, a_j) = 0. \quad (2.5)$$

If the assembly position of Part a_i is determined by several related parts, then

$$\bigwedge_{k=14}^{19} \left[F_k^L(a_i, a_j) \vee F_k^L(a_i, a_l) \vee \dots \vee F_k^L(a_i, a_p) \right] = 1, \quad (2.6)$$

where $1 \leq i, j, l, p \leq n$, $i \neq j \neq l \neq p$, and n is the total number of parts in the subassembly. In the location relation model, we use the symbol \bullet to represent the value 1.

In the subassembly, if there is a group of parts which constrains 6 DOFs of a part, then a logic equation composed by this group of parts can be regarded as a location relation

	F_{14}^L	F_{15}^L	F_{16}^L	F_{17}^L	F_{18}^L	F_{19}^L	F_{20}^I	F_{21}^I	F_{22}^I	F_{23}^I	F_{24}^I	F_{25}^I
(a_1, a_2)	●				●	●	●	●				●
(a_1, a_3)	●				●							●
⋮												
(a_1, a_n)			●		●	●					●	●
(a_2, a_3)		●			●	●						●
(a_2, a_4)		●		●					●	●		
⋮												
(a_i, a_j)			●	●	●							●
(a_i, a_{j+1})		●		●	●				●	●		
⋮												
(a_{n-2}, a_n)		●		●	●	●	●		●	●		
(a_{n-1}, a_n)	●						●				●	

Figure 3: Assembly relation model.

	F_{14}^L	F_{15}^L	F_{16}^L	F_{17}^L	F_{18}^L	F_{19}^L
(a_i, a_{i+1})		●			●	●
(a_i, a_{i+2})		●		●		
⋮						
(a_i, a_j)			●	●	●	
(a_i, a_{j+1})		●		●	●	
⋮						
(a_i, a_{n-1})		●				●
(a_i, a_n)	●					

Figure 4: Location relation model of parts.

equation of this part and its logical value is used to judge whether the position of the part is determined. The formal expression for location relation is shown as follows:

$$B^L(a_i) = a_j \wedge (\vee) a_k \wedge (\vee) \cdots (a_l \wedge (\vee) a_m) \cdots, \tag{2.7}$$

where the parameters, such as a_i, a_j are the parts of the subassembly, and the AND/OR operators reflect location relations between them. Its algorithm rules are defined as follows.

(i) Rule 1: it is assumed that a_i, a_j are two parts belonging to the subassembly. If

$$\bigwedge_{k=14}^{19} F_k^L(a_i, a_j) = 1, \quad (2.8)$$

then

$$B^L(a_i) = a_j. \quad (2.9)$$

(ii) Rule 2: if

$$\bigwedge_{k=14}^{19} \left[F_k^L(a_i, a_j) \vee F_k^L(a_i, a_m) \right] = 1, \quad (2.10)$$

then

$$B^L(a_i) = a_j \wedge a_m. \quad (2.11)$$

(iii) Rule 3: if

$$\bigwedge_{k=14}^{19} F_k^L(a_i, a_j) = 1 \quad (2.12)$$

and meantime

$$\bigwedge_{k=14}^{19} F_k^L(a_i, a_m) = 1, \quad (2.13)$$

then

$$B^L(a_i) = a_j \vee a_m. \quad (2.14)$$

(iv) Rule 4: if

$$B^L(a_i) = a_j \wedge a_l, \quad (2.15)$$

$$B^L(a_i) = a_j \wedge a_m,$$

then

$$B^L(a_i) = a_j \wedge (a_l \vee a_m). \quad (2.16)$$

	F_{20}^I	F_{21}^I	F_{22}^I	F_{23}^I	F_{24}^I	F_{25}^I
(a_i, a_{i+1})						●
(a_i, a_{i+2})			●	●		
⋮						
(a_i, a_j)					●	●
(a_i, a_{j+1})			●	●		
⋮						
(a_i, a_{n-1})			●		●	●
(a_i, a_n)	●				●	

Figure 5: Interference relation model of parts.

- (v) Rule 5: it is prescribed that the position constraints, which are added to Part a_i by Part a_j , are equivalent to those which are added to Part a_j by Part a_i , that is to say

$$\bigwedge_{k=14}^{19} F_k^I(a_i, a_j) = \bigwedge_{k=14}^{19} F_k^I(a_j, a_i). \quad (2.17)$$

In order to simplify the structure of the location relation model, it is stipulated that element (a_i, a_j) must satisfy the condition $i < j$.

- (vi) Rule 6: location relation equations of all the parts in a subassembly are combined together to constitute the system of location relation equations of the subassembly.

2.5. Interference Relation Model and System of Interference Relation Equations

By means of assembly interference information of a part which is extracted from the assembly relation model, its interference relation model is set up shown in Figure 5, in which $a_i, a_j \in S_A$ ($i, j \in \{1, 2, \dots, n\}, i \neq j$), and n represents the total number of parts of subassembly. For the array element (a_i, a_j) , $F_{20}^I - F_{25}^I$ represent interference relations, respectively, along the directions of $\pm X$, $\pm Y$, and $\pm Z$ axes, which are imposed on Part a_i by Part a_j when assembling Part a_i .

In the interference relation model, the function $\bigwedge_{k=20}^{25} F_k^I(a_i, a_j)$ is used to express assembly interference relations between Part a_i and Part a_j . The algorithm rules of interference relation are established as follows.

- (i) Rule 1: when assembling Part a_i into the subassembly, if Part a_j hinders the assembly path of Part a_i , then

$$\bigwedge_{k=20}^{25} F_k^I(a_i, a_j) = 1 \quad (2.18)$$

and the interference relation equation of Part a_i is $B^I(a_i) = a_j$, otherwise

$$\bigwedge_{k=20}^{25} F_k^I(a_i, a_j) = 0, \quad (2.19)$$

$$B^I(a_i) = 0.$$

(ii) Rule 2: it is presumed that $a_i, a_j, a_l \in S_A, i, j, l \in \{1, 2, \dots, n\}$, and $i \neq j \neq l$. If Part a_j and Part a_l jointly hinder the assembly path of Part a_i , then

$$\bigwedge_{k=20}^{25} [F_k^I(a_i, a_j) \vee F_k^I(a_i, a_m)] = 1, \quad (2.20)$$

$$B^I(a_i) = a_j \wedge a_m.$$

(iii) Rule 3: If Part a_j and Part a_m , respectively, interfere with the assembly path of Part a_i , that is to say

$$\bigwedge_{k=20}^{25} F_k^I(a_i, a_j) = 1, \quad (2.21)$$

$$\bigwedge_{k=20}^{25} F_k^I(a_i, a_m) = 1,$$

then

$$B^I(a_i) = a_j \vee a_m. \quad (2.22)$$

(iv) Rule 4: if

$$B^I(a_i) = a_j \wedge a_l, \quad (2.23)$$

$$B^I(a_i) = a_j \wedge a_m,$$

then

$$B^I(a_i) = a_j \wedge (a_l \vee a_m). \quad (2.24)$$

- (v) Rule 5: owing to the direction of interference relation in assembly space, the interference constraints imposed on Part a_i by Part a_j do not mean those which are imposed on Part a_j by Part a_i . Therefore,

$$\bigwedge_{k=20}^{25} F_k^I(a_i, a_j) \neq \bigwedge_{k=20}^{25} F_k^I(a_j, a_i). \quad (2.25)$$

It is obvious that the following expressions are equivalent

$$\begin{aligned} F_{20}^I(a_i, a_j) &= F_{21}^I(a_j, a_i), \\ F_{22}^I(a_i, a_j) &= F_{23}^I(a_j, a_i), \\ F_{24}^I(a_i, a_j) &= F_{25}^I(a_j, a_i). \end{aligned} \quad (2.26)$$

In other words, the interference relations between two parts, which are, respectively, along the positive and negative directions of the same coordinate axis, are equivalent.

- (vi) Rule 6: interference relation equations of all the parts in a subassembly are combined together to constitute the system of interference relation equations of this subassembly.

3. Reasoning Methods of VGCs and Corresponding Tolerance Types

TTRS theory proposed by Salomons [22] divides functional surfaces of parts into seven basic types, which are, respectively, spherical surface, cylindrical surface, plane, helicoidal surface, rotating surface, prismatic surface, and complex surface. On this basis, Hu and Wu [20] proposed VGC theory. The theory considers that geometric constraints are constraints between nominal features. From the viewpoint of manufacturing, geometric constraints between features are variable; therefore, they are called VGC. A VGC consists of three parts: referenced feature (RF), constrained feature (CF), and variational geometric constraint (VGC). VGCs represent constraint relations between RFs and CFs. According to the differences between RF and CF, VGC can be divided into three types: self-referenced VGC (SVGC), cross-referenced VGC (CVGC), and mating VGC (MVGC). There are reasoning relations between the three kinds of VGCs and assembly tolerance types. Based on VGC theory [20] and PS theory [21, 23], this paper constructs the following matrices to describe the reasoning relations between VGCs and corresponding assembly tolerance types.

3.1. Reasoning Matrices of SVGCs and Corresponding Tolerance Types

Each SVGC is a constraint between a real feature and its corresponding associated derived feature (ADF). Functional surfaces of parts can be divided into seven types. Therefore, constraints between functional surfaces and corresponding ADFs can also be classified into

	S	Cy	Pl	H	R	Pr	Co
SC1	●						
SC2		●					
SC3			●				
SC4				●			
SC5					●		
SC6						●	
SC7							●

Figure 6: Reasoning matrix of SVGCs.

	SC1	SC2	SC3	SC4	SC5	SC6	SC7
AT1		●	●		●	●	
AT2			●				
AT3	●	●			●		
AT4		●					
AT5					●	●	●
AT6					●	●	●
AT7	●	●					

Figure 7: Reasoning matrix of tolerance types corresponding to SVGCs.

seven types. By means of the contour matrix of PS, the reasoning matrix of SVGCs is established, as shown in Figure 6. SVGCs are used as the elements of PS, and functional surfaces are used as the contour of PS. S represents spherical surface; Cy represents cylindrical surface; Pl represents plane; H represents helicoidal surface; R represents rotating surface; Pr represents prismatic surface; Co represents complex surface. SC1–SC7 represent seven types of SVGCs corresponding to the related functional surfaces. The solid circle ● represents that the corresponding relation exists between a functional surface and an SVGC.

According to the definition of SVGCs, it is obvious that there are corresponding relations between SVGCs, form tolerances, and dimensional tolerances. By using contour matrix of PS, the reasoning matrix of tolerance types corresponding to SVGCs is set up as shown in Figure 7. Tolerance types are taken as the elements of PS, in which AT1 represents straightness, AT2 represents flatness, AT3 represents circularity, AT4 represents cylindricity, AT5 represents the profile of a line, AT6 represents the profile of a surface, and AT7 represents dimensional tolerance.

3.2. Reasoning Matrices of CVGCs and Corresponding Tolerance Types

Each CVGC is a constraint between two ADFs, in which the two ADFs belong to the same part. Geometric features of a part can be decomposed into point, line, or plane. The interrelations between point, line, and plane can be divided into 27 kinds according to spatial positions of each other. Therefore, 27 kinds of CVGCs can be generated in accordance with them. The reasoning matrix of CVGCs is built shown in Figure 8. In the reasoning matrix, Po represents point, Li represents line, and Pl represents plane. IR-SIR represents spatial relations between ADFs, which are, respectively, inclusion relation, parallel relation, vertical relation in the same plane, intersection relation in the same plane, and space intersection relation. CC1–CC6 are CVGCs taking point as referenced feature, CC7–CC17 are CVGCs taking line as referenced feature, and CC18–CC27 are CVGCs taking plane as referenced feature. The solid circle • represents that there is a corresponding relation between the constrained feature of a CVGC and the spatial position relation between two features of the CVGC.

As with SVGCs, there are corresponding relations between CVGCs and assembly tolerances. Using CVGCs as contour of PS and assembly tolerance types as elements of PS, the reasoning matrix of assembly tolerance types corresponding to CVGCs is set up, as shown in Figure 9. AT8 represents parallelism; AT9 represents verticality; AT10 represents gradient; AT11 represents coaxiality; AT12 represents symmetry; AT13 represents position accuracy; AT14 represents circular run-out; AT15 represents whole runout; AT16 represents location dimension tolerance; AT17 represents location angle tolerance. The solid circle • represents that the corresponding relation between a CVGC and an assembly tolerance type is determined.

3.3. Reasoning Matrix of MVGCs

Constraints, which exist between two real features and, meanwhile, respectively, belong to two different parts, constitute MVGCs. They are actually constraints between contact surfaces of two parts. MVGCs with lower pairs, which are often seen in assembly design, can be divided into seven types. By means of contour matrix of PS, the reasoning matrix of MVGCs is established, shown in Figure 10. Mating surfaces are used as the contour of the matrix, and their corresponding MVGCs are used as the elements of the matrix, in which S represents spherical surface, Cy represents cylindrical surface, Pl represents plane, H represents helicoidal surface, R represents rotating surface, Pr represents prismatic surface, and Co represents complex surface. MC1–MC7, respectively, represent MVGCs which correspond to related mating surfaces. The solid circle • represents that the relation between mating surface and MVGC is determined.

3.4. Mating Tree

In an assembly structure tree, each mating relation uniquely corresponds to two SVGCs and one MVGC. Four features between two assembly parts are taken as nodes, and three VGCs are taken as arc curves. The tree model is called mating tree, which can be represented with the following equation:

$$MT(S1;S2) = T(V;E), \quad (3.1)$$

	Po	Li	Pl	IR	PR	VR	PIR	SIR
CC1	•			•				
CC2	•							•
CC3		•		•				
CC4		•						•
CC5			•	•				
CC6			•					•
CC7	•			•				
CC8	•							•
CC9		•		•				
CC10		•			•			
CC11		•				•		
CC12		•					•	
CC13		•						•
CC14			•	•				
CC15			•		•			
CC16			•			•		
CC17			•				•	
CC18	•			•				
CC19	•							•
CC20		•		•				
CC21		•			•			
CC22		•				•		
CC23		•					•	
CC24			•	•				
CC25			•		•			
CC26			•			•		
CC27			•				•	

Figure 8: Reasoning matrix of CVGCs.

	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9	CC10	CC11	CC12	CC13	CC14	CC15	CC16	CC17	CC18	CC19	CC20	CC21	CC22	CC23	CC24	CC25	CC26	CC27	
AT8									•						•						•				•			
AT9										•						•						•				•		
AT10											•						•						•				•	
AT11	•								•																			
AT12														•							•				•			
AT13		•	•	•	•	•	•	•		•			•	•	•													
AT14									•													•						
AT15									•													•						
AT16		•		•		•		•		•			•		•	•	•	•										
AT17												•						•						•				•

Figure 9: Reasoning matrix of tolerance types corresponding to CVGCs.

	S	Cy	Pl	H	R	Pr	Co
MC1	•						
MC2		•					
MC3			•				
MC4				•			
MC5					•		
MC6						•	
MC7							•

Figure 10: Reasoning matrix of MVGCs.

where MT represents the mating tree. S1 and S2 are two features mating each other. V is the set of two real features and two ADFs. The two real features mate each other and the two ADFs, respectively, correspond to the two real features. E is the set of SVGCs and MVGCs. The two ADFs can separately constrain their corresponding RFs by SVGCs, and the two RFs can constrain each other by MVGCs.

4. Assembly Tolerance Network

By using the reasoning methods described above, we can reason out assembly sequence of all the subassemblies in different layers of assembly structure tree and tolerance types between parts or subassemblies and then add them into the assembly structure tree. This kind of structure tree with assembly sequence and assembly tolerance information is defined as assembly tolerance network. In order to simplify the process of constructing the tolerance network, we take a subassembly of the spindle box of an NC milling machine (shown in Figure 1), namely, the main shaft component, as an example to describe the constructing process. Constructing processes related to other subassemblies and parts are similar to it. The concrete structure of the main shaft component is represented in Figure 11. Assembly sequence of a fastener can be determined by assembly sequence of a part or parts group connected by it. Therefore, this paper does not discuss the assembling of fasteners. The fasteners in Figure 11 are not marked.

- (i) Step 1: extract assembly information. On the premise of automatically recognizing features, the basic information of the main shaft component is extracted from its three-dimensional model in CAD system.
- (ii) Step 2: establish assembly structure tree. According to the rules in Section 2.1, the assembly structure tree of the main shaft component is built up, shown in Figure 12. It has a three-layer structure. The bearing groups consist of two bearings and one space collar (see Figure 11), whose assembly sequences are fixed. Therefore, this paper regards Bearing group 1 and Bearing group 2 as parts to simplify the analysis process. The dust ring is not considered here because it is a flexible part and do not have direct relations with assembly tolerances.
- (iii) Step 3: set up assembly information model. According to the rules in Section 2.2, the assembly information model of the main shaft component is established, shown in Figure 13. P_1 represents the bearing; P_2 represents the end cover 3; P_3 represents the end cover 2; P_4 represents the locknut; P_5 represents the gear; P_6 represents the shaft key; P_7 represents the bearing group 1; P_8 represents the end cover 1; P_9 represents the baffle plate; P_{10} represents the shaft sleeve; P_{11} represents the bearing group 2; P_{12} represents the dust cap 1; P_{13} represents the dust cap 2; P_{14} represents the main shaft; P_{15} represents the taper shank; P_{16} represents the nut; P_{17} represents the pull rod.
- (iv) Step 4: establish assembly relation model. On the basis of Section 2.3, we establish the assembly relation models of the main shaft component and shaft component, shown in Figure 14. The shaft component is represented with C_1 .
- (v) Step 5: establish the system of location relation equations. Extract location information from the assembly relation models in Figure 14, and then set up the system of location relation equations according to the rules in Section 2.4.

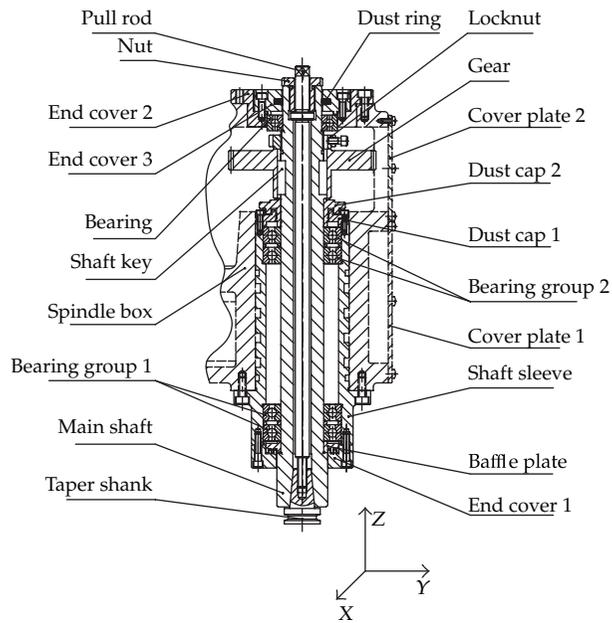


Figure 11: Main shaft component.

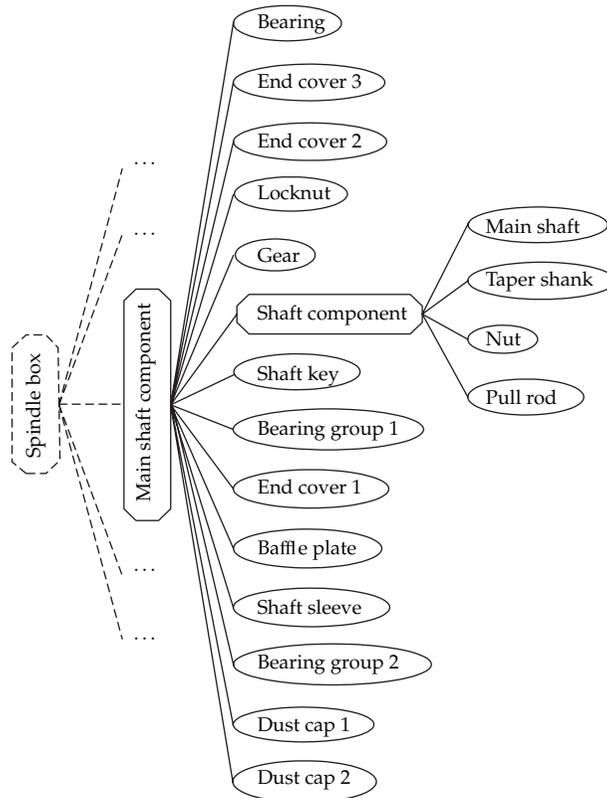


Figure 12: Assembly structure tree of main shaft component.

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}
P_1	•				•	•							•
P_2	•				•						•		
P_3	•				•	•					•		
P_4	•				•					•			
P_5		•			•	•					•		
P_6	•				•						•		
P_7	•				•	•							•
P_8	•				•	•					•		
P_9	•				•	•					•		
P_{10}	•				•	•					•		
P_{11}	•				•	•							•
P_{12}	•				•	•					•		
P_{13}	•				•	•					•		
P_{14}	•				•	•					•		•
P_{15}	•						•			•		•	
P_{16}	•	•								•			
P_{17}	•				•	•					•		

Figure 13: Assembly information model of main shaft component.

The system of location relation equations of the main shaft component is listed as follows:

$$\begin{aligned}
 B^L(P_1) &= C_1; & B^L(P_2) &= P_3 \wedge C_1; & B^L(P_3) &= P_1; & B^L(P_4) &= P_5 \wedge C_1, \\
 B^L(P_5) &= P_6 \wedge P_{13} \wedge C_1; & B^L(P_6) &= C_1; & B^L(P_7) &= (P_9 \vee P_{10}) \wedge C_1, \\
 B^L(P_8) &= P_{10}; & B^L(P_9) &= C_1; & B^L(P_{10}) &= P_7 \wedge P_{11}, \\
 B^L(P_{11}) &= P_{10} \wedge C_1; & B^L(P_{12}) &= P_{10} \wedge P_{11}, \\
 B^L(P_{13}) &= P_{11} \wedge C_1; & B^L(C_1) &= C_1.
 \end{aligned} \tag{4.1}$$

The system of location relation equations of the shaft component is listed as follows:

$$\begin{aligned}
 B^L(P_{14}) &= P_{14}; & B^L(P_{15}) &= P_{14}, \\
 B^L(P_{16}) &= P_{14}; & B^L(P_{17}) &= P_{14} \wedge P_{16}
 \end{aligned} \tag{4.2}$$

	F_{14}^L	F_{15}^L	F_{16}^L	F_{17}^L	F_{18}^L	F_{19}^L	F_{20}^L	F_{21}^L	F_{22}^L	F_{23}^L	F_{24}^L	F_{25}^L
(P_1, P_2)			•	•	•				•			
(P_1, P_3)	•	•		•	•		•	•		•	•	
(P_1, C_1)	•	•		•	•		•	•		•	•	•
(P_2, P_3)			•	•	•		•	•		•	•	•
(P_2, C_1)	•	•		•	•		•	•		•	•	
(P_4, P_5)			•	•	•				•			
(P_4, C_1)	•	•	•	•	•		•	•		•	•	
(P_5, P_{13})			•	•	•							•
(P_5, C_1)	•	•		•	•		•	•		•	•	
(P_6, C_1)	•		•	•	•	•	•		•	•		•
(P_7, P_9)			•	•	•				•			•
(P_7, P_{10})	•	•	•	•	•		•	•	•	•	•	
(P_7, C_1)	•	•	•	•	•		•	•	•	•	•	
(P_8, P_{10})	•	•	•	•	•		•	•		•	•	•
(P_9, C_1)	•	•	•	•	•		•	•	•	•	•	
(P_{10}, P_{11})	•	•	•	•	•		•	•		•	•	•
(P_{10}, P_{12})	•	•	•	•	•		•	•		•	•	•
(P_{11}, P_{13})			•	•	•							•
(P_{11}, C_1)	•	•	•	•			•	•		•	•	
	F_{14}^L	F_{15}^L	F_{16}^L	F_{17}^L	F_{18}^L	F_{19}^L	F_{20}^L	F_{21}^L	F_{22}^L	F_{23}^L	F_{24}^L	F_{25}^L
(P_{14}, P_{15})	•	•		•	•		•	•		•	•	•
(P_{14}, P_{16})	•	•	•	•	•	•	•	•	•	•	•	•
(P_{14}, P_{17})	•	•		•	•		•	•		•	•	•
(P_{16}, C_{17})			•	•	•		•	•		•	•	•

Figure 14: Assembly related models of main shaft component and shaft component. (a) Assembly relation model of main shaft component. (b) Assembly relation model of shaft component.

- (vi) Step 6: establish the system of interference relation equations. Like Step 5, extract the location information from the assembly relation models in Figure 14, and then set up the system of interference equations according to the rules in Section 2.5.

The system of interference relation equations of the main shaft component is listed as follows:

$$B^I(P_1) = P_2; \quad B^I(P_2) = 0; \quad B^I(P_3) = P_2,$$

$$B^I(P_4) = (P_1 \vee P_2 \vee P_3) \wedge C_1,$$

$$\begin{aligned}
B^I(P_5) &= (P_1 \vee P_2 \vee P_3 \vee P_4) \wedge C_1, \\
B^I(P_6) &= (P_1 \vee P_2 \vee P_3 \vee P_4) \wedge P_5 \wedge C_1, \\
B^I(P_7) &= (P_9 \vee C_1) \wedge P_{10}; \quad B^I(P_8) = 0, \\
B^I(P_9) &= (P_1 \vee P_2 \vee P_3 \vee P_4 \vee P_5 \vee P_6 \vee P_{10} \vee P_{11} \vee P_{12} \vee P_{13}) \wedge C_1, \\
B^I(P_{10}) &= (P_8 \vee P_9) \wedge C_1, \\
B^I(P_{11}) &= (P_1 \vee P_2 \vee P_3 \vee P_4 \vee P_5 \vee P_6 \vee P_{12} \vee P_{13}) \wedge C_1, \\
B^I(P_{12}) &= (P_1 \vee P_2 \vee P_3 \vee P_4 \vee P_5 \vee P_6 \vee P_{13}) \wedge C_1, \\
B^I(P_{13}) &= (P_1 \vee P_2 \vee P_3 \vee P_4 \vee P_5) \wedge C_1, \\
B^I(C_1) &= (P_1 \vee P_2 \vee P_3 \vee P_4 \vee P_5 \vee P_6 \vee P_7 \vee P_9 \vee P_{10} \vee P_{12} \vee P_{13}) \wedge (P_{10} \vee P_{11} \vee P_{13}).
\end{aligned} \tag{4.3}$$

The system of interference relation equations of the shaft component is listed as follows:

$$\begin{aligned}
B^I(P_{14}) &= P_{15} \wedge (P_{16} \vee P_{17}), \\
B^I(P_{15}) &= 0, \\
B^I(P_{16}) &= 0, \\
B^I(P_{17}) &= P_{14} \wedge P_{16}.
\end{aligned} \tag{4.4}$$

- (vii) Step 7: generate assembly sequences. Using the reasoning equations generated above and related reasoning method [24, 25], we can obtain ten assembly sequences of the main shaft component and three assembly sequences of the shaft component. The three assembly sequences of the main shaft component are shown as follows:

$$\begin{aligned}
&C_1 \longrightarrow P_9 \longrightarrow P_7 \longrightarrow P_{10} \longrightarrow P_{11} \longrightarrow P_{12} \longrightarrow P_{13} \longrightarrow P_6 \longrightarrow P_5 \longrightarrow P_4 \\
&\quad \longrightarrow P_1 \longrightarrow P_3 \longrightarrow P_2 \longrightarrow P_8, \\
&C_1 \longrightarrow P_9 \longrightarrow P_7 \longrightarrow P_{10} \longrightarrow P_8 \longrightarrow P_{11} \longrightarrow P_{12} \longrightarrow P_{13} \longrightarrow P_6 \longrightarrow P_5 \\
&\quad \longrightarrow P_4 \longrightarrow P_1 \longrightarrow P_3 \longrightarrow P_{28}, \\
&C_1 \longrightarrow P_9 \longrightarrow P_7 \longrightarrow P_{10} \longrightarrow P_{11} \longrightarrow P_8 \longrightarrow P_{12} \longrightarrow P_{13} \longrightarrow P_6 \longrightarrow P_5 \\
&\quad \longrightarrow P_4 \longrightarrow P_1 \longrightarrow P_3 \longrightarrow P_{28}.
\end{aligned} \tag{4.5}$$

The assembly sequences of the shaft component are shown as follows:

$$\begin{aligned}
 P_{14} &\longrightarrow P_{17} \longrightarrow P_{16} \longrightarrow P_{15}, \\
 P_{14} &\longrightarrow P_{17} \longrightarrow P_{15} \longrightarrow P_{16}, \\
 P_{14} &\longrightarrow P_{15} \longrightarrow P_{17} \longrightarrow P_{16}.
 \end{aligned} \tag{4.6}$$

(viii) Step 8: determine mating tree and datum reference frame. SVGCs and MVGCs are combined into mating trees. ADFs of all the datums belonging to the same part are combined together to form a datum reference frame (DRF), between which there are no SVGCs. Taking the shaft component as an example, we use its assembly sequence, namely, $P_{14} \rightarrow P_{17} \rightarrow P_{16} \rightarrow P_{15}$, to obtain mating trees $M(P_{14}, P_{17})$, $M(P_{17}, P_{16})$, $M(P_{16}, P_{14})$, and $M(P_{14}, P_{15})$. The DRFs of P_{14} , P_{15} , P_{16} , and P_{17} are, respectively, DRF_{14} , DRF_{15} , DRF_{16} , and DRF_{17} .

(ix) Step 9: generate assembly feature chains. Taking the mating trees and datum reference frames as nodes, the subassembly sequence is reconstructed, which starts from P_{15} . The result is shown as follows:

$$DRF_{15} \longrightarrow MT(P_{15}, P_{14}) \longrightarrow DRF_{14} \longrightarrow MT(P_{14}, P_{17}) \longrightarrow DRF_{17} \longrightarrow MT(P_{17}, P_{16}) \longrightarrow DRF_{16}. \tag{4.7}$$

By means of the features of the parts in Figure 10, the above expression can be further decomposed into a feature chain as follows:

$$\begin{aligned}
 DRF_{15} &\longrightarrow P_{15_ADF_1} \longrightarrow P_{15_RF_1} \longrightarrow P_{14_RF_1} \longrightarrow P_{14_ADF_1} \\
 &\longrightarrow DRF_{14} \longrightarrow P_{14_ADF_2} \longrightarrow P_{14_RF_2} \longrightarrow P_{17_RF_1} \longrightarrow P_{17_ADF_1} \\
 &\longrightarrow DRF_{17} \longrightarrow P_{17_ADF_2} \longrightarrow P_{17_RF_2} \longrightarrow P_{16_RF_1} \longrightarrow P_{16_ADF_1} \longrightarrow DRF_{16},
 \end{aligned} \tag{4.8}$$

where ADF_1 and ADF_2 are two ADFs which belong to the same part. Likewise, RF_1 and RF_2 are two RFs which also belong to the same part.

(x) Step 10: reason VGC types between features. Using the reasoning matrices of SVGCs and CVGCs in Sections 3.2–3.4, VGC types can be reasoned out and then be added into the above expression. Finally, the expression is further decomposed into VGC chain in which the components are translations and rotations, respectively, along the directions of X , Y , and Z -axis, shown as follows:

$$\begin{aligned}
 DRF_{15} &\xrightarrow[T_x, T_y, T_z, R_x, R_y, R_z]{CC9} P_{15_ADF_1} \xrightarrow[T_x, T_y, T_z, R_x, R_y]{SC5} P_{15_RF_1} \xleftarrow[T_x, T_y, R_x, R_y]{MC5} P_{14_RF_1} \\
 &\xleftarrow[T_x, T_y, T_z, R_x, R_y]{SC5} P_{14_ADF_1} \xleftarrow[T_x, T_y, T_z, R_x, R_y, R_z]{CC9} DRF_{14} \xrightarrow[T_z, R_x, R_y]{CC25} P_{14_ADF_2}
 \end{aligned}$$

$$\begin{aligned}
& \xrightarrow[T_z, R_x, R_y]{SC3} P_{14_RF2} \xleftrightarrow[T_z, R_x, R_y]{MC3} P_{17_RF1} \xleftarrow[T_z, R_x, R_y]{SC3} P_{17_ADF1} \xleftarrow[T_z, R_x, R_y]{CC25} DRF17 \\
& \xrightarrow[T_z, R_x, R_y]{CC25} P_{17_ADF2} \xrightarrow[T_z, R_x, R_y]{SC3} P_{17_RF2} \xleftarrow[T_z, R_x, R_y]{MC3} P_{16_RF1} \\
& \xleftarrow[T_z, R_x, R_y]{SC3} P_{16_ADF1} \xleftarrow[T_z, R_x, R_y]{CC25} DRF16,
\end{aligned} \tag{4.9}$$

where \rightarrow represents the path from a referenced feature to a constrained feature and \leftrightarrow represents that the two features connected by it are mutually a referenced feature and a constrained feature.

- (xi) Step 11: reason tolerance types corresponding to VGCs. It can be seen that the same VGCs can correspond to different tolerance types in the reasoning matrices of tolerance types in Sections 3.2–3.4. Therefore, the related tolerance types should be chosen according to the assembly properties and functional requirements between the mating parts. By virtue of the reasoning matrices, the tolerance type corresponding with each of the VGCs can be reasoned out and then be added into the VGC chain to get the subassembly tolerance chain shown as follows:

$$\begin{aligned}
DRF15 & \xrightarrow[T_x, T_y, T_z, R_x, R_y, R_z]{CC9 \ AT11} P_{15_ADF1} \xrightarrow[T_x, T_y, T_z, R_x, R_y]{SC5 \ AT1 \ AT3} P_{15_RF1} \xleftarrow[T_x, T_y, R_x, R_y]{MC5 \ Fitting \ accuracy} P_{14_RF1} \\
& \xleftarrow[T_x, T_y, T_z, R_x, R_y]{SC5 \ AT1 \ AT3} P_{14_ADF1} \xleftarrow[T_x, T_y, T_z, R_x, R_y, R_z]{CC9 \ AT11} DRF14 \xrightarrow[T_z, R_x, R_y]{CC25 \ AT8} P_{14_ADF2} \xrightarrow[T_z, R_x, R_y]{SC3 \ AT2} P_{14_RF2} \\
& \xrightarrow[T_z, R_x, R_y]{MC3 \ Fitting \ accuracy} P_{17_RF1} \xleftarrow[T_z, R_x, R_y]{SC3} P_{17_ADF1} \xleftarrow[T_z, R_x, R_y]{CC25 \ AT8} DRF17 \xrightarrow[T_z, R_x, R_y]{CC25 \ AT8} P_{17_ADF2} \\
& \xrightarrow[T_z, R_x, R_y]{SC3 \ AT2} P_{17_RF2} \xleftarrow[T_z, R_x, R_y]{MC3 \ Fitting \ accuracy} P_{16_RF1} \xleftarrow[T_z, R_x, R_y]{SC3 \ AT2} P_{16_ADF1} \xleftarrow[T_z, R_x, R_y]{CC25 \ AT8} DRF16.
\end{aligned} \tag{4.10}$$

- (xii) Step 12: establish assembly tolerance network. In the light of the reasoning rules mentioned above, the assembly tolerance types of the spindle box and its all subassemblies can be reasoned out to form assembly tolerance chains. All the assembly tolerance chains reasoned out in Step 11 and the assembly sequence are combined to construct the assembly tolerance network of the product. Figure 15 is a schematic graph of assembly tolerance network of the spindle box, in which we can see that all the subassembly tolerance chains in dashed line frames and assembly sequences represented by the numbers with a circle are combined to form the assembly tolerance network of the whole product.

5. Conclusions

Assembly tolerance network design is determined by many factors, and these factors associate with each other. Therefore, the design is a multiscale issue. Many scholars have applied

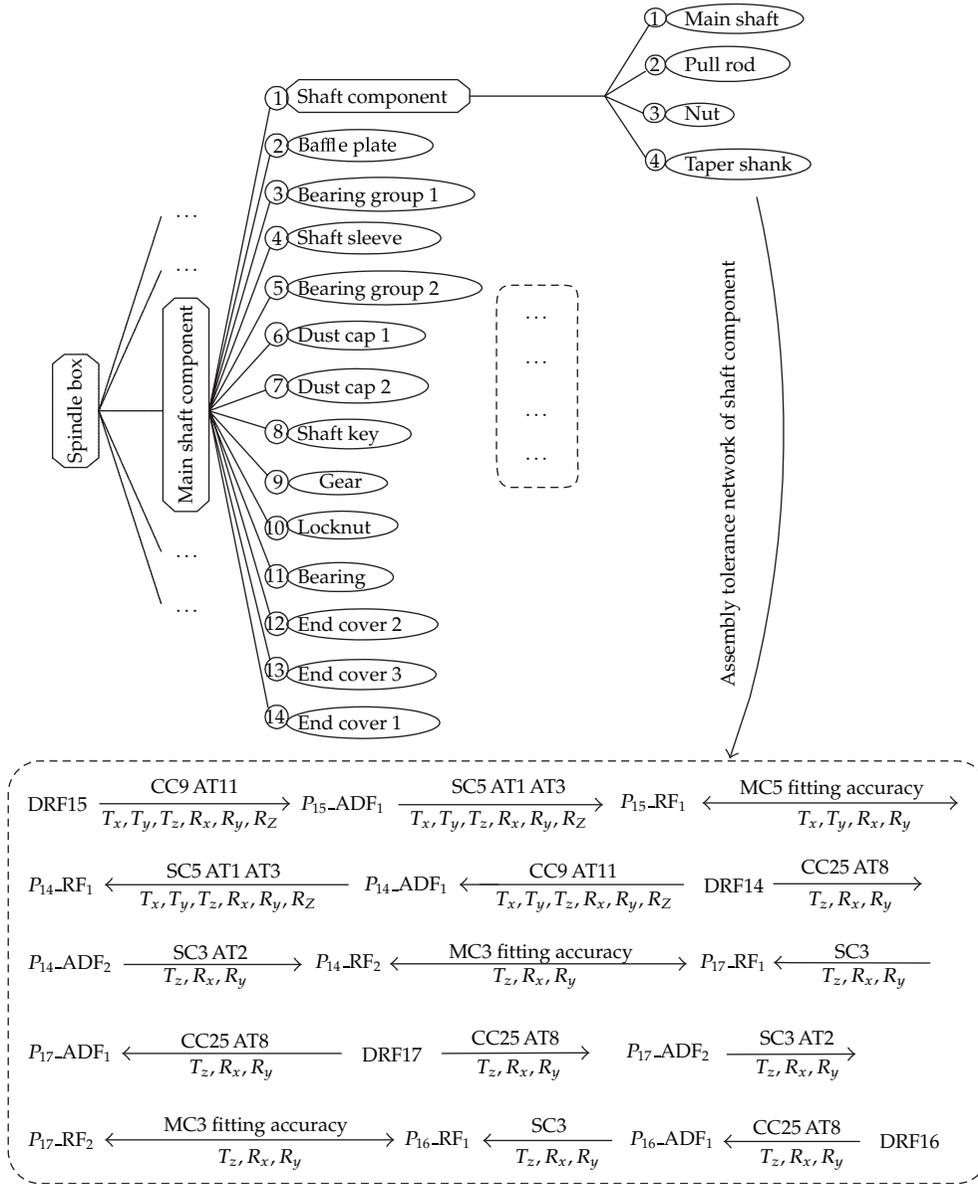


Figure 15: Assembly tolerance network of the spindle box.

various methods to resolve the multiscale issues in different research fields, such as multiscale time schemes [1], multiscale time-frequency representation [3], multiscale curvelet shrinkage [4], multiscale Retinex model [7], and multiscale nonlinear modeling algorithm [10]. Obviously, there are many factors affecting assembly tolerance design and they have great differences between each other. An effective method must be used to represent the reasoning and constraint relations. In this paper, PS theory [21] is introduced to describe the associations between different factors. Some models are established with the unified mathematic expression, including assembly information model, assembly relation model between parts, location

relation model of parts, interference relation model of parts, reasoning matrices of VGCs, and reasoning matrices of tolerance types. On this basis, this paper combines assembly tolerance design with assembly sequence planning to research the automatic generation of assembly tolerances of mechanical assemblies. By means of the product prototype in CAD system as well as the establishment of a hierarchical assembly structure tree model, this paper uses the related reasoning rules to solve the assembly sequences of the product. Then, the automatic generation of assembly tolerance types and the construction of assembly tolerance network are realized based on the variational geometric constraints theory and polychromatic sets theory. The assembly sequences, respectively, belonging to components and product are generated by the same method, which can effectively reduce the scale of the solved problem and avoid the combination explosion in the solving process. All the relation models are established by using the unified mathematic expression, which is convenient to manage knowledge better and can enhance reasoning efficiency when designing assembly tolerance networks. This method lays a good foundation for the automatic generation of assembly tolerances of mechanical assemblies. Meanwhile, the establishment of assembly tolerance network makes a useful exploration for the quantitative design of assembly tolerances of complex product.

Acknowledgments

The authors are particularly grateful to Editor Carlo Cattani for his continual encouragement, patience and sincere help on earlier drafts of this paper. The authors also would like to thank two anonymous reviewers whose insightful comments and constructive suggestions have substantially improved the paper. The authors gratefully acknowledge the support from the major project of the National Natural Science Foundation of China (NSFC) under Grant number 50935006.

References

- [1] J. Kou, S. Sun, and B. Yu, "Multiscale time-splitting strategy for multiscale multiphysics processes of two-phase flow in fractured media," *Journal of Applied Mathematics*, vol. 2011, Article ID 861905, 24 pages, 2011.
- [2] C. Picard, C. Frisson, F. Faure, G. Drettakis, and P. G. Kry, "Advances in modal analysis using a robust and multiscale method," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 392782, 12 pages, 2010.
- [3] H. Zhu, C. Liu, and W. Gaetz, "Estimation of time-varying coherence and its application in understanding brain functional connectivity," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 390910, 11 pages, 2010.
- [4] L. Xiao, L.-L. Huang, and B. Roysam, "Image variational denoising using gradient fidelity on curvelet shrinkage," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 398410, 16 pages, 2010.
- [5] E. G. Bakhoun 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.
- [6] M. Djilas, C. Azevedo-Coste, D. Guiraud, and K. Yoshida, "Spike sorting of muscle spindle afferent nerve activity recorded with thin-film intrafascicular electrodes," *Computational Intelligence and Neuroscience*, vol. 2010, Article ID 836346, 13 pages, 2010.
- [7] S. Chen and A. Beghdadi, "Natural enhancement of color image," *EURASIP Journal on Image and Video Processing*, vol. 2010, Article ID 175203, 19 pages, 2010.
- [8] C. Chen, R. Saxena, and G.-W. Wei, "A multiscale model for virus capsid dynamics," *International Journal of Biomedical Imaging*, vol. 2010, Article ID 308627, 9 pages, 2010.

- [9] L. Florack and H. Van Assen, "A new methodology for multiscale myocardial deformation and strain analysis based on tagging MRI," *International Journal of Biomedical Imaging*, vol. 2010, Article ID 341242, 8 pages, 2010.
- [10] M. N. Nounou and H. N. Nounou, "Reduced noise effect in nonlinear model estimation using multiscale representation," *Modelling and Simulation in Engineering*, vol. 2010, Article ID 217305, 8 pages, 2010.
- [11] M. N. Nounou and H. N. Nounou, "Multiscale latent variable regression," *International Journal of Chemical Engineering*, vol. 2010, Article ID 935315, 8 pages, 2010.
- [12] A. Lucia, "A multiscale gibbs-helmholtz constrained cubic equation of state," *Journal of Thermodynamics*, vol. 2010, Article ID 238365, 10 pages, 2010.
- [13] A. Clement and A. Riviere, "Tolerancing versus nominal modeling in next generation CAD/CAM system," in *Proceedings of 3rd CIRP Seminar on Computer Aided Tolerancing*, Cachan, France, April 1993.
- [14] A. Clement, A. Riviere, and P. A. Serre, "A declarative information model for functional requirements," in *Proceedings of 4th CIRP Seminars on Computer Aided Tolerancing*, Tokyo, Japan, April 1995.
- [15] P. Hoffmann, "Analysis of tolerances and process inaccuracies in discrete part manufacturing," *Computer-Aided Design*, vol. 14, no. 2, pp. 83–88, 1982.
- [16] A. Desrochers and A. Clément, "A dimensioning and tolerancing assistance model for CAD/CAM systems," *The International Journal of Advanced Manufacturing Technology*, vol. 9, no. 6, pp. 352–361, 1994.
- [17] N. Wang and T. M. Ozsoy, "Automatic generation of tolerance chains from mating relations represented in assembly models," *Journal of Mechanical Design*, vol. 115, no. 4, pp. 757–761, 1993.
- [18] J. Xue and P. Ji, "Identifying tolerance chains with a surface-chain model in tolerance charting," *Journal of Materials Processing Technology*, vol. 123, no. 1, pp. 93–99, 2002.
- [19] J. Q. Zhou, Y. F. Xing, X. M. Lai, G. L. Chen, X. Lan, and C. Z. Wei, "Study on dimension chain generation for auto-body tolerance analysis," *Journal of Shanghai Jiaotong University (Science)*, vol. 11, no. 4, pp. 417–422, 2006.
- [20] J. Hu and Z. Wu, "Methods for generation of variational geometric constraints network for assembly," *Journal of Computer-Aided Design and Computer Graphics*, vol. 14, no. 1, pp. 79–82, 2002.
- [21] Z. Li and L. D. Xu, "Polychromatic sets and its application in simulating complex objects and systems," *Computers and Operations Research*, vol. 30, no. 6, pp. 851–860, 2003.
- [22] O. W. Salomons, *Computer support in the design of mechanical products: constraint specification and satisfaction in feature based design for manufacturing*, Ph.D. thesis, University of Twente, The Netherlands, 1995.
- [23] Y. Zhang, Z. B. Li, and J. K. Wang, "Hierarchical reasoning model of tolerance information and its using in reasoning technique of geometric tolerance types," in *the International Conference of Intelligent Robotics and Applications*, Wuhan, China, October 2008.
- [24] S. Zhao and Z. Li, "Formalized reasoning method for assembly sequences based on Polychromatic Sets theory," *International Journal of Advanced Manufacturing Technology*, vol. 42, no. 9-10, pp. 993–1004, 2009.
- [25] S. Zhao and Z. Li, "A new assembly sequences generation of three dimensional product based on polychromatic sets," *Information Technology Journal*, vol. 7, no. 1, pp. 112–118, 2008.

Research Article

The Inverse Problem for a General Class of Multidimensional Hyperbolic Equations

Gani Aldashev^{1,2} and Serik A. Aldashev³

¹ *Department of Economics and CRED, University of Namur (FUNDP), 5000 Namur, Belgium*

² *ECARES, Université libre de Bruxelles, 1050 Bruxelles, Belgium*

³ *Department of Mathematics, Aktobe State University, 030000 Aktobe, Kazakhstan*

Correspondence should be addressed to Gani Aldashev, gani.aldashev@fundp.ac.be

Received 1 July 2011; Accepted 4 August 2011

Academic Editor: Carlo Cattani

Copyright © 2012 G. Aldashev and S. A. Aldashev. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Inverse problems for hyperbolic equations are found in geophysical prospecting and seismology, and their multidimensional analogues are especially important for applied work. However, whereas results have been established for the some narrow classes of hyperbolic equations, no results exist for more general classes. This paper proves the solvability of the inverse problem for a general class of multidimensional hyperbolic equations. Our approach consists of properly choosing the shape of the overidentifying condition that is needed to determine the right-hand side of the hyperbolic PDE and then applying the Fourier series method. We are then able to establish the results of the existence of solution for the cases when the unknown right-hand side is time-independent or space independent.

1. Introduction

One key applied informational problem with heavy involvement of advanced mathematical modelling is the inverse problem. Generally speaking, it can be described as follows: a physical, biological, or an economic model relates some parameters, via the knowledge of certain fundamental laws, to observational data. An inverse problem then consists of recovering the (unknown) parameters of the model from known observational data. Such problems often emerge and are intensively studied in geophysics and seismology (in physical sciences, [1–3]), tomography (in biological sciences, [1, 4–6]) and economic geography, econometrics, and macroeconomics (in economic sciences, [7–10]).

The mathematical problems that emerge in such studies often reduce to the analysis of linear second-order partial differential equations. For instance, in recent analysis of

spatial economic development, economists construct models with continuous time and space dimensions. Capital accumulation in different points of space is modelled via an equation of motion that becomes a parabolic partial differential equation [11, 12]. An important (inverse) problem is then recovering the unknown parameters of the aggregate production function of the economy from observations about the level of capital stock at some moment in time [7].

Instead, inverse problems for hyperbolic equations emerge in the study of the problems of geophysics and seismology [1–3]. For instance, two major types of problems in geophysical prospecting are (i) the structural problems that are concerned with the structure of sedimentary deposits and the Earth's crust (the structures that have to do with oil and gas resources) and (ii) the location problems that are concerned with establishing local non-homogeneities in sedimentary deposits and crust (caused by mineral deposits, e.g., ore). Both problems basically consists of recovering the Earth's inner structure from the surface measurements of physical fields.

In seismology, one of the most important problems is determining the velocity of propagation of seismic waves within the Earth, given the information about the seismic wave fronts on the surface (coming from different earthquakes). The velocities of longitudinal and transverse waves and the elastic properties of the medium are linked through some known formulae. The determination of the propagation velocities given these formulae reduces mathematically to an inverse problem known as the Herglotz-Wiechert inverse kinematic problem of seismics.

In most practical applications, the parameters of the physical model are a multidimensional function (e.g. three-dimensional space plus time). Therefore, considerable attention has been devoted to the analysis of multidimensional inverse problems. Chapter 8 of [3] provides a good survey of the current state of the literature for hyperbolic problems, together with a fine discussion of the mathematical difficulties involved.

So far, however, most of the results have been obtained for relatively narrow classes of equations: the so-called model equations, that is, including only the wave operator, or only some of the lower-order derivatives. After the classic study of the inverse problem for the wave equation [1], some results have been established (and applied techniques developed) for the hyperbolic equations of the following classes: $u_{tt} = \Delta u - p(x)u$ in [13], $u_{tt} = \Delta u + p(x,t)u$ in [14], and $c(x)u_{tt} = \Delta u$ in [15]. To the best of our knowledge, no results have been yet found for more general classes of hyperbolic equations. This paper tries to close this gap by establishing the existence result for a general class of multidimensional hyperbolic equations.

Our approach consists of properly choosing the shape of the overidentifying condition that is needed to determine the right-hand side of the hyperbolic PDE, and then applying the Fourier series method. We are then able to establish the results of the existence of solution for the cases when the unknown right-hand side is time independent or space-independent.

2. Setup of the Problem

Let D_ε be a finite domain of the Euclidean space of E_{m+1} points (x_1, \dots, x_m, t) bounded by the surfaces $|x| = t + \varepsilon$, $|x| = 1 - t$ and the plane $t = 0$, where $|x|$ denotes the length of the vector $x = (x_1, \dots, x_m)$, $0 \leq t \leq (1 - \varepsilon)/2$, and $0 < \varepsilon < 1$. Moreover, let us denote the parts of these surfaces that form the boundary ∂D_ε of the domain D_ε as S_ε , S_1 , and S , respectively.

In this domain, a number of applied problems described in the Introduction can be described mathematically as some special cases of the following general inverse problem.

Problem 2.1. Find the functions $u(x, t) \in C(\overline{D_\varepsilon}) \cap C^2(D_\varepsilon)$ and $g(x, t)$ that are linked in the domain D_ε by the following equation:

$$Lu \equiv \Delta_x u - u_{tt} + \sum_{i=1}^m a_i(x, t) u_{x_i} + b(x, t) u_t + c(x, t) u = g(x, t), \quad (2.1)$$

where the functions $u(x, t)$ should satisfy the conditions

$$u|_S = \tau(x), \quad u_t|_S = v(x), \quad (2.2)$$

$$u|_{S_\beta} = \sigma(x), \quad (2.3)$$

Δ_x being the Laplace operator defined over the variables x_1, \dots, x_m , $m \geq 2$ and S_β being the cone $\beta|x| = t + \varepsilon$, $0 < \beta = \text{const} < 1$.

Note that in this inverse problem the conditions (2.2) are the conditions of the standard Cauchy problem, while the condition (2.3) is the overidentification condition, which is needed to determine the unknown function $g(x, t)$. The appropriate choice of this condition would allow us to establish the results below.

Before deriving our main results, however, it is useful to switch from the Cartesian coordinates x_1, \dots, x_m, t to the spherical ones $r, \theta_1, \dots, \theta_{m-1}, t$, with $0 \leq \theta_1 < 2\pi$, $r \geq 0$, $0 \leq \theta_i \leq \pi$, $i = 2, 3, \dots, m-1$.

We need further some additional notation. Let $\{Y_{n,m}^k(\theta)\}$ be a system of linearly independent spherical functions of degree n , $1 \leq k \leq k_n$, $(m-2)!n!k_n = (n+m-3)!(2n+m-2)$, $\theta = (\theta_1, \dots, \theta_{m-1})$, $W_2^l(D_\varepsilon)$, $l = 0, 1, \dots$ —are Sobolev spaces, and $\tilde{S}_\beta = \{(r, \theta) \in S, \varepsilon < r < (1+\varepsilon)/(1+\beta)\}$.

Also, let us denote as $\tilde{a}_n^k(r, t)$, $a_n^k(r, t)$, $\tilde{b}_n^k(r, t)$, $\tilde{c}_n^k(r, t)$, $\tilde{g}_n^k(r)$, ρ_n^k , $\tilde{\tau}_n^k(r)$, $\tilde{v}_n^k(r)$, and $\tilde{\sigma}_n^k(r)$ the coefficients of the decomposition of the series (2.4), respectively, of the functions $a_i(r, \theta, t)\rho(\theta)$, $a_i(x_i/r)\rho$, $b(r, \theta, t)\rho$, $c(r, \theta, t)\rho$, $g(r, \theta)$, $\rho(\theta)$, $i = 1, \dots, m$, $\tau(r, \theta)$, $v(r, \theta)$, and $\sigma(r, \theta)$, and moreover, $\rho(\theta) \in C^\infty(\Gamma)$, Γ being the unit sphere in E_m .

For our analysis, we will exploit the following lemma that has been shown in [16].

Lemma 2.2. Let $f(r, \theta) \in W_2^l(S)$. If $l \geq m-1$, then the series

$$f(r, \theta) = \sum_{n=0}^{\infty} \sum_{k=1}^{k_n} f_n^k(r) Y_{n,m}^k(\theta), \quad (2.4)$$

as well as the series obtained through its differentiation of order $p \leq l - m + 1$, converges absolutely and uniformly.

Next, one introduces the set of functions

$$B^l(S) = \left\{ f(r, \theta) : f \in W_2^l(S), \sum_{n=0}^{\infty} \sum_{k=1}^{k_n} \left(\|f_n^k(r)\|_{C^2(\varepsilon, 1)}^2 + \|f_n^k(r)\|_{C([\varepsilon, 1])} \right) \times \exp 2(n^2 + n(m-2)) < \infty, l \geq m-1 \right\}. \quad (2.5)$$

Finally, let $a_i(x, t)$, $b(x, t)$, $c(x, t) \in W_2^p(D_\varepsilon) \subset C(\overline{D_\varepsilon})$, $i = 1, \dots, m$, $p \geq m+1$, and $\tau(r, \theta)$, $v(r, \theta) \in B^l(S)$, $\sigma(r, \theta) \in B^l(\tilde{S}_\beta)$.

3. Main Results

We are now ready to establish our main results. These are theorems of the existence of solutions of Problem 1, when the unknown right-hand side of (2.1) is time independent or space independent.

Theorem 3.1. *If $g(r, \theta, t) = g(r, \theta) \in W_2^l(D_\varepsilon) \cap C^1(\overline{D_\varepsilon})$, then the functions $u(r, \theta, t)$ and $g(r, \theta)$ always exist.*

Theorem 3.2. *If $g(r, \theta, t) = g(t) \in C^1([0, (1-\varepsilon)/2])$, then the functions $u(r, \theta, t)$ and $g(t)$ always exist.*

Proof of Theorem 1. The uniqueness of the solution of the (direct) Cauchy problem (2.1), (2.2) is well known (see, for instance, [17]). We will search for the solution of this problem in the form of the series

$$u(r, \theta, t) = \sum_{n=0}^{\infty} \sum_{k=1}^{k_n} \bar{u}_n^k(r, t) Y_{n,m}^k(\theta), \quad (3.1)$$

where $\bar{u}_n^k(r, t)$ are the functions to be determined later.

Substituting (3.1) into (2.1), and first multiplying the resulting expression by $\rho(\theta) \neq 0$ and then integrating over the unit sphere Γ , we get for \bar{u}_n^k (see [18, 19])

$$\begin{aligned} & \rho_0^1 \bar{u}_{0rr}^1 - \rho_0^1 \bar{u}_{0tt}^1 + \left(\frac{m-1}{r} \rho_0^1 + \sum_{i=1}^m a_{i0}^1 \right) \bar{u}_{0r}^1 + \tilde{b}_0^1 \bar{u}_{0t}^1 + \tilde{c}_0^1 \bar{u}_0^1 - \tilde{c}_0^1 \bar{g}_0^1(r) \\ & + \sum_{n=0}^{\infty} \sum_{k=1}^{k_n} \left\{ \rho_n^k \bar{u}_{nrr}^k - \rho_n^k \bar{u}_{ntt}^k + \left(\frac{m-1}{r} \rho_n^k + \sum_{i=1}^m a_{in}^k \right) \bar{u}_{nr}^k + \tilde{b}_n^k \bar{u}_{nt}^k \right. \\ & \left. + \left[\tilde{c}_n^k - \frac{\lambda_n}{r^2} \rho_n^k + \sum_{i=1}^m (\tilde{a}_{in-1}^k - n a_{in}^k) \right] \bar{u}_n^k - \rho_n^k \bar{g}_n^k(r) \right\} = 0, \end{aligned} \quad (3.2)$$

$$\lambda_n = n(n+m-2).$$

Next, let us analyze the infinite system of differential equations

$$\rho_0^1 \bar{u}_{0rr}^1 - \rho_0^1 \bar{u}_{0tt}^1 + \frac{m-1}{r} \rho_0^1 \bar{u}_{0r}^1 = \rho_0^1 g_0^1(r), \quad (3.3)$$

$$\begin{aligned} \rho_1^k \bar{u}_{1rr}^k - \rho_1^k \bar{u}_{1tt}^k + \frac{m-1}{r} \rho_1^k \bar{u}_{1r}^k - \frac{\lambda_1}{r^2} \rho_1^k \bar{u}_1^k &= \rho_1^k g_1^k(r) \\ - \frac{1}{k_1} \left(\sum_{i=1}^m a_{i0}^1 \bar{u}_{0r}^1 + \tilde{b}_0^1 \bar{u}_{0t}^1 + \tilde{c}_0^1 \bar{u}_0^1 \right), \quad n=1, \quad k = \overline{1, k_1}, \end{aligned} \quad (3.4)$$

$$\begin{aligned} \rho_n^k \bar{u}_{nr}^k - \rho_n^k \bar{u}_{ntt}^k + \frac{m-1}{r} \rho_n^k \bar{u}_{nr}^k - \frac{\lambda_n}{r^2} \rho_n^k \bar{u}_n^k &= \rho_n^k g_n^k(r) \\ - \frac{1}{k_n} \sum_{k=1}^{k_n-1} \left\{ \sum_{i=1}^m a_{in-1}^k \bar{u}_{n-1r}^k + \tilde{b}_{n-1}^k \bar{u}_{n-1t}^k \right. \\ \left. + \left[\tilde{c}_{n-1}^k + \sum_{i=1}^m (\tilde{a}_{in-2}^k - (n-1)a_{in-1}^k) \right] \bar{u}_{n-1}^k \right\}, \quad k = \overline{1, k_n}, \quad n = 2, 3, \dots \end{aligned} \quad (3.5)$$

Note that if we sum the expression (3.4) from 1 to k_1 , the expression (3.5) from 1 to k_n , and then add the obtained expressions to (3.3), we obtain (3.2). Therefore, if $\{\bar{u}_n^k\}$, $k = \overline{1, k_n}$, $n = 0, 1, \dots$ is the solution of the system (3.3)–(3.5), then it is also the solution of (3.2). The opposite is not true, however; therefore, we cannot show the uniqueness of the solution of Problem 1.

Now, taking into account the orthogonality of the spherical functions $Y_{n,m}^k(\theta)$ ([16]) and given expression (3.1), we obtain, from the boundary-value conditions (2.2), (2.3):

$$\begin{aligned} \bar{u}_n^k(r, 0) &= \bar{v}_n^k(r), \quad \bar{u}_{nt}^k(r, 0) = \bar{v}_n^k(r), \quad \varepsilon \leq r \leq 1, \\ \bar{u}_n^k(r, \beta r + \varepsilon) &= \bar{\sigma}_n^k(r), \quad \varepsilon \leq r \leq \frac{1-\varepsilon}{1+\beta}, \quad k = \overline{1, k_n}, \quad n = 0, 1, \dots \end{aligned} \quad (3.6)$$

Thus, we have shown that our inverse problem (2.1)–(2.3) reduces to a system of inverse problems for the equations (3.3)–(3.5). We can now look for the solution of these problems.

It is easy to see that each equation of the system (3.3)–(3.5) can be represented in the form

$$\bar{u}_{nr}^k - \bar{u}_{ntt}^k + \frac{m-1}{r} \bar{u}_{nr}^k - \frac{\lambda_n}{r^2} \bar{u}_n^k = \bar{g}_n^k(r) + \bar{f}_n^k(r, t), \quad (3.7)$$

where $\overline{f}_n^k(r, t)$ are determined from the previous equations of this system and, moreover, $\overline{f}_0^k(r, t) \equiv 0$. From (3.7), having done the substitution of variables $\overline{u}_n^k(r, t) = r^{(1-m)/2} u_n^k(r, t)$, and then setting $\xi = (r + t)/2$, $\eta = (r - t)/2$, we obtain

$$Mu \equiv u_{n\xi\eta}^k + \frac{[(m-1)(3-m) - 4\lambda_n]}{4(\xi + \eta)^2} u_n^k = g_n^k(\xi + \eta) + f_n^k(\xi, \eta), \quad (3.8)$$

$$g_n^k(\xi + \eta) = (\xi + \eta)^{(m-1)/2} \overline{g}_n^k(\xi + \eta), \quad f_n^k(\xi, \eta) = (\xi + \eta)^{(m-1)/2} \overline{f}_n^k(\xi + \eta, \xi - \eta),$$

and, given this result, the boundary-value conditions (3.6) will take the form

$$\begin{aligned} u_n^k(\xi, \xi) &= \tau_n^k(\xi), \quad \left(\frac{\partial u_n^k}{\partial \xi} - \frac{\partial u_n^k}{\partial \eta} \right) \Big|_{\xi=\eta} = v_n^k(\xi), \quad \frac{\varepsilon}{2} \leq \eta < \xi \leq \frac{1}{2}, \\ u_n^k(\xi, \alpha\xi + \gamma) &= \sigma_n^k(\xi), \quad \frac{\varepsilon}{2\beta} \leq \xi \leq \frac{1}{2}, \\ \tau_n^k(\xi) &= (2\xi)^{(m-1)/2} \overline{\tau}_n^k(2\xi), \\ v_n^k(\xi) &= \sqrt{2}(2\xi)^{(m-1)/2} \overline{v}_n^k(2\xi), \\ \sigma_n^k(\xi) &= [(1 + \alpha)\xi + \gamma]^{(m-1)/2} \overline{\sigma}_n^k \\ &\quad \times ((1 + \alpha)\xi + \gamma), \\ 0 < \alpha &= \frac{1 - \beta}{1 + \beta} < 1, \quad \gamma = \frac{\varepsilon}{1 + \beta}, \quad k = \overline{1, k_n}, \quad n = 0, 1, \dots \end{aligned} \quad (3.9)$$

Using the general solution of the equation (3.8) (see [20]), it is easy to show that the solution of the Cauchy problem for the equation (3.8) takes the form

$$\begin{aligned} u_n^k(\xi, \eta) &= \frac{1}{2} \tau_n^k(\eta) R(\eta, \eta; \xi, \eta) + \frac{1}{2} \tau_n^k(\xi) R(\xi, \xi; \xi, \eta) \\ &\quad + \frac{1}{\sqrt{2}} \int_{\eta}^{\xi} \left[v_n^k(\xi_1) R(\xi_1, \xi_1; \xi, \eta) - \tau_n^k(\xi_1) \frac{\partial}{\partial N} R(\xi_1, \eta_1; \xi, \eta) \Big|_{\xi_1=\eta_1} d\xi_1 \right] \\ &\quad + \int_{1/2}^{\xi} \int_{\varepsilon/2}^{\eta} \left[g_n^k(\xi_1 + \eta_1) + f_n^k(\xi_1, \eta_1) \right] R(\xi_1, \eta_1; \xi, \eta) d\xi_1 d\eta_1, \end{aligned} \quad (3.10)$$

where $R(\xi_1, \eta_1; \xi, \eta) = P_{\mu}[\frac{(\xi_1 - \eta_1)(\xi - \eta) + (\xi_1 \eta_1 + \xi \eta)}{(\xi_1 + \eta_1)(\xi + \eta)}] = P_{\mu}(z)$ is the Riemann function of the equation $Mu = 0$ (see [21]) and $P_{\mu}(z)$ is the Legendre function, $\mu = n + (m - 3)/2$, $(\partial/\partial N)|_{\xi=\eta} = (1/\sqrt{2})((\partial/\partial \xi) - (\partial/\partial \eta))|_{\xi=\eta}$.

From (3.10), for $\eta = \alpha\xi + \gamma$, using the boundary-value conditions (3.9), we obtain

$$\varphi_n^k(\xi) = \int_{1/2}^{\xi} \int_{\varepsilon/2}^{\alpha\xi + \gamma} g_n^k(\xi_1 + \eta_1) R(\xi_1, \eta_1; \xi, \alpha\xi + \gamma) d\xi_1 d\eta_1, \quad \frac{\varepsilon}{2} \leq \xi \leq \frac{1}{2}, \quad (3.11)$$

where

$$\begin{aligned}
\varphi_n^k(\xi) &= \sigma_n^k(\xi) - \frac{\tau_n^k(\alpha\xi + \gamma)}{2} R(\alpha\xi + \gamma, \alpha\xi + \gamma; \xi, \alpha\xi + \gamma) - \frac{\tau_n^k(\xi)}{2} R(\xi, \xi; \xi, \alpha\xi + \gamma) \\
&\quad - \frac{1}{\sqrt{2}} \int_{\alpha\xi + \gamma}^{\xi} \left[v_n^k(\xi_1) R(\xi_1, \xi_1; \xi, \alpha\xi + \gamma) - \tau_n^k(\xi_1) \frac{\partial}{\partial N} R(\xi_1, \eta_1; \xi, \alpha\xi + \gamma) \Big|_{\xi_1 = \eta_1} d\xi_1 \right] \\
&\quad - \int_{1/2}^{\xi} \int_{\varepsilon/2}^{\alpha\xi + \gamma} f_n^k(\xi_1 + \eta_1) R(\xi_1, \eta_1; \xi, \alpha\xi + \gamma) d\xi_1 d\eta_1,
\end{aligned} \tag{3.12}$$

which, after the double differentiation reduces to the following loaded Volterra integral equation of the second kind (see [22]):

$$\begin{aligned}
&\frac{1}{2\alpha} \frac{d^2 \varphi_n^k}{d\xi^2} g_n^k((\alpha + 1)\xi + \gamma) + \frac{1}{2\alpha} \int_{\varepsilon/2}^{\alpha\xi + \gamma} \frac{\partial}{\partial \xi} \left[g_n^k(\xi_1 + \xi) R(\xi_1, \xi; \xi, \alpha\xi + \gamma) d\xi_1 \right. \\
&\quad \left. + \frac{1}{2\alpha} \int_{\varepsilon/2}^{\alpha\xi + \gamma} g_n^k(\xi_1 + \xi) \frac{\partial}{\partial \xi} R(\xi_1, \eta_1; \xi, \alpha\xi + \gamma) \Big|_{\eta_1 = \xi} d\xi_1 \right. \\
&\quad \left. + \frac{1}{2} \int_{1/2}^{\xi} \frac{\partial}{\partial \xi} \left[g_n^k(\alpha\xi + \gamma + \eta_1) R(\alpha\xi + \gamma, \eta_1; \xi, \alpha\xi + \gamma) d\eta_1 \right. \right. \\
&\quad \left. \left. + \frac{1}{2} \int_{1/2}^{\xi} g_n^k(\alpha\xi + \gamma + \eta_1) \frac{\partial}{\partial \xi} R(\xi_1, \eta_1; \xi, \alpha\xi + \gamma) \Big|_{\xi_1 = \alpha\xi + \gamma} d\eta_1 \right. \right. \\
&\quad \left. \left. + \frac{1}{2\alpha} \int_{1/2}^{\xi} \int_{\varepsilon/2}^{\alpha\xi + \gamma} g_n^k(\xi_1 + \eta_1) \frac{\partial^2}{\partial \xi^2} R(\xi_1, \eta_1; \xi, \alpha\xi + \gamma) d\xi_1 d\eta_1. \right. \right.
\end{aligned} \tag{3.13}$$

It is handy to rewrite (3.13) as

$$g_n^k(\xi) = F(g_n^k). \tag{3.14}$$

Given that $|P_\mu(z)| \leq C$, $|P'_\mu(z)| \leq C$, $|P''_\mu(z)| \leq C$ [23], $C = \text{const}$, the integral operator F maps the full metric space $C^1(\bar{J})$ with the norm $\|g_n^k\| = \max_{\bar{J}} |g_n^k(\xi)| + \max_{\bar{J}} |dg_n^k/d\xi|$ into itself, where J is the interval $(\varepsilon/2, 1/2)$.

Next, let g_{1n}^k and g_{2n}^k be arbitrary elements of the space $C^1(\bar{J})$. It is easy to see that for $g_n^k = g_{1n}^k - g_{2n}^k$, the following estimate is valid:

$$\begin{aligned}
|F(g_n^k)| &= \left| -\frac{1}{2\alpha} \int_{\varepsilon/2}^{\alpha\xi+\gamma} \left[(g_n^k(\xi_1 + \xi))' R(\xi_1, \xi; \xi, \alpha\xi + \gamma) + g_n^k(\xi_1 + \xi) \frac{\partial}{\partial \xi} R(\xi_1, \xi; \xi, \alpha\xi + \gamma) \right] d\xi_1 \right. \\
&\quad - \frac{1}{2\alpha} \int_{\varepsilon/2}^{\alpha\xi+\gamma} g_n^k(\xi_1 + \xi) \frac{\partial}{\partial \xi} R(\xi_1, \eta_1; \xi, \alpha\xi + \gamma) \Big|_{\eta_1=\xi} d\xi_1 \\
&\quad + \frac{1}{2} \int_{\xi}^{1/2} \left[\alpha (g_n^k(\alpha\xi + \gamma + \eta_1))' R(\alpha\xi + \gamma, \eta_1; \xi, \alpha\xi + \gamma) \right. \\
&\quad \quad \left. + g_n^k(\alpha\xi + \gamma + \eta_1) \frac{\partial}{\partial \xi} R(\alpha\xi, \eta_1; \xi, \alpha\xi + \gamma) \right] d\eta_1 \\
&\quad + \frac{1}{2} \int_{\xi}^{1/2} g_n^k(\alpha\xi + \gamma + \eta_1) \frac{\partial}{\partial \xi} R(\xi_1, \eta_1; \xi, \alpha\xi + \gamma) \Big|_{\xi_1=\alpha\xi+\gamma} d\eta_1 \\
&\quad + \frac{1}{2\alpha} \int_{1/2}^{\xi} \int_{\varepsilon/2}^{\alpha\xi+\gamma} g_n^k(\xi_1 + \eta_1) \frac{\partial^2}{\partial \xi^2} R(\xi_1, \eta_1; \xi, \alpha\xi + \gamma) d\xi_1 d\eta_1 \Big| \\
&\leq \frac{3M}{2\alpha} \|g_n^k\| \left[(\alpha\xi + \gamma - \frac{\varepsilon}{2}) + \left(\frac{1}{2} - \xi\right) + \left(\frac{1}{2} - \xi\right) (\alpha\xi + \gamma - \frac{\varepsilon}{2}) \right], \\
M &= \max \left(\max_{\bar{J}^* \bar{J}} |R|, \max_{\bar{J}^* \bar{J}} \left| \frac{\partial R}{\partial \xi} \right|, \max_{\bar{J}^* \bar{J}} \left| \frac{\partial^2 R}{\partial \xi^2} \right| \right).
\end{aligned} \tag{3.15}$$

Furthermore, it is evident that

$$|F^2(g_n^k)| \leq \left(\frac{3}{2\alpha} M \right)^2 \|g_n^k\| \left[\frac{(\alpha\xi + \gamma - (\varepsilon/2))^2}{2} + \frac{((1/2) - \xi)^2}{2} + \frac{((1/2) - \xi)^2 (\alpha\xi + \gamma - (\varepsilon/2))^2}{2} \right]. \tag{3.16}$$

Continuing this process, we obtain

$$|F^n(g_n^k)| \leq \left(\frac{3}{2\alpha} M \right)^n \|g_n^k\| \left[\frac{(\alpha\xi + \gamma - (\varepsilon/2))^n}{n!} + \frac{((1/2) - \xi)^n}{n!} + \frac{((1/2) - \xi)^n (\alpha\xi + \gamma - (\varepsilon/2))^n}{n!} \right], \tag{3.17}$$

where F^n is the n th degree of the operator F . From here, one can see that we can find such n that

$$|F^n(g_n^k)| \leq C \|f\|, \quad C = \text{const} < 1. \tag{3.18}$$

The inequality (3.18) implies that the operator F^n is a contraction.

Thus, the operator F has a fixed point [24]. This fixed point is the solution of (3.14), that is, (3.13).

Therefore, having first solved the problem (3.3), (3.6), (3.7) (for $n = 0$), and then (3.4), (3.6), (3.7) (for $n = 1$) and so on, we can find sequentially all $\bar{u}_n^k(r, t)$, $k = \overline{1, k_n}$, $n = 0, 1, \dots$.

Thus, we have shown that

$$\int_{\Gamma} \rho(\theta)(Lu - g(x))d\Gamma = 0. \quad (3.19)$$

Let $f(r, \theta, t) = R(r)\rho(\theta)T(t)$, and, moreover, $R(r) \in V_0$ is dense in $L_2(t+\varepsilon, 1-t)$, $\rho(\theta) \in C^\infty(\Gamma)$ is dense in $L_2(\Gamma)$, and $T(t) \in V_1$ is dense in $L_2(0, (1-\varepsilon)/2)$. Then, $f(r, \theta, t) \in V$, $V = V_0 \otimes \Gamma \otimes V_1$ is dense in $L_2(D_\varepsilon)$ (see, e.g., [25]).

Thus, from (3.19), it follows that

$$\int_{D_\varepsilon} f(r, \theta, t)(Lu - g(x))dD_\varepsilon = 0 \quad Lu = g(x), \quad \forall(x, t) \in D_\varepsilon. \quad (3.20)$$

Therefore, the problem (2.1)–(2.3) has the solutions of the form

$$u(r, \theta, t) = \sum_{n=0}^{\infty} \sum_{k=1}^{k_n} r^{(1-m)/2} u_n^k(r, t) Y_{n,m}^k(\theta), \quad (3.21)$$

where $u_n^k(r, t)$ are determined from (3.10), in which $g_n^k(\xi, \eta)$ are found from (3.13).

Taking into account the restrictions on the functions $\tau(r, \theta)$, $v(r, \theta)$, $\sigma(r, \theta)$, one can analogously prove (as shown, e.g., in [18, 19]) that the obtained solution $u(r, \theta, t)$ in the form (3.21) belongs to the required class.

This completes the proof of Theorem 3.1. Theorem 3.2 is proven analogously. \square

4. Conclusion

In this paper, we have proven the existence theorems for the solution of the inverse problem for a general class of multidimensional hyperbolic PDEs, for the cases of time-independent and space-independent unknown right-hand sides of the equation. The potential importance of this work comes from the fact that multidimensional inverse problems for hyperbolic equations are extremely important in applied work, but so far, only the results for relatively narrow classes of equations have been established.

Turning to the limitations of our work, we should note that so far, we have not been able show the uniqueness or stability of the solution. Given the importance of the well-posedness problem in applied fields, we believe that this is an important avenue for future work.

Acknowledgments

The authors thank two referees and Professor Carlo Cattani (Editor) for useful suggestions. They acknowledge financial support from the Belgian Science Policy Program on Inter-University Poles of Attraction (PAI-UAP P6/07) and National Bank of Belgium grant for the research project “Spatial Economic Dynamics”.

References

- [1] M. Lavrentiev, V. Romanov, and V. Vasiliev, *Multidimensional Inverse Problems for Differential Equations*, Springer, New York, NY, USA, 1970.
- [2] V. Romanov, *Some Inverse Problems for the Equations of Hyperbolic Type*, Nauka, Novosibirsk, Russia, 1972.
- [3] V. Isakov, *Inverse Problems for Partial Differential Equations*, Springer, New York, NY, USA, 2006.
- [4] 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.
- [5] S. Y. Chen, H. Tong, and C. Cattani, “Markov models for image labeling,” *Mathematical Problems in Engineering*, vol. 2011, Article ID 814356, 18 pages, 2011.
- [6] S. Y. Chen and Q. Guan, “Parametric shape representation by a deformable NURBS model for cardiac functional measurements,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 480–487, 2011.
- [7] R. Engbers, M. Burger, and V. Capasso, *Inverse Problems in Geographical Economics: Parameter Identification in the Spatial Solow Model*, University of Muenster, Muenster, Germany, 2011.
- [8] M. Carrasco, J.-P. Florens, and E. Renault, “Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization,” in *Handbook of Econometrics*, J. Heckman and E. Leamer, Eds., vol. 7, chapter 77, pp. 5633–5751, North-Holland, Amsterdam, The Netherlands, 2007.
- [9] C. A. Favero and R. Rovelli, “Macroeconomic stability and the preferences of the Fed: a formal analysis, 1961-98,” *Journal of Money, Credit and Banking*, vol. 35, no. 4, pp. 545–556, 2003.
- [10] E. Castelnuovo and P. Surico, “Model uncertainty, optimal monetary policy and the preferences of the fed,” *Scottish Journal of Political Economy*, vol. 51, no. 1, pp. 105–126, 2004.
- [11] P. Brito, *The Dynamics of Growth and Distribution in a Spatially Heterogeneous World. Working Paper*, ISEG, Technical University of Lisbon, Lisbon, Portugal, 2004.
- [12] R. Boucekine, C. Camacho, and B. Zou, “Bridging the gap between growth theory and the new economic geography: the spatial Ramsey model,” *Macroeconomic Dynamics*, vol. 13, no. 1, pp. 20–45, 2009.
- [13] J.-P. Puel and M. Yamamoto, “Generic well-posedness in a multidimensional hyperbolic inverse problem,” *Journal of Inverse and Ill-Posed Problems*, vol. 5, no. 1, pp. 55–83, 1997.
- [14] A. G. Ramm and Rakesh, “Property C and an inverse problem for a hyperbolic equation,” *Journal of Mathematical Analysis and Applications*, vol. 156, no. 1, pp. 209–219, 1991.
- [15] C. Daveau, D. Douady, and A. Khelifi, “On a hyperbolic coefficient inverse problem via partial dynamic boundary measurements,” *Journal of Applied Mathematics*, Article ID 561395, 14 pages, 2010.
- [16] S. G. Mikhlin, *Multidimensional Singular Integrals and Integral Equations*, Pergamon, New York, NY, USA, 1965.
- [17] V. Smirnov, *A Course in Higher Mathematics*, vol. 4, part 2, Addison-Wesley, Boston, Mass, USA, 1964.
- [18] S. Aldashev, *Boundary-Value Problems for Multi-Dimensional Hyperbolic and Mixed Equations*, Gylym Press, Almaty, Kazakhstan, 1994.
- [19] S. A. Aldashev, “On Darboux problems for a class of multidimensional hyperbolic equations,” *Differentsialnye Uravneniya*, vol. 34, no. 1, pp. 64–68, 1998.
- [20] A. V. Bitsadze, *Equations of the Mixed Type*, Pergamon, New York, NY, USA, 1964.
- [21] E. T. Copson, “On the Riemann-Green function,” *Archive for Rational Mechanics and Analysis*, vol. 1, pp. 324–348, 1958.
- [22] A. Nakhushhev, *The Equations of Mathematical Biology*, Nauka Press, Moscow, Russia, 1995.
- [23] H. Bateman and A. Erdelyi, *Higher Transcendental Functions*, vol. 1, McGraw-Hill, New York, NY, USA, 1955.

- [24] L. V. Kantorovich and G. P. Akilov, *Functional Analysis*, Pergamon, New York, NY, USA, 2nd edition, 1982.
- [25] A. Kolmogorov and S. Fomin, *Elements of the Theory of Functions and Functional Analysis*, Dover Publications, Mineola, NY, USA, 1999.

Research Article

Noise Estimation for Single-Slice Sinogram of Low-Dose X-Ray Computed Tomography Using Homogenous Patch

Zhiwu Liao,¹ Shaoxiang Hu,² Ming Li,³ and Wufan Chen²

¹ Visual Computing and Virtual Reality Key Laboratory Of Sichuan Province, Sichuan Normal University, Chengdu 610101, China

² School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

³ School of Information Science & Technology, East China Normal University, no. 500, Dong-Chuan Road, Shanghai 200241, China

Correspondence should be addressed to Shaoxiang Hu, hushaox@126.net

Received 29 June 2011; Accepted 21 July 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Zhiwu Liao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a new method to estimate noise for a single-slice sinogram of low-dose CT based on the homogenous patches centered at a special pixel, called center point, which has the smallest variance among all sinogram pixels. The homogenous patch, composed by homogenous points, is formed by the points similar to the center point using similarity sorting, similarity decreasing searching, and variance analysis in a very large neighborhood (VLN) to avoid manual selection of parameter for similarity measures. Homogenous pixels in the VLN allow us find the largest number of samples, who have the highest similarities to the center point, for noise estimation, and the noise level can be estimated according to unbiased estimation. Experimental results show that for the simulated noisy sinograms, the method proposed in this paper can obtain satisfied noise estimation results, especially for sinograms with relatively serious noises.

1. Introduction

With continued technology advancement and wider applications, use of computed Tomography (CT) is increasing. However radiation exposure and associated risk of cancer for patients receiving CT examination have been an increasing concern in recent years. Thus, minimizing X-ray exposure to patients has been one of the major efforts in the CT field [1–3].

A simple and cost-effective means to achieve low-dose CT applications is to lower X-ray tube current (mA) as low as achievable. However, dose reduction generally leads to

an increased level of noise in the measured projection data (sinogram) and the subsequent reconstructed images.

Sinograms acquired from low-dose CT are corrupted by many factors, including Poisson noise, logarithmic transformation of scaled measurements, and prereconstruction corrections for system calibration [4]. All these factors complicate noise modeling on the sinogram. Up to now, various noise models for sinogram have been developed [4–15].

One common model for noise estimation is Gaussian distribution with variance depending on the sinogram [4–6]. The model is developed by repeatedly acquired projection measurements of a physical phantom at a fixed angle for 900 times by a GE spiral CT scanner [5, 6].

Another model is Poisson distribution based on quantum noise for CT which is due to the limited number of photons collected by the detector [7, 8]. Although a CT detector is not based on photon-counting but energy integrated that generates a signal proportional to the total energy deposited in the detector, a photon-counting model is still a good approximation and is widely used for characterizing noise properties of the sinogram.

More accurate noise model, compound Poisson model (CPM), which takes into account both the polychromatic X-ray beam and energy integration, has been investigated in [9–11]. However, for its complex expression, the exact form of the likelihood is not amendable. Therefore, many researchers have often approximated the CPM by either the Poisson distribution or the Gaussian distribution [12, 13].

Based on the above models, dose-reduction simulation using synthetic-noise generators, enables ethical studies of low-dose procedures to improve the quality of reconstruction images. Dose-reduction simulation has been reported using Poisson or Gaussian noise models [14, 15]. However, above simulation methods assume that the parameters either for Poisson or for Gaussian are known. It is unreal in clinical X-ray CT systems. Thus in order to get satisfactory reconstruction images, noise estimation becomes a key problem in low-dose CT imaging both for sinogram and for reconstructed images.

In this paper, we focus on how to estimate parameters for added signal independent Gaussian noise (SIGN) on sinogram. Although it does not coincide noise models introduced in this section, it is still a valuable model in investigating the properties of the noise for sinogram and the relations between common noise models for sinogram which will be discussed in Section 2.

Noise estimation for a single image is an important and difficult problem in image denoising for complex structures of images and is studied in the early 90s for the last century [16–19]. The main start point for these methods is that the noise level should be estimated using the smoother versions of images. However, it leads to their high computation burden and overestimate of noise levels.

Recently, noise estimation from a single image both for Gaussian and Poisson noise becomes a hot spot in computer vision [20–24]. However, most of these methods must pose complex prior such as, camera parameters and so forth [21, 22], or have many parameters chosen by hand [20, 24] or segment images previously [20, 23] which hampers their application in noise estimation for sinogram.

In this paper, we insist that noise estimation should be performed in a local homogenous patch and present a method, which does not need presegment or pose complex prior. Moreover, in order to get more reliable estimated results and to improve the overestimate when noise levels are low, motivated by nonlocal means and some most recent results [25–41], a very large neighborhood is adopt to find similar points. Nonlocal means estimates each real gray level of an image based on block similarity in

a nonlocal neighborhood with size 21×21 and very recently is used in low-dose CT imaging [25–32].

In the very large neighborhood, the homogenous patches are determined by finding similar points to the center point through similarity sorting, decreasing similarity searching and variance analysis. The noise parameters will be estimated on this homogenous patch using unbiased estimate. Since very large neighborhood provides more reliable estimation, proposed method can get satisfied noise estimation results.

The remainder of this paper is arranged as the noise models will be discussed in Section 2; the motivation and method will be given in Sections 3 and 4, respectively; and then the experimental results will be given in Section 5; finally, it is the conclusion, future works, and acknowledgment.

2. Noise Models

For clinical X-ray system, its detected that X-ray intensity follows a compound Poisson distribution [9–11]. However, for its complex expression, it has no analytic formula for the likelihood function. Various approximations have been proposed for it. Two common models include Poisson distribution and Gaussian distribution.

In this section, we will introduce signal-independent Gaussian noise (SIGN), Poisson noise, and signal-dependent Gaussian noise, as well as their relations and the reason for addressing SIGN.

2.1. Signal-Independent Gaussian Noise (SIGN)

SIGN is a common noise for imaging system. Poisson noise and signal-dependent Gaussian noise can be converted to SIGN using scale transforms which will be discussed in Section 2.3.

Let the original projection data be $\{x_i\}$, $i = 1, \dots, m$, where i is the index of the i th bin. The signal has been corrupted by additive noise $\{n_i\}$, $i = 1, \dots, m$, and one noisy observation

$$y_i = x_i + n_i, \quad (2.1)$$

where n_i is an observation for the random variable N_i as normal $N(0, \sigma_N^2)$ and independent to the Gaussian random variable X_i where the uppercase letters denote the random variables and the lower-case letters denote the observations for respective variables.

2.2. Poisson Model and Signal-Dependent Gaussian Model

The photon noise is due to the limited number of photons collected by the detector [42]. For a given attenuating path in the imaged subject, $N_0(i, \alpha)$ and $N(i, \alpha)$ denote the incident and the penetrated photon numbers, respectively. Here, i denote the index of detector channel or bin and α is the index of projection angle. In the presence of noises, the sinogram should be considered as a random process and the attenuating path is given by

$$r_i = -\ln \left[\frac{N(i, \alpha)}{N_0(i, \alpha)} \right], \quad (2.2)$$

where $N_0(i, \alpha)$ is a constant and $N(i, \alpha)$ is Poisson distribution with mean N .

Thus we have

$$N(i, \alpha) = N_0(i, \alpha) \exp(-r_i). \quad (2.3)$$

Both its mean value and variance are N .

Gaussian distributions of ployenergetic systems were assumed based on limited theorem for high-flux levels and followed many repeated experiments in [6]; we have

$$\sigma_i^2(\mu_i) = f_i \exp\left(\frac{\mu_i}{\gamma}\right), \quad (2.4)$$

where μ_i is the mean and σ_i^2 is the variance of the projection data at detector channel or bin i , γ is a scaling parameter and f_i is a parameter adaptive to different detector bins.

The most common conclusion for the relation between Poisson distribution and Gaussian distribution is that the photon count will obey Gaussian distribution for the case with large incident intensity and Poisson distribution with feeble intensity [6]. In addition, in [42], the authors deduce the equivalency between Poisson model and Gaussian model. Therefore, both theories indicate that these two noises have similar statistical properties and can be unified into a whole framework.

2.3. Scale Transformations

As a statistical method for treating nonnormally distributed or signal dependent noise, scale transformations are widely-used to stabilize the variance [6, 31]. In [4], authors indicate that the variance of their signal-dependent Gaussian model represented in (2.4) can be stabilized by

$$g(x) = \frac{c}{k} \ln(kx + b) + c_1, \quad (2.5)$$

where c is the expected constant standard deviation for transformed data and c_1 is an arbitrary constant or

$$g(x) = \frac{2c}{\sqrt{4ac - b^2}} \arctan \frac{2ax + b}{\sqrt{4ac - b^2}} + c_2, \quad (2.6)$$

where c_2 also is an arbitrary constant.

Armando Manduca et al. also indicate that the Anscombe transform can convert Poisson distribution data to data with an approximately normal distribution with a constant variance. Thus we can use $Q = 2\sqrt{N} + 3/8$ to convert data with distribution expressed in (2.3) to a normal distribution with a constant variance.

Since both Poisson noise and signal-dependent Gaussian noise can be converted to SIGN, the noise estimation for low-dose CT can start from SIGN to focus on our new method itself.

3. Backgrounds and Motivation

Following the SIGN model discussed in Section 2.1, we will introduce backgrounds and motivation for proposed method.

Noise estimation for low-dose CT projection data is an estimation problem. Using the terms in Section 2.1, the goal for SIGN estimation is to get unbiased estimated value $\hat{\sigma}_N$ of σ_N from the noisy observation y_i . Here σ_N is the parameter for population distribution of the noise.

However, only one noisy observation y_i is provided. Thus we have to share statistical information each other. One reasonable way for sharing statistical information is to assume independent identical distributions (iid) for nearby similar pixels among noisy projection data (sinogram). In this paper, the similar point is the point with high block similarity to the center point. This section will discuss the motivation for finding samples of unbiased estimator.

3.1. Unbiased Estimation

Suppose we have a statistical model parameterized by θ , and a statistic $\hat{\theta}$ which serves as an estimator of θ based on any samples x . That is, we assume that population data follows Gaussian distribution with a fixed and unknown constant θ , and then we construct some estimator $\hat{\theta}$ that maps samples to values that we hope are close to θ . Then the bias of this estimator is defined to be

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - E(\theta) = E(\hat{\theta} - \theta), \quad (3.1)$$

where $E(\cdot)$ denotes expected value over the distribution, that is, averaging overall possible samples x .

An estimator is said to be unbiased if its bias is equal to zero for parameter θ . Suppose $x = \{x_1, \dots, x_n\}$ are observations for a Gaussian random variable with expectation μ and variance σ , the unbiased estimators for these samples are

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n x_i, \\ S^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2. \end{aligned} \quad (3.2)$$

In theory, the sample estimators shown in (3.2) are asymptotically unbiased and efficient for the sample size is moderate or large. Thus unbiased estimation requires providing enough samples to approximate the population distribution.

3.2. Measure Similarity

In order to measure similarity between two different points $x_{i,j}$ and $x_{s,t}$ of noisy sinogram where $i \neq s$ or $j \neq t$, we should measure two 7×7 blocks centered at $x_{i,j}$ and $x_{s,t}$, respectively. The similarity is measured as

$$S[(i, j), (s, t)] = \sum_{k=-3}^{k=3} \sum_{m=-3}^{m=3} [x_{i+k, j+m} - x_{s+k, t+m}]. \quad (3.3)$$

Generally, we can predefine a threshold T to find the similar points of $x_{i,j}$ in a nonlocal neighborhood with 21×21 centered at $x_{i,j}$

$$\begin{aligned} x_{s,t} \text{ is a similar point of } x_{i,j} & \text{ if } S\{(i, j), (s, t)\} \leq T, \\ x_{s,t} \text{ is not a similar point of } x_{i,j} & \text{ if } S\{(i, j), (s, t)\} > T, \end{aligned} \quad (3.4)$$

where $s \neq i$ or $t \neq j$ and $s, t = -10, \dots, 10$.

3.3. Motivation

Since only samples with moderate or large size can approximate the population distribution well, we must find enough samples to estimate the noise correctly. Moreover, iid assumption requires to share statistical information among similar points of noisy sinogram. Both requirements coincide with the start point of existing methods. That is, noise estimation should be performed at smoother versions of the images to suppress the influence of the complex structures of images.

However, estimation noise using smoother versions of the images has high computation burden and overestimate noise levels [24]. In this paper, we propose a method to estimate noise in a VLN by similarity sorting and variance analysis.

Just as discussed in this section, the key objective is to find enough similar points for noise estimation. Unlike existing global methods, which estimate noise levels using whole images, the proposed method try to estimate noise levels in a more local way to reduce computation burden and increase flexibility for estimation.

In order to accomplish the above objectives (enough samples with a more local structure), VLN is used to find similar points (samples). Moreover, the VLN should be put in a suitable position to ensure the enough samples can be found. That is, VLN should be put in a homogenous region. Since the center of VLN can determine the position of VLN, how to put VLN in a suitable position can be converted to how to locate the center point in a homogenous region.

The noise level of low-dose CT is low. In this situation, only variance is enough for describing the local homogeneity roughly. That is, large variance relates to a square near singularities while small variance relates to a homogenous square. Therefore, by comparing variances of squares with fixed size for all pixels in sinogram, the center point is defined as the pixel with the smallest variance among all pixels.

After determining the center point, we propose a new method for noise estimation based on similarity sorting, similarity decreasing searching, and variance analysis. Its main start point is from how to avoid threshold setting in finding similar points to the center point

since the threshold selection depends on the noise level which makes it a “chicken and egg” problem.

Similarity sorting provides a similarity decreasing sequence (SDS). Thus the samples must be formed by the up at the front points in the SDS since points with smaller similarities to the center point maybe the outliers for the estimation.

In order to find the largest number samples with the highest similarities in the VLN, similarity decreasing searching combining variance analysis is used. That is, the samples are formed by adding 50 points each time according to the order of the SDS. Thus it ensures that each addition adds the points with the highest similarities in residual SDS.

Moreover, in order to find the largest number samples used in estimation, we must find when the outliers are added. It can be achieved by variance analysis for each addition. When the variance for an addition becomes large suddenly, it means some outliers are added. Therefore, the real samples are the last before this addition. The detailed algorithm will be introduced in Section 4 and two-number examples can be found in Section 5.3.

4. The Method

In this section, we will give the framework for noise estimation. Just as shown in Figure 1, it includes three steps.

- (1) Find the center point: It is the key step to locate the VLN, which should be put in a homogenous region. Thus the center point is defined as the point with the smallest local variance in all sinogram pixels.
- (2) Determine the homogenous patch: The homogenous patch is composed by similar points to the center point (samples), and these similar points are searched using similarity sorting, similarity decreasing searching and variance analysis.
- (3) Estimate noise: The parameters are estimated using unbiased estimator by the samples on the homogenous patch.

By these steps, the largest number of samples with the highest similarities can be obtained. These samples form a very large homogenous patch for noise estimation. Thus even using simple unbiased estimate, satisfied results can be obtained. The details for these three steps will be given in the remainder of this section.

4.1. Find Center Pixel

The center pixel should locate at the center of a homogenous region. However, in noise, finding a large *regular* homogenous region correctly and putting a pixel of sinogram in the center of this region is not a trivial task.

Motivated by [27], irregular pixel patch composed by similar pixels to the center point is a reasonable choice. Thus the choice for the center point becomes an easy task which only needs to ensure it is the center of a small homogenous square.

Since the variance can describe how far the numbers lie from the mean, small variances will relate to homogenous squares while large variances will relate to squares near singularities. Thus the center point is chosen as the pixel with the smallest variance in a 7×7

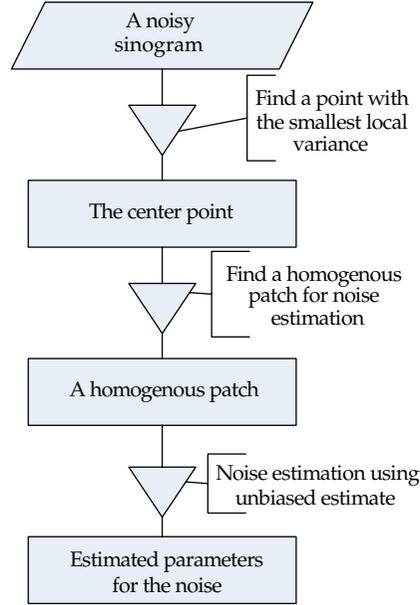


Figure 1: The flow chart for the proposed method.

square centered at the point among all pixels of the sinogram. That is, the position of the center point should be

$$(i, j) = \arg \min_{\substack{(i,j) \\ x_{i,j} \in X}} \sum_{m=-3}^{m=3} \sum_{n=-3}^{n=3} (x_{i-m,j-n} - \bar{x}_{i,j})^2, \quad (4.1)$$

where X is the sinogram and

$$\bar{x}_{i,j} = \frac{1}{49} \sum_{s=-3}^{s=3} \sum_{t=-3}^{t=3} x_{i-s,j-t} \quad (4.2)$$

is the mean.

By this way, we can find a point centered at a homogenous square and then extend it to an irregular homogenous patch.

4.2. Determine Similar Points (Homogenous Patch)

After finding the center point $x_{i,j}$, the irregular homogenous patch P composed by the similar points to $x_{i,j}$ should be determined.

Each similar point is found by computing the similarity between itself and $x_{i,j}$ using (3.3). However, unlike existing methods which use a predefined threshold value T shown in Section 4.1, our method proposes a different scheme by similarities sorting and variance analysis in a VLN.

The main advantage for this scheme is that it avoids threshold setting in finding similar points. Since the threshold should be set according to different noise level, noise estimation and threshold setting become a “chickens and eggs” question.

The most straight motivation for the proposed method is that the small variance indicates only the homogenous samples, while the large variance indicates not only the homogenous samples but also outliers. Thus the variances can be considered as a sign for outliers. It reminds us that if we sort the points according to their similarities to $x_{i,j}$ in a VLN and then compute the variances for increasing adding points until the big variance is reached, the most reliable noise estimation can be obtained by the as large as possible number for the samples.

Thus we can add samples according to the fact that samples with bigger similarities will be added earlier, and compute the variance after each addition. When the added variance become big suddenly, it means that there are some outliers that are mixed in estimation. Thus the real noise level is estimated using the last before this addition.

In summary, the steps for finding similar points are as follows.

- (1) Compute the similarity between the center point $X_{i,j}$ and each point in the 41×41 square using (3.3).
- (2) Sort the similarities to form a similarity descending sequence (SDS).
- (3) Add 50 samples each time to form collections of samples and then estimate the standard deviations (SDs) of the collection of samples to form an SDs sequence.
- (4) Compute the variance difference using the $i + 1$ th and i th, $i = 1, \dots, 31$, elements of the SDs sequence to form a difference variances sequence and from the 21 addition, find the variance becomes big suddenly (is bigger than 5), and denote its index k .
- (5) The estimated variance of the noise is $k - 1$ th element of the SDs sequence obtained on the step 3.

5. Experimental Results

In this section, we will compare our method to some well-known noise estimation methods. Just as discussed in Section 1, most of existing methods for noise estimation are not suitable for sinogram for their complex prior, segmentation for images or practical setting parameters. Only methods which estimate noise by filtering images to suppress singularities previously can be used for noise estimation of sinogram [16–19, 24]. We will introduce two of these methods firstly in Section 5.2 and then compare them with the proposed method in Section 5.3.

5.1. Data

Five test images are used in this paper: a thorax phantom acquired from a GE HiSpeed multislice CT scanner (see Figure 2(a)), two phantom data produced on Matlab (see Figures 2(b) and 2(c)), and two images acquired with a 16 multidetector row CT unit (Somatom Sensation 16; Siemens Medical Solutions) using 120 kVp, 5 mm slice thickness (see Figures 2(d) and 2(e)).

For Figure 2(a), the distance from the center of rotation (COR) to the curved detector is 408.075 mm. The detector array is on an arc concentric to the X-ray source with a distance

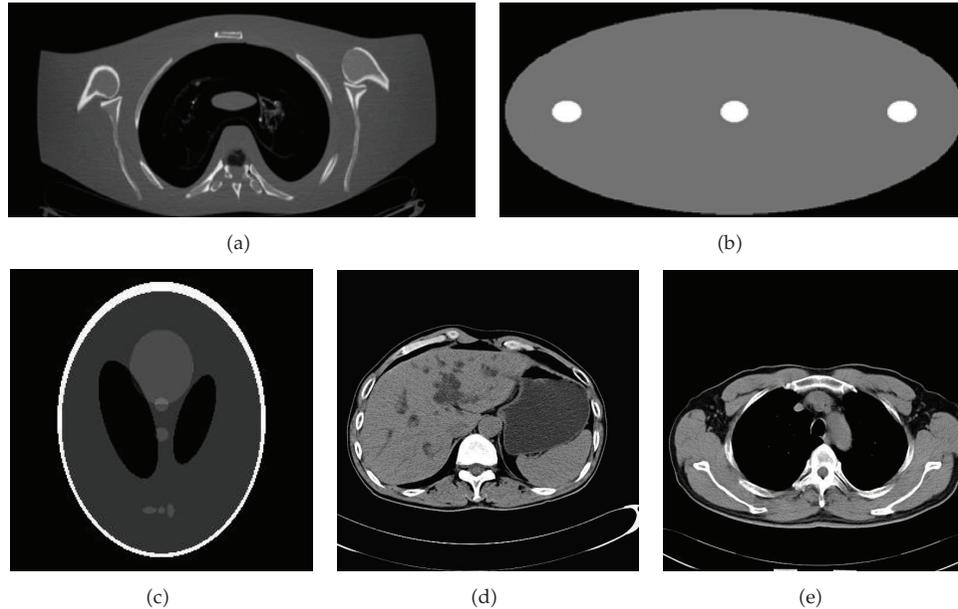


Figure 2: Test images. (a) Thorax phantom data acquired at 120 kVp, 150 mA. (b) A phantom data produced by Matlab. (c) Modified Shepp-Logan head phantom whose size is 256×256 . (d) A CT image of an abdomen of a 58-year-old man with 120 kVp, 150 mAs. (e) A CT image of a chest of a 62-year-old woman with 120 kVp, 150 mAs.

between the X-ray source and the COR of 541.00 mm. The detector cell spacing is 1.0239 mm, and the slice thickness is 1.0 mm. Its projection data is shown on Figure 3(a).

The scanning parameters for Figures 2(d) and 2(e) are: gantry rotation time, 0.5 second; table feed per gantry rotation, 24 mm; and pitch, 1 : 1. The CT doses were controlled by a fixed tube currents 150 mAs. Two respective sinogram data are shown on Figures 3(d) and 3(e).

In addition, two projection data for Figures 2(b) and 2(c) are shown in Figures 3(b) and 3(c), respectively. The projection data are produced in Matlab using radon transform for the equally distributed 90 angles for two phantoms.

The SIG noises are added on the projection data in Figure 3 directly. Since the interesting SIG noises are with small variances (SIG noises with large variances are not accepted in clinic situations), the variances are form 25 to 225 (the standard deviations are from 5 to 15).

5.2. Comparison Methods

Noise estimation is not a trivial task in image processing because of the complex structures for images. Some researchers suggest that the noise levels should be estimated by filtering images previously to suppress the image structures [17, 18, 24]. The methods compared with the proposed method are as follows.

Wavelets

This method suppresses image structures using wavelets [18]. Since coefficients in the HH subbands of wavelets only preserving high frequency energy for images, the image

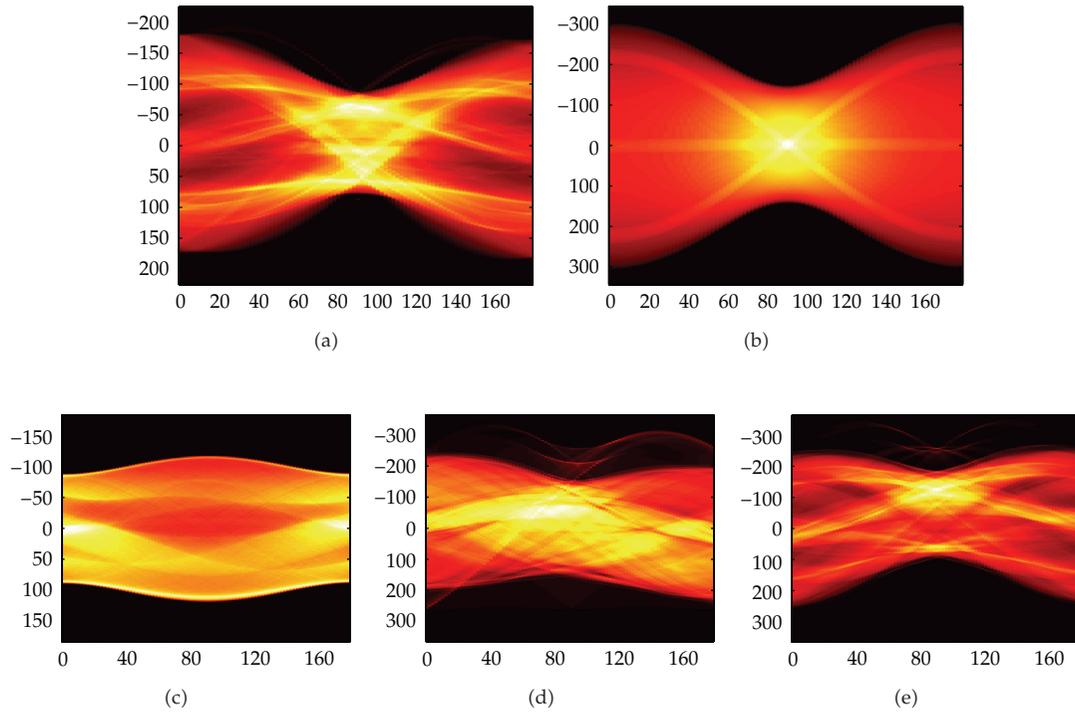


Figure 3: The relative projection data for the test images.

structures, which mainly reside in low frequency coefficients, can be considered to be suppressed in the HH subbands. Thus based on robust median estimator, the level of the noise can be estimated as

$$\sigma_N = \frac{\text{median}(|W_{HH}(X)|)}{0.6745}, \tag{5.1}$$

where X is the image, W is the wavelet operator, HH indicates the coefficients in the HH subbands, $|\cdot|$ is the absolute value, and median is the median estimator.

Fast Estimation (FE)

This method is proposed in [17]. It only has two steps for noise estimation: the first step is filtering the image using Laplacian operator and the noise level is estimated using

$$\sigma_N = \sqrt{\frac{\pi}{2}} \frac{1}{6(w-2)(h-2)} \sum |X \otimes L|, \tag{5.2}$$

where w and h are the width and height of the image, respectively, \otimes is the convolution operator, L is the Laplacian operator:

$$\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}. \quad (5.3)$$

5.3. Experimental Results

In this section, three methods, proposed method (PM), wavelets, and Fast estimation (FE), are compared. In order to compare these three methods on a fair stage, the parameters used in these methods are fixed for all noise levels and all test images. According to this standard, the method proposed recently in [24] is not included because one of its parameter must be adjusted and selected with the assumption that the real noise levels are known.

The wavelet used for wavelets is Symlets with support 4. In summary, the parameters for PM are the squares used for finding the center point which are 7×7 squares (see Section 4.1); the size of the square computing similarity between the considering pixel and the center point is 7×7 (see Section 3.2); the size for very large neighborhood is 41×41 (see Section 4.2 step 1); adding 50 samples each time to form the SDs sequence (see Section 4.2 step 3); the threshold of absolute difference between adjacent standard deviations for finding the variances become suddenly 5; in order to avoid the influence of the outliers, the comparison for the variance is from the 21th addition for samples (see Section 4.2 step 4).

Firstly, in order to show the role of variance analysis for the SDs sequence, we will use two Figures, Figures 2(c) and 2(d), for analyzing the SDs sequence (see Table 1), where "Real SD" is the real standard deviation (SD) for the added noise, "No" is the times adding samples, "ESD" is the estimated standard deviation for each sample adding, "Diff" is the difference between the i th and $i-1$ th estimated variances, the bold digits indicate the difference between two adjoint variances beyond the 5 (bigger diff), and the italic digits are the final estimated SDs.

According to the steps in Subsection 4.2, if the index for bigger diff is i , the final estimated SD is the $i-1$ th SD. For example, in the second column, which is Figure 2(c) added noise with real SD 5, we can find that the bigger diff is 19.68 and its index is 26 (see the last no blank row of the first and second columns). Thus the index for final estimated SD is 25 represented by an italic digit, and the SD is 4.90 (see the first and second columns with no. 25).

The final estimated SDs can be get through four steps (see Section 4.2). The following two tables (Tables 2 and 3) give the final estimated SDs for sinograms in Figure 3 added with noise whose SDs are from 5 to 15.

Table 2 gives the final estimated SDs for phantoms (see Figures 3(a)–3(c)), "Real SD" is the real standard deviation (SD) for the added noise, and "PM," "W," and "FE" are the abs. of the proposed method, wavelets [18], and fast estimation [17], respectively. The bold digits represent the best estimate using three estimators while the digits with * are the worst estimate. In order to ignore small differences between two estimators, two digits will not be signed if the difference between two biases (difference of estimated SDs and the real SD) is

Table 1: Variance analysis to estimate the noise level.

Real SD NO.	Figure 2(c)						Figure 2(d)					
	5		10		15		5		10		15	
	ESD	Diff	ESD	Diff	ESD	Diff	ESD	Diff	ESD	Diff	ESD	Diff
21	4.93	0.26	9.82	1.07	14.67	0.61	4.95	0.52	9.65	-0.26	14.42	3.75
22	4.94	0.10	9.86	0.78	14.78	3.00	4.96	0.15	9.69	0.90	14.45	0.95
23	4.91	-0.29	9.82	-0.79	14.74	-0.98	4.98	0.19	9.75	1.20	14.50	1.53
24	4.88	-0.29	9.75	-1.42	14.62	-3.48	4.98	0.19	9.75	1.20	14.52	0.34
25	4.90	0.22	9.81	1.19	14.71	2.63	5.13	0.80	9.82	0.53	14.68	4.80
26	6.61	19.68	10.76	19.64	15.34	18.71	5.20	0.69	9.94	2.42	14.77	2.42
27							5.25	0.51	10.02	1.52	14.87	3.17
28							5.30	0.54	10.04	0.42	14.89	0.53
29							5.94	7.24	10.35	6.17	14.98	2.73
30											15.32	10.38

Table 2: Estimated noise levels for the phantoms.

Figure 3 Real SD	(a)			(b)			(c)		
	PM	W [18]	FE [17]	PM	W [18]	FE [17]	PM	W [18]	FE [17]
5	5.57*	4.93	5.31	4.90	4.91	6.11*	5.05	4.98	5.40*
6	6.51*	5.91	6.30	5.88	5.89	7.10*	6.03	5.93	6.39*
7	7.45*	6.87	7.28	6.86	6.87	8.10*	7.03	6.89	7.38*
8	8.40*	7.84	8.27	7.84	7.85	9.10*	8.02	7.85	8.38*
9	9.37*	8.81	9.26	8.82	8.83	10.08*	9.02	8.82	9.37*
10	10.35*	9.77	10.25	9.81	9.81	11.08*	10.01	9.79	10.37*
11	11.33*	10.75	11.24	10.79	10.80	12.07*	11.01	10.76	11.37*
12	12.26	11.72	12.23	11.77	11.79	13.07*	12.00	11.75	12.37*
13	13.24	12.67	13.22	12.75	12.77	14.06*	13.00	12.73	13.37*
14	14.24	13.64*	14.21	13.73	13.75	15.06*	14.00	13.69	14.37*
15	15.41	14.60	15.20	14.71	14.74	16.06*	14.99	14.69	15.38*

smaller than 0.05. For example, in (b), “PM” and “W” have very similar estimated results, they are not signed.

From Table 2, we can see that “PM” and “W” have very similar estimated results for (b), while “FE” has the worst estimated results in all noise levels; estimated results using “PM” are the best in most of noise levels, especially in relatively serious noises of (c) and the worst estimator is “FE” also, which overestimates noise levels for both subfigures.

Table 3 gives the final estimated SDs for real projection data (see Figures 3(d)-3(e)). The all signs on Table 3 are the same as they are on Table 2.

From Table 3, we also can find that “PM” and “W” have very similar estimated results, while “FE” is the worst estimator in all noise levels for its overestimated SDs. For (e), “PM” has better performance in relatively serious noises, while “W” has better performance in relatively low noises.

It should be indicated that the estimated results using “PM” on (a) are not satisfied especially in low noises. It reminds us, that there are also future works which can be done for improving the estimated results for the proposed method which will be discussed in Section 7.

Table 3: Estimated noise levels for the real projection data.

Figure 3 Real SD	(c)			(d)		
	PM	W [18]	FE [17]	PM	W [18]	FE [17]
5	5.30	5.22	6.12*	5.01	5.07	5.83*
6	6.23	6.18	7.08*	6.21	6.05	6.80*
7	7.17	7.18	8.04*	7.11	7.03	7.77*
8	8.12	8.16	9.01*	8.06	8.01	8.75*
9	9.08	9.14	10.00*	9.04	8.99	9.73*
10	10.04	10.14	10.97*	10.00	9.99	10.71*
11	11.25	11.13	11.95*	11.09	10.97	11.70*
12	12.15	12.12	12.93*	12.08	11.95	12.69*
13	13.10	13.11	13.92*	13.04	12.93	13.68*
14	14.03	14.11	14.91*	13.99	13.91	14.67*
15	15.09	15.11	15.90*	14.98	14.92	15.66*

6. Conclusion

In this paper, we propose a new method to estimate noise for sinograms of low-dose CT. The proposed method can obtain estimated results both for phantoms and real projection data, especially in relatively serious noises, which demonstrate its potential for noise estimation of sinograms of low-dose CT.

Based on the similarity sorting and variance analysis in a very large neighborhood whose scale is 41×41 , we can find enough similar samples to obtain reliable estimated results which make this proposed local method have very similar estimated results to the best existing global methods.

In addition, avoiding convolution for suppressing image structures and relatively homogenous local structure makes the proposed method also be easily generalized to the more complex noises, such as, Poisson noise and Gaussian compound noise. Thus the proposed method is also a promising method for real sinograms of low-dose CT.

7. Future Works

Although this paper proposes a new powerful method for simulated sinogram noise estimation, it may be improved as follows.

- (1) How to find a center point in a large homogenous patch to ensure that there are enough points to obtain reliable estimation. We try to use multiresolution method to solve this problem.
- (2) How to determine the number of the similar points according to the size of the samples and farther variance analysis.
- (3) How to generalize the framework to more complex noise estimation, such as Poisson noise or Poisson and Gaussian compound noise.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (nos. 60873102, 60873264, and 61070214), National Key Basic Research Program Project of China

(nos. 2010CB732501, 2011CB302801/2011CB302802), China Postdoctoral Science Foundation of China (no. 20070410843), and Open Foundation of Visual Computing and Virtual Reality Key Laboratory Of Sichuan Province (no. J2010N03).

References

- [1] O. W. Linton, A. Fred, and F. A. Mettler, "National conference on dose reduction in CT, with an emphasis on pediatric patients," *American Journal of Roentgenology*, vol. 181, no. 2, pp. 321–329, 2003.
- [2] Y. Zhang, J. Zhang, and H. Lu, "Statistical sinogram smoothing for low-dose CT with segmentation-based adaptive filtering," *IEEE Transactions on Nuclear Science*, vol. 57, no. 5, part 1, pp. 2587–2598, 2010.
- [3] H. Lu, X. Li, I. T. Hsiao, and Z. Liang, "Analytical noise treatment for low-dose CT projection data by penalized weighted least-square smoothing in the K-L domain," in *Proceedings of the SPIE Medical Imaging*, vol. 4682, pp. 146–152, San Diego, Calif, USA, 2002.
- [4] O. Demirkaya, "Reduction of noise and image artifacts in computed tomography by nonlinear filtration of the projection images," in *Proceedings of the SPIE Medical Imaging*, vol. 4322, no. 2, pp. 917–923, San Diego, Calif, USA, 2001.
- [5] J. Wang, H. Lu, Z. Liang et al., "An experimental study on the noise properties of x-ray CT sinogram data in Radon space," *Physics in Medicine and Biology*, vol. 53, no. 12, pp. 3327–3341, 2008.
- [6] H. Lu, I.-T. Hsiao, X. Li, and Z. Liang, "Noise properties of low-dose CT projections and noise treatment by scale transformations," in *Proceedings of the IEEE Nuclear Science Symposium Conference Record*, vol. 3, pp. 1662–1666, 2001.
- [7] J. Hsieh, "Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise," *Medical Physics*, vol. 25, no. 11, pp. 2139–2147, 1998.
- [8] K. Sauer and B. Liu, "Non-stationary filtering of transmission tomograms in high photon counting noise," *IEEE Transactions on Medical Imaging*, vol. 10, no. 3, pp. 445–452, 1991.
- [9] B. R. Whiting, "Signal statistics of x-ray computed tomography," *Proceedings of the SPIE Physics of Medical Imaging*, vol. 4682, pp. 53–60, 2002.
- [10] B. R. Whiting, P. Massoumzadeh, and O. A. Earl, "Properties of preprocessed sinogram data in x-ray computed tomography," *Medical Physics*, vol. 33, no. 9, pp. 3290–3303, 2006.
- [11] P. Gravel, G. Beaudoin, and J. A. De Guise, "A method for modeling noise in medical images," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1221–1232, 2004.
- [12] I. A. Elbakri and J. A. Fessler, "Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography," *Proceedings of SPIE Image Processing*, vol. 5032, pp. 1839–1850, 2003.
- [13] J. Xu and B. M. W. Tsui, "Electronic noise modeling in statistical iterative reconstruction," *IEEE Transactions on Image Processing*, vol. 18, no. 6, pp. 1228–1238, 2009.
- [14] J. R. Mayo, K. P. Whittall, A. N. Leung et al., "Simulated dose reduction in conventional chest CT: validation study," *Radiology*, vol. 202, no. 2, pp. 453–457, 1997.
- [15] M. Gies, W. A. Kalender, H. Wolf, and C. Suess, "Dose reduction in CT by anatomically adapted tube current modulation. I. Simulation studies," *Medical Physics*, vol. 26, no. 11, pp. 2235–2247, 1999.
- [16] S. I. Olsen, "Estimation of noise in images: an evaluation," *Graphical Models and Image Processing*, vol. 55, no. 4, pp. 319–323, 1993.
- [17] J. Immerkaer, "Fast noise variance estimation," *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 300–302, 1996.
- [18] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [19] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [20] F. Russo, "A method for estimation and filtering of Gaussian noise in images," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, no. 4, pp. 1148–1154, 2003.
- [21] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, 2008.
- [22] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang, "Noise estimation from a single image," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 1, pp. 901–908, 2006.

- [23] A. De Stefano, P. White, and W. Collis, "Training methods for image noise level estimation on wavelet components," *EURASIP Journal on Applied Signal Processing*, no. 16, pp. 2400–2407, 2004.
- [24] S. C. Tai and S. M. Yang, "A fast method For image noise estimation using laplacian operator and adaptive edge detection," in *Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP '08)*, vol. 12–14, pp. 1077–1081, St Julians, Malta, March 2008.
- [25] B. Antoni, C. Bartomeu, and M. J. Michel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 123–139, 2008, Special Section: Selection of Papers for CVPR 2005, Guest Editors: Cordelia Schmid, Stefano Soatto and Carlo Tomasi.
- [26] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.
- [27] Z. Liao, S. Hu, and W. Chen, "Determining neighborhoods of image pixels automatically for adaptive image denoising using nonlinear time series analysis," *Mathematical Problems in Engineering*, vol. 2010, Article ID 914564, 14 pages, 2010.
- [28] K. Vladimir, F. Alessandro, E. Karen, and A. Jaakko, "From local kernel to nonlocal multiple-model image denoising," *International Journal of Computer Vision*, vol. 86, no. 1, pp. 1–32, 2010.
- [29] Y. Li, Y. Chen, W. Chen, L. Luo, and X. Yin, "Improving low-dose X-ray CT images by weighted intensity averaging over large-scale neighborhoods," in *Proceedings of the International Conference on Medical Image Analysis and Clinical Application (MIACA '10)*, pp. 1–4, 2010.
- [30] Z. S. Kelm, D. Blezek, B. Bartholmai, and B. J. Erickson, "Optimizing non-local means for denoising low dose CT," in *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '09)*, pp. 662–665, 2009.
- [31] B. I. Yi-ming, M. A. Jian-hua, L. I. U. Nan et al., "Anscombe transform and BM3D filtering based projection restoration for low-dose CT reconstruction," *Computer Engineering and Applications*, vol. 46, no. 13, pp. 216–220, 2010.
- [32] J. Ma, J. Huang, Y. Chen et al., "Generalized Gibbs Prior based high quality low-dose X-CT reconstruction," *Computer Engineering and Applications*, vol. 44, no. 16, pp. 4–6, 2008.
- [33] S. Y. Chen, Hanyang Tong, and Carlo Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
- [34] 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.
- [35] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 480–487, 2011.
- [36] S. Y. Chen, J. Zhang, H. Zhang et al., "Myocardial motion analysis for determination of tei-index of human heart," *Sensors*, vol. 10, no. 12, pp. 11428–11439, 2010.
- [37] E. G. Bakhoun and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.
- [38] E. G. Bakhoun 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.
- [39] J. Vandemeulebroucke, S. Rit, J. Kybic, P. Clarysse, and D. Sarrut, "Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs," *Medical Physics*, vol. 38, no. 1, pp. 166–178, 2011.
- [40] J. L. Ong and A. K. Seghouane, "From point to local neighborhood: polyp detection in CT colonography using geodesic ring neighborhoods," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1000–1010, 2011.
- [41] J. C. Ross, S. J. E. Rail, K. G. Diaz et al., "Automatic lung lobe segmentation using particles, thin plate splines, and maximum a posteriori estimation," *Medical Image Computing and Computer-Assisted Intervention*, vol. 6363, part 3, pp. 163–171, 2010.
- [42] T. Li, X. Li, J. Wang et al., "Nonlinear sinogram smoothing for low-dose X-ray CT," *IEEE Transactions on Nuclear Science*, vol. 51, no. 5, pp. 2505–2513, 2004.

Research Article

Emergent Behaviors in Social Networks of Adaptive Agents

Florin Leon

Faculty of Automatic Control and Computer Engineering, "Gheorghe Asachi" Technical University of Iasi, Boulevard Mangeron 27, 700050 Iasi, Romania

Correspondence should be addressed to Florin Leon, florinleon@gmail.com

Received 24 June 2011; Accepted 13 July 2011

Academic Editor: Carlo Cattani

Copyright © 2012 Florin Leon. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Designing multiagent systems that can exhibit coherent group behavior based on a small number of simple rules is a very challenging problem. The evolution of mobile computing environments has created a need for adaptive, robust systems, whose components should be able to cooperate in order to solve the tasks continuously received from users or other software agents. In this paper, an interaction protocol for a task allocation system is proposed, which can reveal the formation of social networks as an emergent property. The agents can improve their solving ability by learning and can collaborate with their peers to deal with more difficult tasks. The experiments show that the evolution of the social networks is similar under a great variety of settings and depends only on the dynamism of the environment. The average number of connections and resources of the agents follows a power law distribution. Different configurations are studied in order to find the optimal set of parameters that leads to the maximum overall efficiency of the multiagent system.

1. Introduction

Complexity theory is a relatively recent research field which mainly studies the way in which critically interacting components self-organize to form potentially evolving structures exhibiting a hierarchy of emergent system properties [1]. Thus, complexity theory includes the new domain of self-organizing systems, which searches for general rules about the forms and evolution of system structures, and the methods that could predict the potential organization resulting from changes to the underlying components [2]. It is hypothesized that a dynamic system, independent of its type or composition, always tends to evolve towards a state of equilibrium, that is, an attractor. This process decreases the uncertainty about the system state and therefore the system's entropy, leading to self-organization [3].

Four necessary requirements have been identified for systems to exhibit a self-organizing behavior under external pressure [4].

- (i) *Mutual causality*: at least two components of the system have a circular relationship, each influencing the other.
- (ii) *Autocatalysis*: at least one of the components is causally influenced by another component, resulting in its own increase.
- (iii) *Far from equilibrium condition*: the system imports a large amount of energy from outside the system, uses it to renew its own structures (autopoiesis), and dissipates rather than accumulates the increasing entropy back into the environment.
- (iv) *Morphogenetic changes*: at least one of the components of the system must be open to external random variations from outside the system, and the components of the system are changed [5].

Currently, many applications of self-organizing systems have been proposed [6], among which we could mention the self-organization of broker agent communities for information exchange [7], solutions to constraint satisfaction problems such as the automated timetabling problem [8], flood forecasting [9], land use allocation based on principles of eco-problem solving [10], and traffic simulation using holons [11]. An interesting approach is the use of the emergent structures given by Conway's game of life [12] to define functional components such as adders, sliding block-based memories, and even the pattern of a Turing machine [13].

1.1. Emergent Properties

The emergent properties are characteristic of complex systems, which usually are formed of a relatively small variety of fairly homogenous agents acting together according to simple local rules of interaction and communication, with no global or centralized control. Despite their homogeneity, complex systems are irreducible, because emergent global properties result from agents operating in a self-organizing manner [14]. Emergent properties are often associated with properties such as fault tolerance, robustness, and adaptability.

Surprise or novelty is characteristic of emergence. It can be stated that the language of low-level design L and the language of high-level observation H are distinct, and the causal link between the elementary interactions programmed in L and the behaviors observed in H is nonobvious, and therefore surprising, to the observer [15]. Thus, emergence can be considered a process that leads to the appearance of structure not directly described by the defining constraints and forces that control the system, and over time something new appears at scales not directly specified by the local rules of interaction [16].

Some researchers studied the possibility of emergence engineering [17], that is, designing the local rules such that a desired high-level behavior should be produced. They concluded that different methodologies than the usual software development should be employed in order to get close to this desideratum, because the main idea is to implement or generate the system without actually knowing how it works. They suggest three approaches in this respect:

- (i) by imitating phenomena usually considered as emergent;
- (ii) by using an incremental design process and incrementally adapting the system to make it produce a behavior that will be closer to the behavior one ultimately expects;
- (iii) by developing self-adaptive systems.

1.2. Complex Networks

Many complex phenomena are related to a network-based organization. Complex networks are often encountered in the real world, for example, technological networks such as the power grid, biological networks such as the protein interaction networks or the neural network of the roundworm *Caenorhabditis elegans*, and social networks such as scientific collaboration of human communication networks [18]. There are also many other types of physical or chemical systems that demonstrate nonlinear behavior [19–21].

Many real complex networks share distinctive features that make them different from regular (lattice) and random networks. One such feature is the “small world” property, which means that the average shortest path length between the vertices in the network is small; it usually scales logarithmically with the size of the network, and the network exhibits a high clustering effect. This model proposed by Watts and Strogatz [22] was designed as the simplest possible model that accounts for clustering while retaining the short average path lengths of the Erdős and Rényi model [23].

A well-known example is the so-called “six degrees of separation” in social networks [24]. Another is the scale-free property of many such networks, that is, the probability distribution of the number of links per node $P(k)$ satisfies a power law $P(k) \sim k^{-\gamma}$ with the degree exponent γ in the range $2 < \gamma < 3$ [25]. The World Wide Web network has been shown to be a scale-free graph [26].

The generation of scale-free graphs can be performed based on two basic rules [27].

- (i) *Growth*: at each time step, a new node is added to the graph.
- (ii) *Preferential attachment*: when a new node is added to the graph, it attaches preferentially to nodes with high degrees.

Complex networks can also reveal self-similar properties, and thus their fractal dimension could be computed using methods such as “box counting” and “cluster growing” [28].

1.3. Agent-Based Simulation of Complex Phenomena

Software agents can be naturally used to play the role of autonomous entities capable of self-organization. Agents are appropriate to simulate such systems, in order to better understand models or create new ones [29]. However, agents can be increasingly useful for the development of distributed systems whose components can self-organize and work in a decentralized manner for the realization of the global functionality [30].

The ability of a multiagent system to dynamically reorganize its structure and operation at run-time is highly valuable for many application domains. Therefore, allocating tasks to agents in multi-agent systems is a significant research issue. There are two main directions of study in this respect. On the one hand, the centralized methods for task allocation assume that there is a central controller to assign tasks to agents [31]. The centralized approach can make the allocation process efficient and effective in a small network since the central planner has a global view of the system and it knows which agents are good at which tasks. On the other hand, the decentralized style is more scalable and robust but the communication overhead increases. Some proposed mechanisms form groups of agents before allocating tasks which may result in the increase of computation and communication cost [32]. Other distributed protocols only allow agents to request help

for tasks from their directly linked neighbors which may increase the possibility of task allocation failure because of limited resources available from the agents' neighbors [33]. Another technique is to address the task allocation problem in a loosely coupled peer-to-peer system [34]. Other researchers aim at minimizing the team cost subject to the constraint that each task must be executed simultaneously by a given number of cooperative agents which form coalitions [35]. Task allocation can also be observed in the biological world, for example, ants of different sizes take on tasks of different difficulties: larger ants carry larger food, while smaller ones perform more manageable tasks [36].

2. Model Description

The proposed multiagent system is composed of a set of agents A , which are "physically" distributed over a square grid. However, this localization does not prevent agents from forming relations with any other agents, based on the common interest of solving tasks, as it will be shown in Section 3.

The tasks are considered to be defined by a series of p attributes. A task has specific values of these attributes, considered to be *complexity levels*, each within a certain domain. Let T be the set of tasks $T = \{t_i\}$ and F the set of attributes or features $F = \{c_j\}$. Then:

$$t_i = \{c_1, \dots, c_p\}, \quad (2.1)$$

with $c_j \in D_j$, for all $j \in \{1, \dots, p\}$ and $p = |F|$.

Each agent has *competence levels* l_a^j associated with each attribute j , which describe how much knowledge an agent has regarding a particular feature of a task.

For example, in a software industry environment, attributes can relate to the use of databases, developing graphical user interfaces, or working with specific algorithms. A programming task can contain such features in varying degrees, with different complexity levels. An agent specialized on databases has a high competence level for the first attribute, although these tasks may also contain issues related to the other attributes, possibly with lower complexity levels. However, if other types of tasks should be addressed, the agent can gradually specialize in other development areas.

The tasks are generated in the following way. First, the number of nonnull attributes $p_{nn}^i \in \mathbb{N}$, $1 \leq p_{nn}^i \leq p$, is determined, by using a power law or a uniform distribution:

$$P(p_{nn}^i) \approx \frac{1}{(p_{nn}^i)^k} \quad (2.2)$$

or

$$P(p_{nn}^i) \approx U(1, p). \quad (2.3)$$

When using the power law distribution, most of the tasks will have one or two non-null attributes, and therefore they will have a large degree of specialization. Thus, agents can specialize in performing some type of tasks. When using a uniform distribution, there will be a larger number of "interdisciplinary" tasks; therefore the agents will have to interact more in order to solve them in a cooperative way.

Based on the distribution of non-null attributes, tasks are continuously generated during the simulation, and the corresponding values of the non-null attributes are generated from a uniform distribution:

$$c_j \sim U(1, L_{\max}), \quad (2.4)$$

where L_{\max} is the maximum complexity level allowed in a certain simulation epoch. In Section 4, we will analyze the situations where L_{\max} is constant throughout the whole simulation, or gradually increases.

When an agent receives a task, it first verifies whether the complexity levels of the attributes are less or equal to its respective competence levels: $c_j \leq l_a^j$, for all $j = 1, \dots, p$.

In this case, the agent can solve the task by itself and receives a monetary payment of:

$$M_a^i = \sum_{j=1}^p c_j^2. \quad (2.5)$$

If the complexity level of a task attribute is greater by 1 than the corresponding competence level, then the agent can learn in order to reach the proper level. There are two prerequisites. First, each agent has an individual "willingness to learn" $W_a^l \sim U(W_{\min}, W_{\max})$. The probability to pass from a competence level to another is given by the following probability:

$$P_a^l = \left(W_a^l\right)^{l_a^j}. \quad (2.6)$$

Accordingly, the probability to go from a low level to the next is much higher than the probability to go from a high level to the next. For example, if $W_a^l = 0.95$, the probability to go from level 1 to level 2 is 0.95, while the probability to go from level 9 to level 10 is 0.63.

Secondly, a payment is required of the agent for the learning step:

$$M_a^l = c_j^2. \quad (2.7)$$

For this process to work, each agent has an equal initial amount of "money" M^0 .

If the agent is fit to solve the whole task following a possible learning phase, the agent solves it and it receives the same payment as that described in (2.5).

If the agent is unable to solve a task by itself, it seeks other agents to solve the parts of the task it cannot handle by itself.

3. Interaction Protocol and Social Network Formation

The environment randomly distributes a number of tasks fewer than the number of agents in the system: $|T| < |A|$. Since some agents will be able to solve their tasks, either individually or cooperatively, while others will not, in the subsequent epochs, tasks will be given

preferentially to agents that previously succeeded more. Thus, each agent has a probability to solve tasks defined as

$$P_a^s = \frac{T_a^s}{T_a^r}, \quad (3.1)$$

where T_a^s is the number of tasks solved by agent a and T_a^r is the total number of tasks received by agent a .

In this setting, the agents are sorted by their probabilities to solve tasks, P_a^s , and only the first $|T|$ receives new tasks. Ties between agents with the same P_a^s are broken randomly. The initial values are $T_a^s = 1$, $T_a^r = 2$, for all $a \in A$, and therefore the initial probability to solve is $P_a^s = 0.5$, for all $a \in A$.

However, the agents who fail to solve tasks at first become disadvantaged because they no longer receive new tasks and therefore they no longer have the direct incentives to improve their competence levels. For this reason we introduced a perturbation rate R , such that, when $R = 0$, the tasks are distributed deterministically, as presented above, and, when $R = 1$, tasks are randomly distributed.

Each agent has a set of connections to other agents, that is, “friends,” initially void. When an agent cannot solve a task by itself, it begins a breadth-first search on the connection graph. As agents are “physically” situated on a lattice, the immediate neighbors are always added to the search queue after the immediate friends. Search is performed without allowing agents to be visited twice in order to avoid infinite loops.

Every time an agent is not able to solve the originator’s task either directly or by learning, its own friends are added to the search queue. In this way, the entire graph is ultimately explored, and the closest agents are always selected to solve the tasks, since breadth-first search is complete and optimal. We ignore the exponential memory requirements of the algorithm because we consider that this aspect is irrelevant to our study; the graph itself has quite a small depth, similar to the networks that exhibit the “small world” phenomenon.

Once an agent is found which can solve one or more attributes of a task, and only when the whole originator’s task can be solved, the agents on the path from the originator to the solver are sequentially connected. Finally, the originator and the solver are directly connected.

Several aspects must be underlined: there is no limitation to the number of tasks an agent can solve in one round. It may have only one individually assigned task, but several task attributes received as subcontracting. The originator agent u receives a commission rate α for the subcontracted task attributes, and the solver agent v receives the rest:

$$\begin{aligned} M'_u &= M_u + \alpha \cdot \sum_{j \in \Delta} c_j^2, \\ M'_v &= M_v + (1 - \alpha) \cdot \sum_{j \in \Delta} c_j^2, \end{aligned} \quad (3.2)$$

where Δ is the set of task attributes transferred to agent v for solving.

There is no payment for the intermediary agents, and this is one of the main differences between this approach and the classical contract net protocol [37]. In this paper, we consider

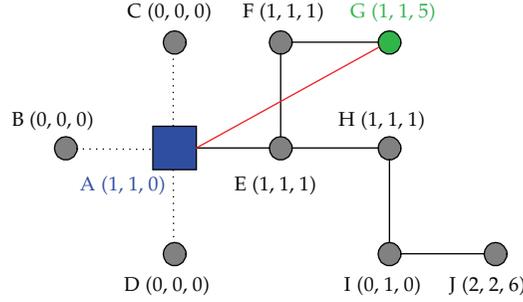


Figure 1: Illustration of the interaction protocol and formation of new connections.

that the most important feature of the social network is to *transfer information*, not actual solutions for tasks accompanied by monetary payments.

Once a task is solved, it is considered to be solved by the originator agent u , whose P_u^s increases. Every connection, once established, is active a limited number of epochs θ_T after which it disappears. If an agent does not succeed in solving a task during one epoch and does not receive another one in the next epoch, it can keep its task for a maximum number of epochs θ_C . If an agent cannot eventually solve a task, its P_u^s consequently decreases.

Figure 1 presents a small example that can illustrate the interaction protocol. Let us assume that agent A receives a task with complexity levels $\{2, 1, 5\}$. Its own competence levels are $\{1, 1, 0\}$, therefore

- (i) it can learn to solve attribute 1, If the learning process is successful, its competence levels will become $\{2, 1, 0\}$,
- (ii) it can directly solve attribute 2,
- (iii) it must search for another agent to solve attribute 3.

Finding a capable agent for attribute 3 involves a breadth-first search (BFS) process. The agents connected with agent A are automatically added to the search queue, in this case agent E . The other neighbors are also added to the queue, that is, B , C , and D . The neighbor links are not considered social network connections, but, in case an agent is isolated at a certain point, its physical neighbors may help it to find other agents. Therefore, its start queue is $\{E, B, C, D\}$. As search nodes are expanded, the “friends” of the agents are added to the end of the queue. In the second stage, the queue becomes $\{B, C, D, F, H\}$. The nodes corresponding to the agents already visited are not expanded again. As B , C , and D have no new connection, the node corresponding to the F agent is expanded and the queue becomes $\{H, G\}$. So far, no agent expanded is capable of solving attribute 3. Then, H is expanded, and the queue is $\{G, I, J\}$. Next, G is expanded and it is capable of solving attribute 3. Therefore, A and G become directly connected, and A pays G a fraction $1 - \alpha$ of the price for solving attribute 3.

Although agent J is also capable of solving attribute 3, the optimality of the BFS algorithm ensures that the nearest capable agent (in the social network, not necessarily in the physical environment) is always selected.

4. Parameter Influence on System Behavior: Case Studies

In this section, the results of some simulations are presented, which illustrate the behavior of the multiagent system. We focus on the common patterns that emerge from the social

network created by the system in execution, while varying some of the parameters of the simulation.

For all the case studies, the following parameters remain unchanged. The initial amount of “money” available to the agents $M^0 = 100$. The limits of the “willingness to learn” are $W_{\min} = 0.9$ and $W_{\max} = 1$. The number of epochs while a connection is active after being created is $\theta_T = 5$. The maximum number of epochs while an agent can keep a task without being able to solve it is $\theta_C = 3$. The total number of tasks distributed during every epoch is $|T| = |A|/2$.

Figure 2 shows the evolution of the social network after 10, 100, 350, and 500 epochs, from left to right, top to bottom, respectively. The maximum complexity level $L_{\max} = 10$ remains constant throughout the simulation. As it can be seen, the number of agents is 400, arranged on a grid of size 20×20 . The perturbation rate $R = 0$, the commission $\alpha = 0.1$, and the non-null attributes follow a power law distribution (2.2) with exponent $k = 2$. One can notice that at the beginning local connections form on short physical distances on the grid. Gradually, these small clusters connect and global networks emerge. Depending on the initial conditions (the actual values of task complexity levels, agent competence levels, and the first distribution of tasks), which are randomly generated and can vary from a simulation to another, one global network or two/few networks can appear. Their number also greatly depends on the number of agents (reflected on the size of the grid). After a great increase, the number of connections gradually decreases, because the agents tend to learn and adapt to the unchanging environment, depicted by a constant L_{\max} . Thus, as agents increase their individual competence levels, they no longer need other agents to solve their tasks. Finally, when most of the agents become independent, the number of active connections approaches 0.

Figure 3 shows the status of the multiagent system after 100 epochs, with the same parameters, when the number of agents varies (100, 256 and 2500), in order to demonstrate the scaling behavior of the system. One can see that the formation of global social network(s) is similar. Also, the evolution of the system as the number of epochs increases is the same: the agents enhance their competence and no longer need their peers, and therefore the connections go through a gradual dissolution process.

This tendency of the system is altered when the maximum complexity level of the tasks is no longer constant but grows as the simulation proceeds. Figure 4 shows the status of the multiagent system after 500 epochs, when the environment is dynamic and L_{\max} constantly increases: $L_{\max} = 10 + N_{\text{epochs}}/20$. Because the competence levels of the agents no longer become saturated, the shape of the social network is maintained throughout the simulation and its dissolution no longer takes place.

This situation is more similar to the real-world problem of task allocation, where the environment continuously changes and individuals must change as well in order to stay adapted and solve the increasingly competitive tasks they are faced with. It is this individual change and adaptation itself that causes the dynamism of the environment, due to the interconnectedness of the individuals.

Figure 5 shows the status of the social network after 100 epochs, when the perturbation rate $R = 0.5$. When the perturbation rate was 0, the agents who solved the tasks were preferred when assigning new tasks. In this case, the evolution of the system greatly depended on the initial assignment of tasks. While unsuccessful agents were eliminated from receiving new tasks, it is likely that some competent agents remained in the system without being assigned in the first phases of the simulation, and this also decreased their chances to being assigned later on. With a 50% randomization, as all the agents have a chance to receive tasks, the social

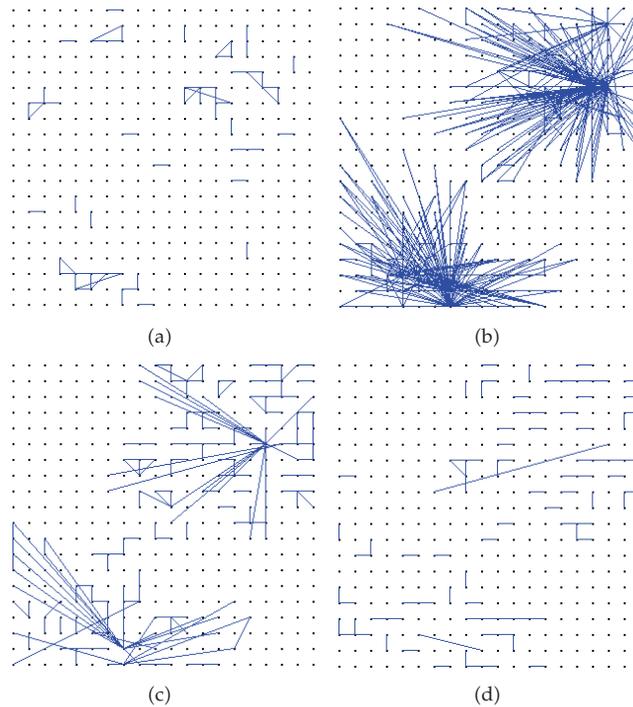


Figure 2: The evolution of the social network after 10, 100, 350, and 500 epochs when the environment is static.

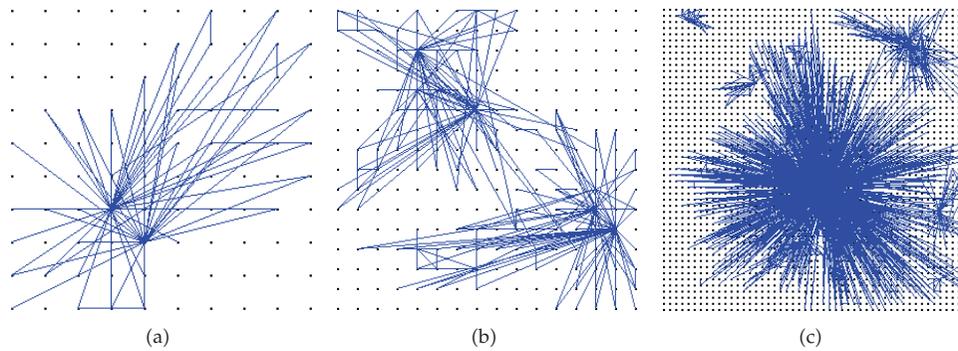


Figure 3: The status of the social network after 100 epochs for different numbers of agents: 100, 256, and 2500.

network is greatly extended. Because the system is evolving, with the same linear increase of L_{\max} in the number of epochs passed, the shape of the system remains similar later in the simulation.

Figure 6 shows the social network of the agents after 100, 350, and 500 epochs, respectively, when the environment is not dynamic ($L_{\max} = 10$ and remains constant) and the perturbation rate is $R = 0.5$. The resulting behavior is an intermediate one between those displayed in Figures 2 and 5. Although more connections are formed, the agents still tend to saturate their competence levels. However, the random perturbation involves more agents in

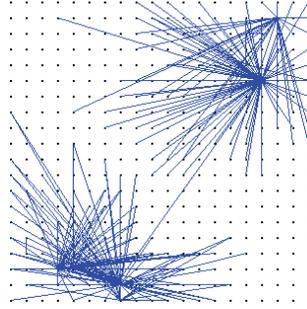


Figure 4: The status of the social network after 500 epochs when the environment is dynamic.

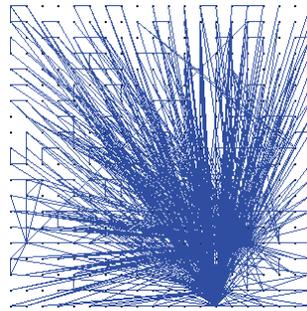


Figure 5: The status of the social network after 100 epochs when the perturbation rate $R = 0.5$.

the task allocation process, and, therefore, connections still remain active after 500 epochs, but in a different, mostly neighbor-based configuration, because it becomes increasingly likely that a competent agent will be found in the physical neighborhood of an agent.

Figure 7 shows the social network after 100, 350, and 500 epochs, respectively, when the environment is dynamic ($L_{\max} = 10 + N_{\text{epochs}}/20$), and the perturbation rate is $R = 1$. In this case, the social network covers almost all the agents, and this configuration remains relatively stable until the end of the simulation.

The overall behavior of the multiagent system is similar when the task attributes are generated following a uniform distribution (2.3). When the environment is dynamic (Figure 8), the network configuration is preserved. When it is static (Figure 9), the connections eventually dissolve. However, compared to the power law distribution case, here the evolution of the system is faster, because the tasks involve more cooperation on the part of the individual agents, and therefore the agents specialize more quickly on the different types of attributes. After only 350 epochs, almost all the connections disappear.

In the following paragraphs, we will analyze some of the statistical properties of the multiagent system, taking into account the way in which the average number of connections, the average wealth, and the average competence levels of the agents vary during 500 epochs of the simulation.

In Figure 10, the evolution of the average number of connections is displayed, when the type of environment and the type of task attribute generation differ: evolving (dynamic) or nonevolving (static) for the former and power law or uniform distribution for the latter. The shape of the function is similar in all four cases, with a sharp increase toward the beginning of the simulation. As it was shown above, the uniform distribution delays the

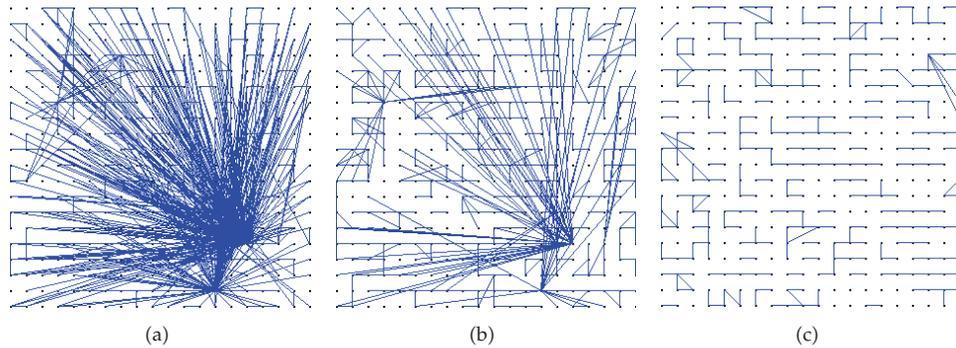


Figure 6: The evolution of the social network after 100, 350, and 500 epochs when the environment is static and the perturbation rate $R = 0.5$.

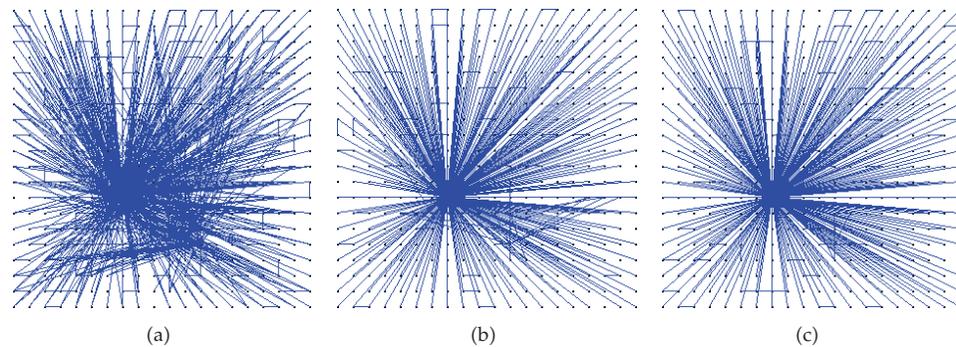


Figure 7: The evolution of the social network after 100, 350, and 500 epochs when the environment is dynamic and the perturbation rate $R = 1$.

formation of the maximum sized network however, when the network is formed, it has a greater number of connections. As the system converges, the average number of connections in the evolving setting is greater than that for the power law distribution. Also, for the non-evolving case, the average number of connections decreases more quickly than in the power law distribution scenario.

These average values show that only a few agents have a great number of connections, while most of the agents have one or no active connections at all. In fact, the histogram of the number of agents having a certain number of connections reveals a power law distribution, as shown in Figure 11. The ordinate axis (the number of agents) has a logarithmic scale. The abscise axis shows the five equal size intervals ranging from 0 to the maximum number of connections.

Regarding the evolution of the average competence levels, Figure 12 displays a typical situation. It is important to note that the shape of this function is approximately the same in all scenarios.

The evolution of the maximum competence level (Figure 13) has a similar profile at the beginning of the simulation, but it grows fairly linear, in small increments, unlike the average value, which resembles a logarithmic expansion. Therefore, it is clear that the difference between the top agents and the rest will become increasingly larger over time, and this will amplify the disparity concerning the resource allocation in the system.

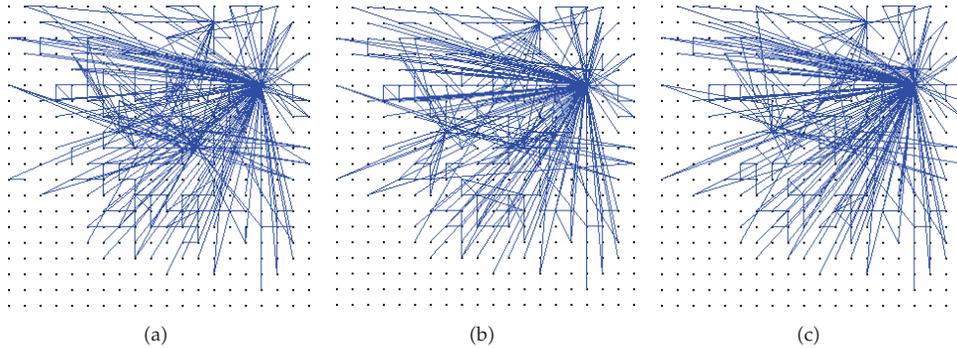


Figure 8: The evolution of the social network after 100, 350, and 500 epochs when the task attributes are generated from a uniform distribution and the environment is dynamic.

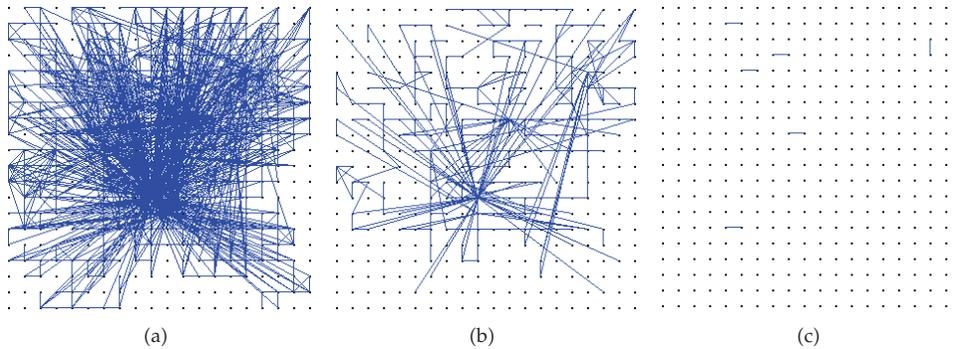


Figure 9: The evolution of the social network after 35, 100, and 350 epochs when the task attributes are generated from a uniform distribution and the environment is static.

An interesting fact was observed, that the “willingness to learn” W_a^l is not crucial to become a top agent. There are “best agents” with the greatest competence levels and the most money, which do not have $W_a^l = 1$, but even a value close to W_{\min} . It seems that the most important fact is the number of opportunities to learn (by receiving tasks directly or indirectly, from other agents in its social network), which forces the agent to eventually learn and become the most competent.

Regarding the average wealth of the agents, which results from the monetary payments for solving tasks, a similar distribution is encountered: there is one (or very few agents) on top, with a large amount of money, and most agents are on the bottom of the scale. Since all the agents start with an equal wealth, at some point in the simulation the ratio reverses, a phenomenon displayed in Figures 14 and 15. The figures show the histogram with five intervals, from 0 to the maximum amount of money, and the bars show on the ordinate the number of corresponding agents on a logarithmic scale. At the beginning (time frame 1, but not epoch 1), most agents are in interval $I5$, close to the maximum. During only a few epochs, the majority of the agents lose their relative wealth compared to the top agents and move into interval $I1$, close to the minimum. The phenomenon takes place more quickly when the commission $\alpha = 0.1$ (Figure 14) than when the commission $\alpha = 0.95$ (Figure 15). Therefore, a higher commission leads to a more egalitarian distribution of the wealth in the first phase of the simulation although, in the end, the results are similar.

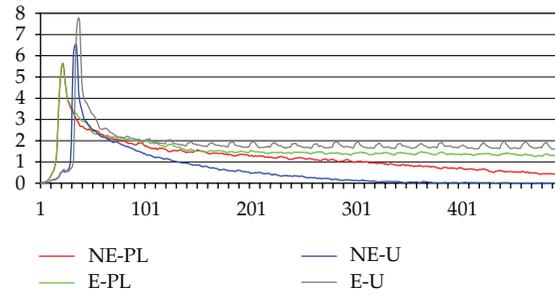


Figure 10: The evolution of the average number of connections.

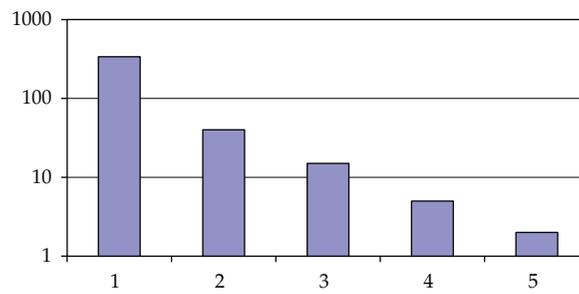


Figure 11: Power law distribution of the number of connections.

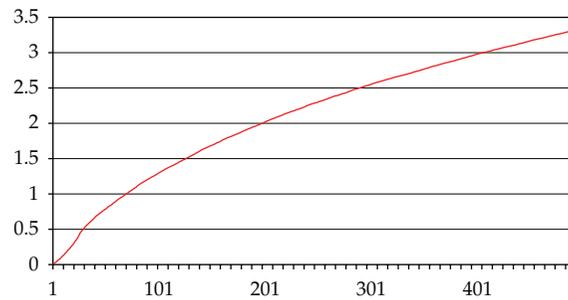


Figure 12: The evolution of average competence levels.

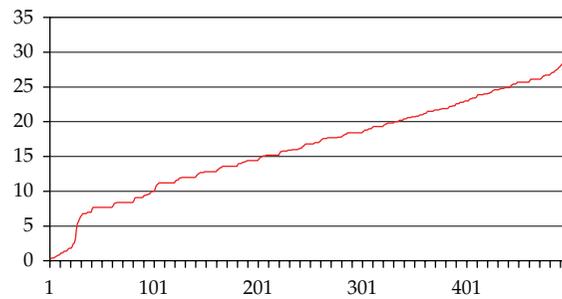


Figure 13: The evolution of maximum competence levels.

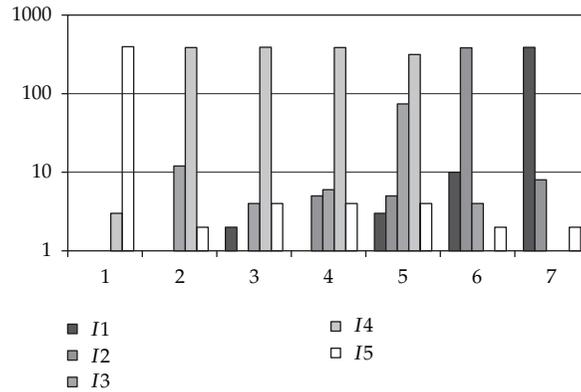


Figure 14: The change histogram of the average wealth with a commission rate $\alpha = 0.1$.

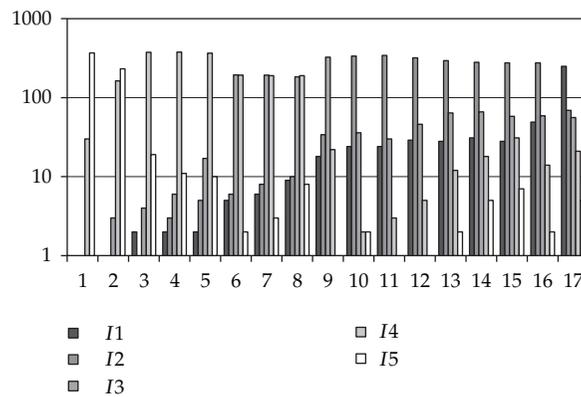


Figure 15: The change histogram of the average wealth with a commission rate $\alpha = 0.95$.

Also, it was observed that when the number of tasks decreases (e.g., from 50% to 20% of the number of agents), the resulting wealth distribution is more balanced.

It is important to note that these figures show only relative values. The absolute value of the money actually increases for all the agents involved.

Next, we address the reciprocal relations between the statistical quantities under study.

In Figure 16, the average wealth of the agents is displayed as a function of the average competence level. It can be seen that the amount of money increases exponentially with knowledge, because an agent with higher competence can receive and solve tasks from all the agents in its social network.

Figure 17 presents the average number of connections as a function of the average competence level. At the beginning, the number of connections must increase if the agents have low competence levels because they must find other agents to help them solve their tasks. As the average knowledge increases, the number of connection decreases and stabilizes because the agents become more independent and can solve the tasks on their own.

The shapes of the functions in Figures 16 and 17 are the same irrespective of the parameters used in the simulation.

Finally, we analyze the efficiency of the overall task allocation system, defined as the average probability to solve tasks P_a^s of all the agents in the system.

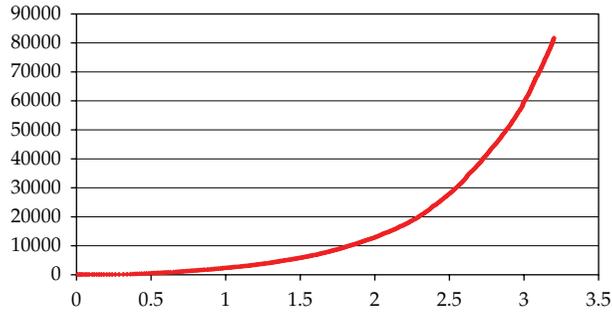


Figure 16: The average amount of money as a function of the average competence level.

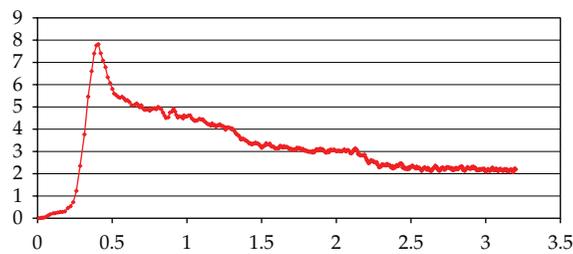


Figure 17: The average number of connections as a function of the average competence level.

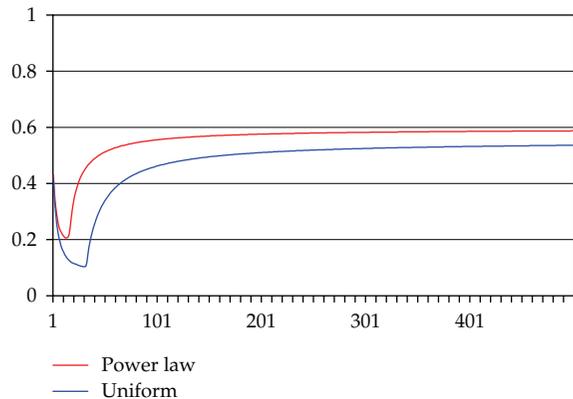


Figure 18: The evolution of the overall system efficiency with different distributions for task attribute generation.

In Figure 18, one can notice that the efficiency is slightly larger when the task attributes are generated using a power law, compared to the uniform distribution. This is because tasks generated in this way are “easier,” as they require less collaboration or learning for solving them.

Figure 19 displays the evolution of the overall efficiency for different values of the perturbation rate R . Although giving tasks preferentially to previously successful agents could seem like a useful heuristic, in fact it appears that the inclusion of more agents in the task solving process leads to better global results. It is especially important to keep the perturbation rate positive $R > 0$ because the difference between the cases when $R = 0.5$ and

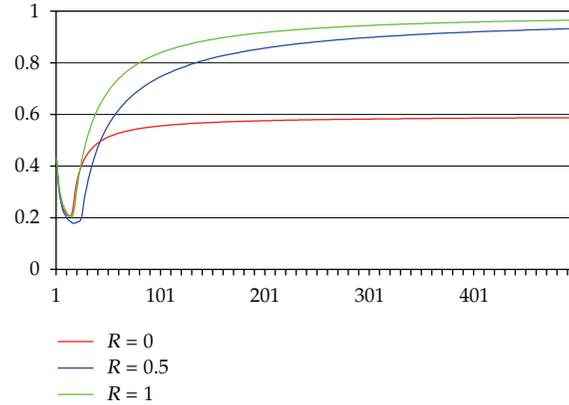


Figure 19: The evolution of the overall system efficiency for different perturbation rates R .

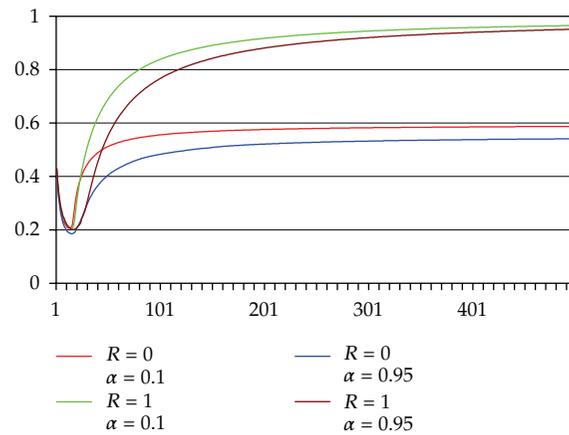


Figure 20: The evolution of the overall system efficiency for different perturbation rates R and commission rates α .

$R = 1$ is smaller than the difference between the cases when $R = 0$ and $R = 0.5$. A small change in the preferential allocation mechanism can lead to a greater change in the overall efficiency. Therefore, this is another nonlinear effect of the model.

When the commission is considered, it can be seen in Figure 20 that the efficiency is a little larger when the commission is smaller. This great change in the way in which money is distributed between the originator and the solver agent could be expected to have a great effect in the overall wealth distribution and system efficiency (i.e., the most money could belong to mediocre agents). However, it was found that, even when the commission is high, the competent agents eventually have the greatest wealth. Therefore, it seems that the proposed protocol is very robust to speculations.

5. Conclusions

According to the proposed model, the agents can gradually enhance their knowledge by learning when presented with tasks to be solved. They also manifest a collaborative

behavior for solving more complex problems. The social network acts as a means to transfer information, unlike the classic contract net protocol which transfers actual tasks and “money” for subcontracting. The model also includes monetary payments for agents which successfully accomplish their goals. Under different configurations of the environment, it is shown that the agents form evolving social networks and the system exhibits several types of emergent properties:

- (i) a power law distribution of the number of connections and agent “wealth”;
- (ii) a dynamic or evolving environment is required to maintain the social networks;
- (iii) a non-deterministic task distribution increases the overall system efficiency;
- (iv) small changes in task distribution cause great changes in the resulting system efficiency;
- (v) robustness to speculations (for different commission rates).

A future development of the research would be to apply the generic model proposed in this paper to a real-world task allocation problem.

Acknowledgment

This work was supported by CNCSIS-UEFISCSU, project no. PNII-IDEI 316/2008, *Behavioural Patterns Library for Intelligent Agents Used in Engineering and Management*.

References

- [1] C. Lucas, “Self-organizing systems: frequently asked questions,” Version 3, September 2008, <http://www.calresco.org/sos/sosfaq.htm>.
- [2] J. M. E. Gabbai, H. Yin, W. A. Wright, and N. M. Allinson, “Self-organization, emergence and multi-agent systems,” in *Proceedings of the IEEE International Conference on Neural Networks and Brain Proceedings (ICNN&B '05)*, pp. 1858–1863, Beijing, China, October 2005.
- [3] F. Heylighen, “Knowledge Management, Organizational Intelligence and Learning, and Complexity,” in *The Encyclopedia of Life Support Systems*, L. D. Kiel, Ed., Eolss Publishers, Oxford, 2001.
- [4] P. Glansdorff and I. Prigogine, *Thermodynamic Study of Structure, Stability and Fluctuations*, Wiley, 1971.
- [5] N. S. Contractor and D. R. Seibold, “Theoretical frameworks for the study of structuring processes in group decision support system—adaptive structuration theory and self-organising systems theory,” *Human Communication Research*, vol. 19, no. 4, pp. 528–563, 1993.
- [6] C. Bernon, V. Chevrier, V. Hilaire, and P. Marrow, “Applications of self-organising multi-agent systems: an initial framework for comparison,” *Informatica*, vol. 30, no. 1, pp. 73–82, 2006.
- [7] F. Wang, “Self-organising communities formed by middle agents,” in *Proceedings of the 1st International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS '02)*, pp. 1333–1339, Bologna, Italy, July 2002.
- [8] G. Picard, C. Bernon, and M. P. Gleizes, “ETTO: emergent timetabling by cooperative self-organisation,” in *Proceedings of the 3rd International Workshop on Engineering Self-Organising Applications (ESOA '05)*, pp. 31–45, Utrecht, The Netherlands, 2005.
- [9] J. P. Georgé, M. P. Gleizes, P. Glize, and C. Régis, “Real-time simulation for flood forecast: an adaptive multi-agent system STAFF,” in *Proceedings of the Symposium on Adaptive Agents and Multi-Agent Systems (AISB '03)*, pp. 7–11, University of Wales, Aberystwyth, UK, 2003.
- [10] A. Dury, F. Le Ber, and V. Chevrier, “A reactive approach for solving constraint satisfaction problems: assigning land use to farming territories,” in *Intelligent Agents V Agents Theories, Architectures and Languages*, J. P. Muller, M. P. Singh, and A. S. Rao, Eds., vol. 1555 of *Lecture Notes in Artificial Intelligence*, pp. 397–412, Springer, 1998.

- [11] S. Rodriguez, V. Hilaire, and A. Koukam, "Holonc modeling of environments for situated multi-agent systems," in *Proceedings of the 2nd International Workshop on Environments for Multi-Agent Systems (E4MAS '05)*, Selected Revised and Invited Papers, pp. 18–31, Utrecht, The Netherlands, 2006.
- [12] M. Gardner, *On Cellular Automata, Self-Reproduction, and the Game "Life"*, Scientific American, 1971.
- [13] P. Rendell, *Turing Universality of the Game of Life*, Collision-Based Computing, Springer, London, UK, 2002.
- [14] S. Stepney, F. A. C. Polack, and H. R. Turner, "Engineering emergence," in *Proceedings of the 11th IEEE International Conference on Engineering of Complex Computer Systems (ICECCS '06)*, pp. 89–97, Stanford, Calif, USA, August 2006.
- [15] E. M. A. Ronald, M. Sipper, and M. S. Capcarrère, "Testing for emergence in artificial life," in *Advances in Artificial Life: 5th European Conference*, D. Floreano, J. D. Nicoud, and F. Mondada, Eds., vol. 1674 of *Lecture Notes in Artificial Intelligence*, pp. 13–20, Springer, Heidelberg, Germany, 1999.
- [16] J. P. Crutchfield, "The calculi of emergence: computation, dynamics, and induction," *Physica D*, vol. 75, no. 1–3, pp. 11–54, 1994.
- [17] J. Deguet, L. Magnin, and Y. Demazeau, "Emergence and software development based on a survey of emergence definitions," in *Emergent Intelligence of Networked Agents*, A. Namatame, S. Kurihara, and H. Nakashima, Eds., vol. 56 of *Studies in Computational Intelligence*, pp. 13–21, Springer, 2007.
- [18] L. Guo and X. Cai, "The fractal dimensions of complex networks," *Chinese Physics Letters*, vol. 26, no. 8, Article ID 088901, 2009.
- [19] E. G. Bakhoun and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.
- [20] E. G. Bakhoun 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.
- [21] F. Leon, S. Curteanu, C. Lisa, and N. Hurduc, "Machine learning methods used to predict the liquid-cristalline behavior of some copolyethers," *Molecular Crystals & Liquid Crystals*, vol. 469, pp. 1–22, 2007.
- [22] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [23] P. Erdős and A. Rényi, "On random graphs: I," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.
- [24] S. Milgram, "The small world problem," *Psychology Today*, vol. 2, pp. 60–67, 1967.
- [25] R. Albert, H. Jeong, and A. L. Barabási, "Diameter of the world-wide web," *Nature*, vol. 401, no. 6749, pp. 130–131, 1999.
- [26] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 99, no. 3, pp. 7314–7316, 2002.
- [27] M. E. Gaston and M. desJardins, "Social networks and multi-agent organizational performance," in *Proceedings of the 18th International Florida Artificial Intelligence Research Society Conference (FLAIRS '05)*, Special Track on AI for Social Networks, Social Networks for AI, Clearwater, Fla, USA, May 2005.
- [28] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," *Nature*, vol. 433, no. 7024, pp. 392–395, 2005.
- [29] F. Leon, *Intelligent Agents with Cognitive Capabilities*, Tehnopress, Iasi, Romania, 2006.
- [30] G. Di Marzo Serugendo, M. P. Gleizes, and A. Karageorgos, "Self-organisation and emergence in MAS: an overview," *Informatica*, vol. 30, no. 1, pp. 45–54, 2006.
- [31] X. Zheng and S. Koenig, "Reaction functions for task allocation to cooperative agents," in *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*, pp. 559–566, Estoril, Portugal, 2008.
- [32] K. Lerman and O. Shehory, "Coalition formation for large-scale electronic markets," in *Proceedings of the 4th International Conference on Multi-Agent Systems*, pp. 167–174, Boston, Mass, USA, 2000.
- [33] M. D. Weerd, Y. Zhang, and T. Klos, "Distributed task allocation in social networks," in *Proceedings of the 6th Autonomous Agents and Multiagent Systems*, pp. 500–507, Honolulu, Hawaii, USA, 2007.
- [34] D. Ye, Q. Bai, M. Zhang, K. T. Win, and Z. Shen, "An efficient task allocation protocol for P2P multi-agent systems," in *Proceedings of the IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA '09)*, pp. 11–18, Chengdu, China, August 2009.
- [35] X. Zheng and S. Koenig, "Greedy approaches for solving task-allocation problems with coalitions," in *Proceedings of the Workshop on Formal Models and Methods for Multi-Robot Systems (AAMAS '08)*, pp. 35–40, Estoril, Portugal, 2008.

- [36] A. Campbell and A. S. Wu, "Multi-agent role allocation: issues, approaches, and multiple perspectives," *Autonomous Agents and Multi-Agent Systems*, vol. 22, no. 2, pp. 317–355, 2010.
- [37] R. G. Smith, "The contract net protocol: high-level communication and control in a distributed problem solver," *IEEE Transactions on Computers*, vol. 29, no. 12, pp. 1104–1113, 1980.

Review Article

Traffic Dynamics on Complex Networks: A Survey

**Shengyong Chen,¹ Wei Huang,¹ Carlo Cattani,²
and Giuseppe Altieri²**

¹ College of Computer Science and Technology, Zhejiang University of Technology, 310023 Hangzhou, China

² Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy

Correspondence should be addressed to Shengyong Chen, sy@ieee.org

Received 20 May 2011; Accepted 28 June 2011

Academic Editor: Gani Aldashev

Copyright © 2012 Shengyong Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traffic dynamics on complex networks are intriguing in recent years due to their practical implications in real communication networks. In this survey, we give a brief review of studies on traffic routing dynamics on complex networks. Strategies for improving transport efficiency, including designing efficient routing strategies and making appropriate adjustments to the underlying network structure, are introduced in this survey. Finally, a few open problems are discussed in this survey.

1. Introduction

Large real communication networked systems have become a hot research topic for a rather long time. Typical examples include the Internet, which is an enormous network of many routers connected by physical or wireless links with information packets flowing on them, and high-way network, which is composed of cities and high-ways between cities. The rapid development of society causes the immense increase of traffic amounts in many real networked communication systems. Congestion may firstly occur on some communication units (such as routers on the Internet [1, 2] or cities in high-way networks) and then spread to more other units. Therefore, it is important to recover the free-flow state from the congested state, and therefore, demands for high transport efficiency are becoming more stringent in recent years. Up till now, there have been many studies on understanding and controlling traffic congestion on communication systems [3–18]. However, many early studies only assumed that the communication networked systems are completely random or even ignored the existence of network structure. Since the pioneering discovery of small-world phenomena [19] and scale-free feature [20, 21] at the end of 20th century, researchers began to realize that

traffic dynamics are highly relevant to the underlying network structure. For example, in 2005, Zhao et al. found that for two networks with different network structures, even when the two networks have the same average connectivity, the same delivering capacity of each node, and the same number of nodes, traffic congestion may tend to occur in one network than in another [22]. Therefore, traffic dynamics on complex network have attracted a lot of interest in both applied physics and computational science.

Actually, many empirical studies have revealed that the transport performances are not only relevant to the characteristics of underlying network structure, but also significantly affected by routing strategies. In this light, in order to improve transport performances on real networks, one can either design efficient routing strategies, or make appropriate changes to the underlying network structure. Designing efficient routing strategies is considered to be “soft” strategies, because it does not require any topological changes. Making changes to network structure is considered to be “hard” strategies, because it requires topological changes. One can add a few links to existing networks or delete a few links from existing networks to realize the modification of network structure.

In this survey, we mainly review recent progresses for traffic dynamics on complex networks. We have to mention that this survey is mainly from the perspective of physics. The remainder of this survey is organized as follows. In Section 2, we introduce some basic models and concepts for traffic routing dynamics. In Sections 3 and 4, we present some important work on “soft” strategies and “hard” strategies, respectively. Finally, conclusions are drawn in Section 5 with a brief discussion for future work.

2. Models and Concepts for Traffic Dynamics

2.1. Traffic Routing Model

Processes of random walk on complex networks have been extensively studied recently due to its wide applications in real networks [19, 23, 24]. However, the accumulation of packets on routers is not involved in the process of random walk, and therefore, the process of random walk cannot reflect real traffic system completely. Under the background of complex network, a basic traffic dynamics model has been proposed and frequently used to mimic the traffic transport in communication networked systems [7, 22, 25–27]. In this basic model, all nodes in a network are equally considered as hosts and routers [26, 28] for generating and delivering packets. The whole traffic dynamics model is an iterated process. In a general setting, packets are generated with a give rate ρ at randomly selected nodes at each time step. Each packet is designated with a randomly chosen node, different from its source, as the destination to which the packet will be delivered. Each packet is delivered from one node to another following a given routing strategy. Each node has its own delivering capacity C , that is, the maximal number of packets each node can deliver at one time step. An arrived packet will be placed at the end of the queue if this node already has some packets to be delivered to their respective destinations. The packets in queue, which may be created locally at some previous time steps or they are delivered by other nodes in an earlier time, work on a “first-in-first-out” basis. Finally, a packet will be removed from the network once it reaches its destination.

The basic model has also been generalized to more realistic models, which incorporate the fact that hub nodes usually have high delivering capacities or can generate more packets at each unit time step than those low-degree nodes can do. For example, degree-dependent delivering capacity was assumed in the form $1 + k_i^\theta$ [29] or $1 + \beta k_i$ [22, 30]. In [30, 31], the

author assumed that degree-dependent packet generation rate in the form λk_i . In [32, 33], the authors assumed that the bandwidth B of each link, that is, the maximal capacity of each link for delivering packets, is limited and varies from link to link. Moreover, in [34], the authors assumed that sources and destinations are not homogeneously selected. They considered two situations of packet generation: (i) packets are more likely generated at high-degree nodes and (ii) packets are more likely generated at low-degree nodes. Similarly, they considered two situations of packet destination: (a) packets are more likely to go to high-degree nodes and (b) packets are more likely to go to low-degree nodes.

2.2. Traffic Capacity

One of the most important measurements for transport performance of traffic is the traffic capacity ρ_c , that is, the critical packet generation rate. At ρ_c , the network undergoes a phase transition from free-flow state to congested state. When the packet generation rate ρ is below ρ_c , the number of generated and delivered packets are balanced, and therefore, the network is in free-flow state. But when ρ goes beyond ρ_c , the number of packets keeps on increasing with time and can lead to congestion finally, simply because nodes cannot delivering too many packets at each time step due to limited delivering capacity. The traffic capacity is usually described by an order parameter [6]

$$\eta(\rho) = \lim_{t \rightarrow \infty} \frac{C\langle \Delta W \rangle}{\rho N \Delta t}, \quad (2.1)$$

where $\langle \Delta W \rangle = W(t+1) - W(t)$ and $\langle \dots \rangle$, the average over time windows of width Δt . $W(t)$ is denoted as the number of packets in the network, and N is the network size. When $\rho < \rho_c$, $\Delta W = 0$, and $\eta = 0$, this indicates that the network system is under the free-flow state. On the other hand, when $\rho > \rho_c$, η is above zero, which indicates that packets are accumulating in the network and the network will become congested. Thus, ρ_c is the maximal packet generation rate under which the network system can remain in the free-flow state.

2.3. Effective Betweenness

In order to provide a theoretical estimate of traffic capacity, the concept "betweenness" is introduced here. Betweenness centrality [35–38], or betweenness for short, measures how central a node is in the network. Originally, the betweenness of a node l is defined as the number of total shortest paths that pass through the node l . Let $\sigma_{ij}(l)$ be the number of paths going through node l and following the shortest paths between node i and node j . The node betweenness, B_l , can be expressed as $B_l = \sum_{ij} \sigma_{ij}(l)$, where $i, j \neq l$. Later on, the definition of node betweenness was extended. The path between any two distinct nodes may not be along the shortest path between the two nodes but is guided by a given searching algorithm. This is the so-called effective node betweenness [26, 39]. When the searching algorithm is able to find the shortest path between any two distinct nodes, the effective node betweenness recovers the original definition of node betweenness [35–38]. The effective node betweenness can be defined as [26, 39]

$$B_j = \sum_{i,m} b_{ij}^m, \quad (2.2)$$

where b_{ij}^m is the average times that a packet generated at i and with destination m that passes through j . In a network with N nodes, we define

$$\mathbf{b}^m = \begin{pmatrix} b_{11}^m & \cdots & b_{1N}^m \\ \vdots & \ddots & \vdots \\ b_{N1}^m & \cdots & b_{NN}^m \end{pmatrix}. \quad (2.3)$$

According to [26, 39],

$$\mathbf{b}^m = (\mathbf{I} - \mathbf{p}^m)^{-1} \mathbf{p}^m, \quad (2.4)$$

where \mathbf{I} is an identity matrix and

$$\mathbf{p}^m = \begin{pmatrix} p_{11}^m & \cdots & p_{1N}^m \\ \vdots & \ddots & \vdots \\ p_{N1}^m & \cdots & p_{NN}^m \end{pmatrix}, \quad (2.5)$$

where p_{ij}^m is the probability that a packet whose destination points to node m goes from node i to a new node j in the next movement. Under the transportation rule regulated, from [26, 39], we have

$$\begin{aligned} p_{ij}^m &= a_{im} \delta_{jm} + (1 - a_{im} - \delta_{im}) P_{i \rightarrow j, m} \\ &= a_{im} \delta_{jm} + (1 - a_{im} - \delta_{im}) \frac{\tilde{k}_j^\alpha \tilde{d}_{j, m}^\beta}{\sum_{l \in \text{Nei}(i)} \tilde{k}_l^\alpha \tilde{d}_{l, k}^\beta}, \end{aligned} \quad (2.6)$$

where a_{im} is the element in the adjacency matrix of the network ($a_{im} = 1$ if there is a direct edge between node i and node m ; otherwise, $a_{im} = 0$) and δ_{jm} is the delta function. It should be noted that $p_{mj}^m = 0$ for all j , which means that a packet will be removed from the network when it reaches its destination, and $p_{im}^m = 1$ for all $i \neq m$, which means a packet will be directly forwarded to the destination node m if node m is a direct neighbor of node i .

Then, we can make a theoretical estimate for the traffic capacity ρ_c . The number of packets that arrive at node i is, on average, $\rho B_i / (N - 1)$. If $\rho B_i / (N - 1) > C$, packets will be accumulated on the node i , and congestion will thereafter occur. To avoid congestion on the network, $\rho B_i / (N - 1) \leq C$ should hold for each node. Therefore, the traffic capacity can be estimated as [22, 26]

$$\rho_c = \frac{C(N - 1)}{B_{\max}}, \quad (2.7)$$

where B_{\max} is the maximal effective betweenness in the network.

3. “Soft” Strategies

3.1. Shortest Path Routing Strategy

Under the shortest path routing strategy, each packet is always transported along the topological shortest path between the packet’s source and destination. Different from routing strategies with stochastic factors, under the shortest path routing strategy, each packet has a fixed delivering path once the network is constructed. Therefore, the shortest path routing strategy is widely used in real communication systems [2, 40–42] due to its economical and technical costs. However, under the shortest path routing strategy, packets are easy to pass through hub nodes, which can easily lead to congestions on hub nodes [29, 43, 44]. This fact consequently motivates the designing of many other efficient routing strategies, which will be introduced in the following part.

3.2. Review of Improved Routing Strategies for Traffic Dynamics

It has been demonstrated that networks cannot handle heavy traffic if packets are always transported along the shortest path from source and destination [29, 43, 44]. To improve transport efficiency of packets in the network, Echenique et al. proposed a strategy in [45, 46], called traffic-awareness strategy, in which the factor of waiting time at the neighboring nodes is also considered in addition to the fact of shortest path length from the neighboring node to destination. The waiting time is regarded as the number of packets in the queue at a neighboring node at the time of decision [45, 46]. The authors investigated the maximal time $\langle T_{\max} \rangle$ it takes for a packet to travel from its source to destination and the traffic capacity ρ_c in complex heterogeneous networks. Results show that compared to the shortest path routing strategy, $\langle T_{\max} \rangle$ is shortened and ρ_c is enhanced by using the traffic-awareness strategy. Moreover, in [46], the authors mentioned that under the shortest path routing strategy, traffic jams appear and grow smoothly, as the amount of generated packets increases. However, as soon as traffic awareness conditions are taken into account, the jamming transition is reminiscent of a first-order phase transition. Therefore, we can say that traffic-awareness routing strategy can delay the appearance of congestion at the cost of a sudden jump to a highly jammed phase due to the lack of early warnings.

Later on, Zhang et al. in [30] proposed an efficient routing strategy that is based on the projected waiting time along the shortest path from a neighboring node to the destination. Figure 1 illustrates the comparing results of three routing strategies in heterogeneous complex networks. λ and β are parameters related to degree-dependent packet generation rate and degree-dependent packet delivering capacity. Results show that jamming is harder to occur using the strategy in [30], when compared with both the shortest path strategy and traffic-awareness strategy [45, 46]. The strategy in [30] has the advantage of spreading the packets among the nodes according to the degree of nodes.

In [29], the authors incorporated both the global shortest path length information and local degree information in the transport process of traffic, via two tunable parameters, α and β , to guide the routing of packets. In detail, at each time step, all packets move from their current position, i , to the next node in their path, j , with a probability Q_{ij} defined as

$$Q_{ij} = \frac{k_j^\alpha \exp[-\beta(d_{it} - d_{jt} - 1)]}{\sum_{l \in \Omega_i} k_l^\alpha \exp[-\beta(d_{it} - d_{lt} - 1)]}, \quad (3.1)$$

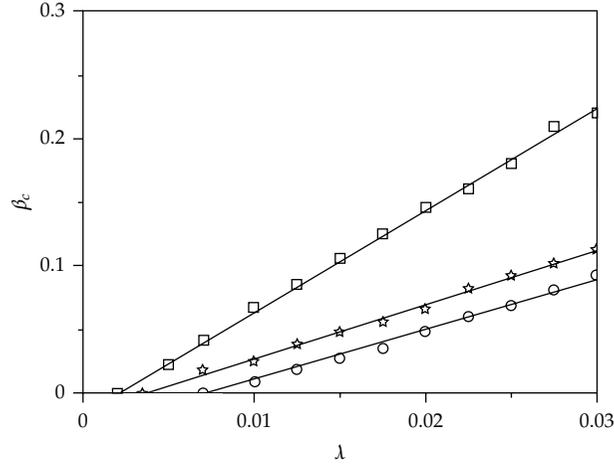


Figure 1: The critical value $\beta_c(\lambda)$ for three different routing strategies: the strategy proposed in [30] (circles), traffic-awareness strategy [45, 46] with $h = 0.8$ (stars) and the shortest path strategy (squares). The lines are guides to eye.

where Ω_i is the set of neighboring nodes of i , k_j is the degree of node j , and d_{it} is the shortest path length between node i and node t . The parameters α and β are tunable parameters with varying range $\alpha \in (-\infty, \infty)$ and $\beta \in [0, \infty)$. Through numerical simulations, the authors in [29] showed that with appropriate selection of the tunable parameters, the strategy in [29] is superior to the shortest path routing strategy in enhancing the traffic capacity. Furthermore, the authors of [29] pointed out that the strategy in [45, 46] is sensitive to the total number of tasks assigned to the networks. In the case of heavy traffic, the contribution of shortest path length for delivering packets can be ignored, because the diameter of most real communication networks grows only logarithmically with system size [1] and shortest path length is only bound to the diameter value. However, the strategy in [29] is insensitive to the amount of traffic flow in the network. In [39], based on the strategy which considered both shortest path length ingredient and degree ingredient in heterogeneous networks, the authors proposed an improved routing strategy with memory information in scale-free networks. This result was inspired by [47], in which researchers found that when studying local search in power-law communication networks, the search efficiency can be effectively enhanced if retracing the last step is disallowed in random networks. In the strategy with memory information [39], when node A receives a new packet from node B , node A has to record that the new packet is from node B . At the next time step, the packet node A is forbidden to be forwarded back to node B . In this way, the chance that packets are transported back and forth between two distinct nodes is greatly reduced. Figure 2 shows the results of performances in enhancing traffic capacity for both strategies with and without memory information. In Figure 2, α regulates the degree ingredient involved for forwarding packets, and β regulates the shortest path length ingredient involved for forwarding packets. γ denotes the degree exponent of scale-free networks. Results show that traffic capacity can be effectively enhanced by using the strategy with memory information.

In [44], the authors proposed a kind of efficient routing strategy. In order to find the optimal routing strategy, the authors define “the efficient path.” For any path between node i

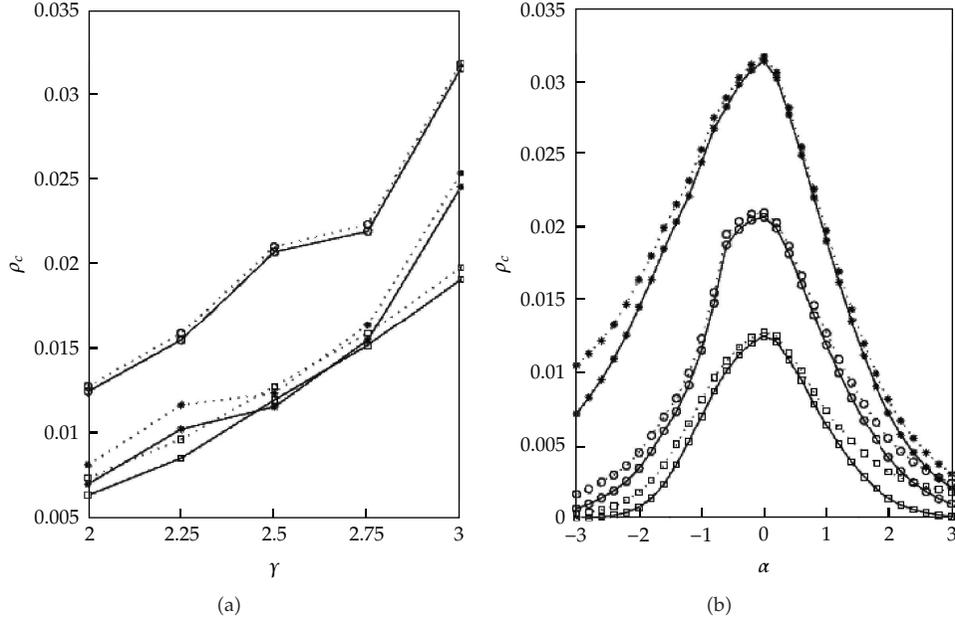


Figure 2: (a) The traffic capacity rate ρ_c versus the degree exponent γ with β equal to 1. The cases of $\alpha = -1$, 0, and 1 are marked with asterisks, circles, and squares, respectively. (b) The traffic capacity ρ_c versus α . The cases of $\gamma = 3$, 2.5, and 2 are marked with asterisks, circles, and squares, respectively. In both figures, dotted and solid lines are for the routing strategies with and without memory information. Network size is set to 500 in both figures [39].

and node j as $P(i \rightarrow j) := i \equiv x_0, x_1, \dots, x_{n-1}, x_n \equiv j$, define

$$L(P(i \rightarrow j) : \beta) = \sum_{i=0}^{n-1} k(x_i)^\beta, \quad (3.2)$$

where $k(x_i)$ is the degree of node x_i and β is a tunable parameter. The efficient path between i and j is corresponding to the route that makes the sum $L(P(i \rightarrow j) : \beta)$ minimal. It is obvious that $L(\beta = 0)$ recovers the shortest path routing strategy. It is expected that the system behaves better under the routing rule with $\beta > 0$ than under the shortest path routing strategy. In [44], the authors concerns how R_c [28], that is, the critical number of generated packets at each time step, varies with β . Here, $R_c = N\rho_c$. The simulation results for the critical value R_c as a function of β on BA scale-free networks are illustrated in Figure 3. It can be found that R_c firstly increases with β and then decreases, with the maximal R_c corresponding to $\beta = 1$. As compared to the shortest path routing strategy ($\beta = 0$), the traffic capacity is greatly improved, from $R_c \approx 4$ when $\beta = 0$ to $R_c \approx 45$ when $\beta = 1$, more than ten times. However, the authors also found that the average total delivering path length is maximal when $\beta = 1$. Therefore, they concluded that the system capability in processing information is considerably enhanced at the cost of increasing the average total delivering path length when using the strategy [30].

The routing strategies mentioned above, allow each node to have the whole network's global topological information, which may be practical for small or medium size networks

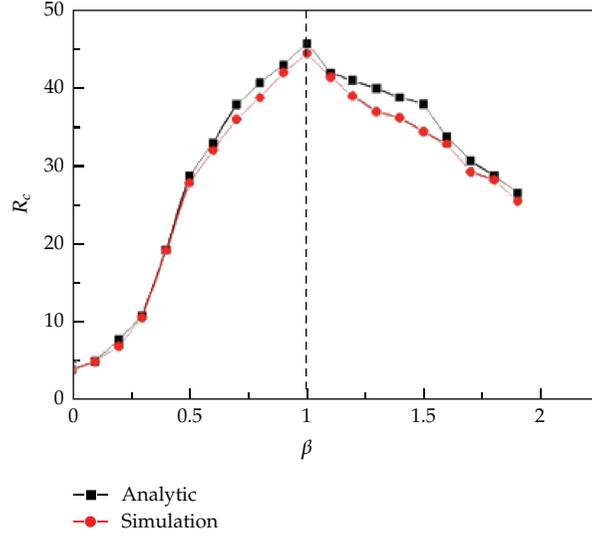


Figure 3: The critical R_c versus β for scale-free networks with size $N = 1225$. Both simulations and theoretical analysis indicate that the maximal value of R_c corresponding to $\beta \approx 1$. The data shown here is the average over 10 independent runs [44].

but not for large real communication such as the Internet, WWW [48], peer-to-peer networks [25, 49], or urban transportation system [50, 51]. However, strategies based on local information (each node only knows the information of its neighbors) are favored in large networks due to heavy communication cost on searching global information in networks. Inspired by the study on network search or network navigation on complex networks [47, 52–55], the authors in [56, 57] present a traffic model in which packets are routed only based on local topological information with a single tunable parameter α . The optimized value of α was sought out to maximize the traffic capacity. In this model, each node performs a local search among its neighbors. If the packet's destination is within the searched area, it is delivered directly to its destination. Otherwise, it is forwarded to a node i , one of the neighbors of the searching node, according to the preferential probability

$$\Pi_i = \frac{k_i^\alpha}{\sum_j k_j^\alpha}, \quad (3.3)$$

where the sum runs over all neighbors of the searing node i , k_i is the degree of node i , and α is a tunable parameter. With the delivering capacity of each node set as a constant C , the simulation results show that the optimal performance of the system corresponds to $\alpha = -1$, which can be observed from Figure 4. Theoretical analyses for the optimal value $\alpha = -1$ were obtained in [56]. The authors of [56] also pointed out that choosing the optimal $\alpha = -1$ not only maximizes the traffic capacity but also minimizes the average delivering time. When the delivering capacity of each node is proportional to its degree k , the optimal value of α changes to $\alpha = 0$, as demonstrated in Figure 5. This result indicates that under the conditions of heterogeneous delivering capacity of each node, the random walks strategy is the best choice for routing packets.

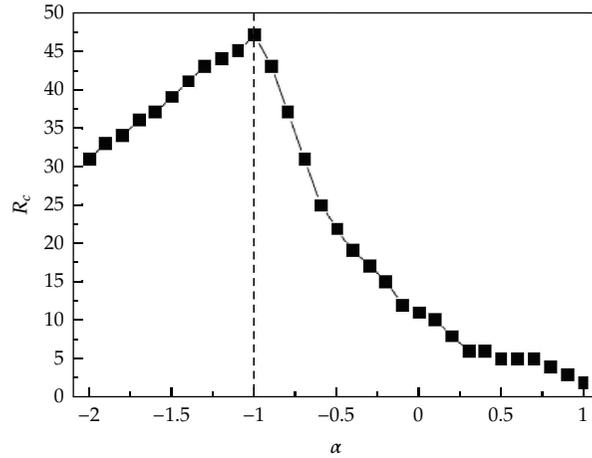


Figure 4: Critical R_c versus α on a BA scale-free network with network size $N = 1000$ and constant delivery capacity of each node $C = 10$. The maximal of R_c corresponds to $\alpha = -1$ marked by a dotted line [56].

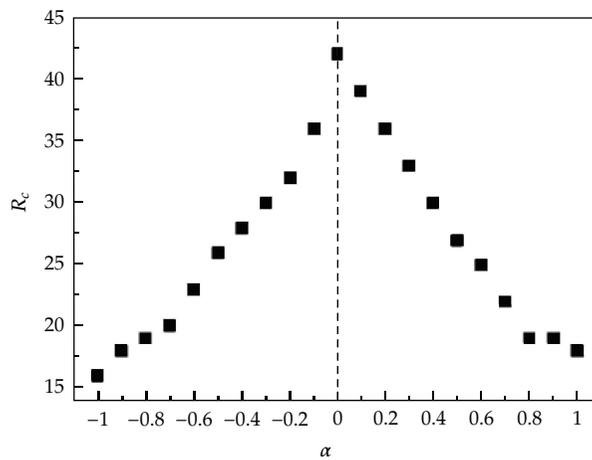


Figure 5: The critical generating rate R_c versus α in the case of delivery capacity of each node proportional to its degree $C_i = k_i$. Here, k_{\min} is set to be 3 and network size $N = 1000$. The maximum of R_c corresponds to $\alpha = 0$ marked by a dotted line [56].

Actually, the area of information that each router can access can significantly affect the performances of local routing strategy in traffic transport on complex networks. References [58, 59] found that the next-to-nearest routing algorithm can perform much better than the nearest routing algorithm under the random routing strategy. The next-to-near routing algorithm means that a packet can directly be delivered to its destination if the destination is one of the next-to-nearest neighbors of the searching node.

The strategy in [56, 57] is only based on local static information, which can only enhance the traffic capacity but cannot considerably reduce the transport time. In [49], the authors proposed a new routing strategy based on local static and dynamical information in scale-free networks. In the routing model of [49], it was assumed that each node has

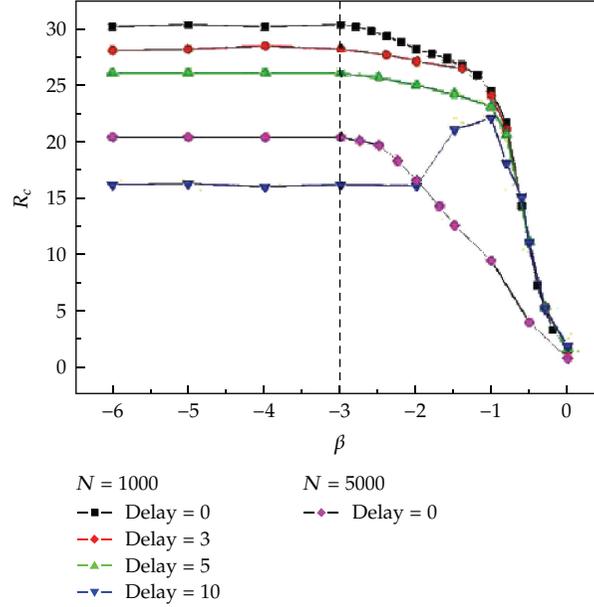


Figure 6: The network capacity R_c versus parameter β for different time delay and for different network size N . The other parameter is $C = 5$ [49].

an identical delivering capacity C . A packet is directly delivered to its destination if the destination is one of the direct neighbors of the searing node l . Otherwise, the probability of a neighbor node i to which the packet will be delivered is

$$P_{l \rightarrow i} = \frac{k_i(n_i + 1)^\beta}{\sum_j k_j(n_j + 1)^\beta}, \quad (3.4)$$

where the sum runs over all neighbors of the searching node i , n_i is the number of packets in the queue of i and β is a tunable parameter. Also, this model considers the effect of transmission delay on the traffic dynamics. The delay is defined as the number of time steps in receiving updated dynamical information from the neighbors [49]. Figure 6 shows that in the case of no delay, the traffic capacity is considerably reduced when decreasing β until $\beta \approx -3$. The maximal traffic capacity is only 23 when choosing $C = 5$ under the case of only adopting local topological information [56, 57]. Results show that the traffic capacity is further improved by using the strategy in [49] as compared to the strategy in [56, 57]. The average transport time $\langle T \rangle$ is reported in Figure 7. It can be seen that by adopting the strategy in [56, 57], $\langle T \rangle$ is approximately independent of R . When R is not too large, $\langle T \rangle$ is much shorter by adopting the strategy in [49] than the strategy in [56, 57]. In real communication system, the local dynamical information can be obtained by using the keep-alive messages that router continuously exchange with their peers [45].

In [60], the authors proposed another global routing strategy which considers both global topological information and queue length of each node. In this global strategy, the

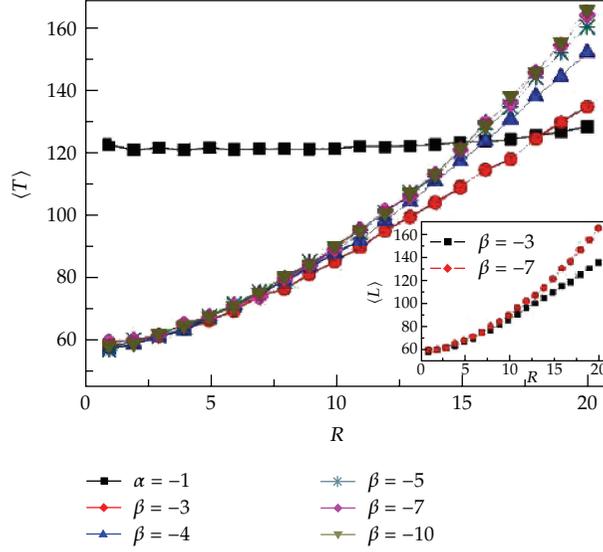


Figure 7: Average transport time $\langle T \rangle$ versus R for strategies in [49, 56, 57], respectively. $\alpha = -1$ corresponds to the optimal value of the strategy in [56, 57] and β is the parameter of the strategy in [49]. The network size is set to $N = 1000$ and the node delivering capacity is set to $C = 5$.

packets are delivered along the path in which the sum of queue length of nodes is a minimum. The path between node i (source) and node j (destination) can be denoted as

$$P_{ij} = \min \sum_{m=0}^l [1 + n(x_m)], \quad (3.5)$$

where $n(x_m)$ is the queue length of node and l is the path length. Figures 8(a)–8(c) show the evolution of the total packet number N_p under the shortest path strategy [29, 43, 44], the efficient strategy proposed in [44], and the global dynamic paths with the network size $N = 500$ and average degree $\langle k \rangle = 4$. Figure 8(d) compares the relation of order parameter η versus R under the three routing strategies. It can be seen that the traffic capacity under the shortest path strategy [29, 43, 44] is $R_c = 3$. The efficient strategy in [44] has $R_c = 20$ and the global dynamic strategy can reach up to $R_c = 41$. Therefore, the global dynamic strategy can achieve the highest traffic capacity.

Figure 9 shows the relation of traffic capacity R_c versus the average degree $\langle k \rangle$ and versus the network size N under the three routing strategies [60]. Results show that the traffic capacity under the global dynamic routing strategy is the largest. Because the node queue length changes from time to time, it is computationally time consuming to find the global paths at each time step. Therefore, the authors in [60] introduced a time delay δT for the update of the global queue information and the corresponding paths. They found that with the increment of time delay, the traffic capacity remains almost the same, but the total packet number in system, the traveling time, and the waiting time will increase.

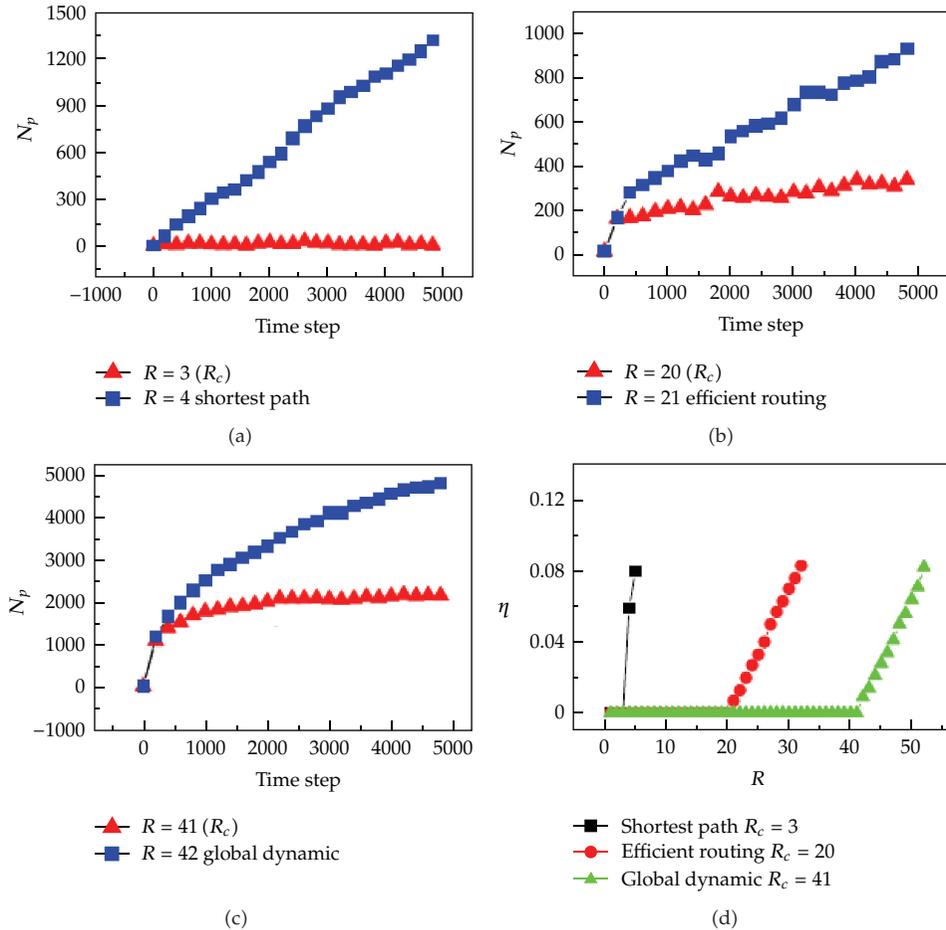


Figure 8: Evolution in the packet number in the network under different routing strategies [60]. (a) Shortest path routing strategy, (b) Efficient routing strategy in [44], and (c) The global dynamic routing strategy. (d) The order parameter η versus R under the three routing strategies.

4. “Hard” Strategies

4.1. Removing Links

“Hard” strategies mean network topological structure is appropriately changed so that transport efficiency can be improved. Adding or rewiring links are more costly than soft strategies (i.e., designing efficient routing strategies), because adding or rewiring links usually have to consume much financial, manpower or even energy cost. On the contrary, removing links from networks is usually easy to be implemented at low cost. For example, in a high-way network system, some road ways are usually closed at rush hours to alleviate congestion, especially when crowds of people are rushing to offices in the morning or rushing back home in the afternoon. To realize the closure of roadways, traffic administrators need only to block the entries to the roadways, which is easy to be implemented. Another example is the Internet. Network administrators can easily isolate some connections among computers through computer software.

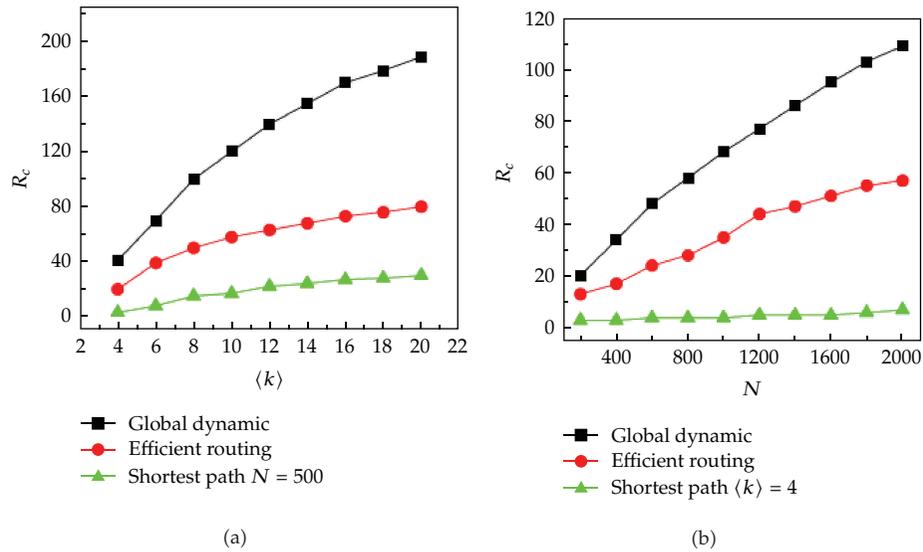


Figure 9: (a) The traffic capacity R_c versus average degree $\langle k \rangle$ with the same network size $N = 500$ under the three routing strategies. (b) The traffic capacity R_c versus the network size N with the same average degree $\langle k \rangle = 4$ under the three routing strategies [60].

In fact, the link removal strategy has been extensively studied to enhance or optimize dynamics of different kinds on complex networks. In [61], the authors pointed out that the removals of nodes or links could alleviate or even mitigate cascades of overloading on networks. In [62], the link-removal strategy was applied in the metabolic network. They pointed out that the removals of metabolic reactions, which represent links in the metabolic network, could improve metabolic performances and rescue defective metabolic networks. In [63], the authors found that the removals of links could also help to enhance synchronization in complex networks of dynamical systems.

Literatures on enhancing transport capacity by removing links in communication networks have also been reported. The strategies in [64, 65] were proposed to enhance the traffic capacity by removing certain links among hub nodes in scale-free networks. The strategy proposed in [64], also called high-degree-first (HDF) strategy, first ranks the links according to the value of the product $(k_m \times k_n)$, where k_m and k_n are the links' end-node degrees. Then, links are closed according to this order from large to small. Because hub nodes are usually more important and bear more traffic loads, the links with larger values of $(k_m \times k_n)$ are easier to jam. Hence, removing highly congested links can lead to the redistribution of traffic loads along links so as to enhance the overall packet handling and delivering capacity. For convenience, the number of removed links was denoted as L_c . Figure 10 shows the increment of R_c versus L_c under the shortest path routing strategy. Results show that on closing the links according to the order of $k_m \times k_n$, traffic capacity can be remarkably enhanced.

The authors of [64] also investigated the relationship of R_c versus L_c under the local routing strategy [56]. As shown in Figure 11, when L_c increases from 0 to 600, the maximal traffic capacity R_c^{\max} always emerges at $\alpha = -1$, but its value decreases from 10 to 5. However, when α is far from -1 , L_c increases with the number of removed links. In Internet, it has been found that there are fluctuations in information flow [26]. Therefore, the link-removal

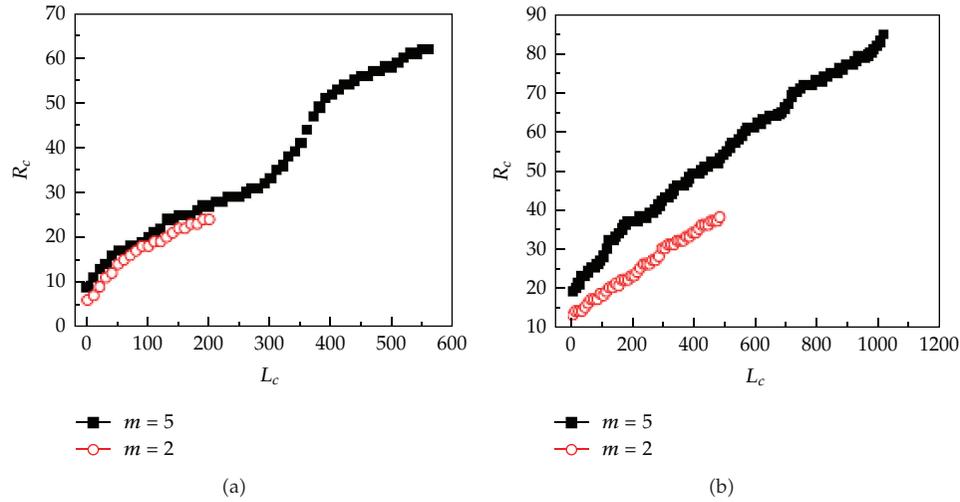


Figure 10: R_c versus L_c under the shortest path routing strategy with the network size (a) $N = 1000$ and (b) $N = 5000$. The data are obtained by averaging R_c over ten network realizations [64].

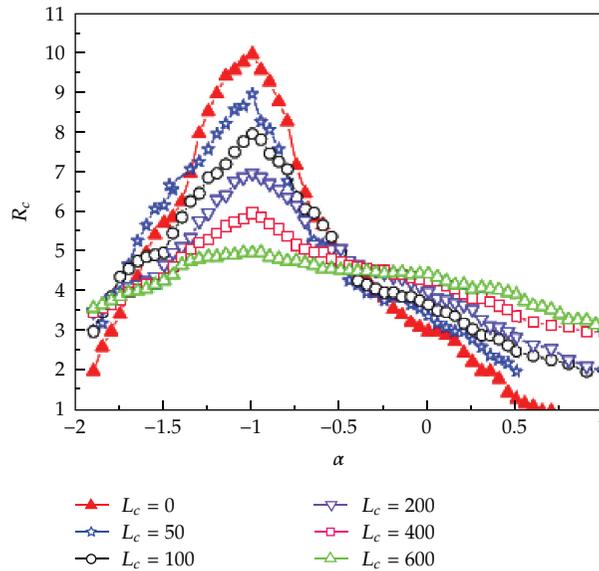


Figure 11: R_c versus L_c under the local routing strategy with the network size $N = 1000$. The data are obtained by averaging R_c over ten network realizations [64].

strategy can be applied to alleviate traffic congestion at times of high flux, and links can be recovered at times of low flux. The essence of the strategy in [65], also called high-betweenness-first (HBF) strategy, is much the same as that of the strategy in [64], but the strategy in [65] ranks the links according to the value of $(B_m \times B_n)$, where B_m and B_n are the links' end-node betweennesses.

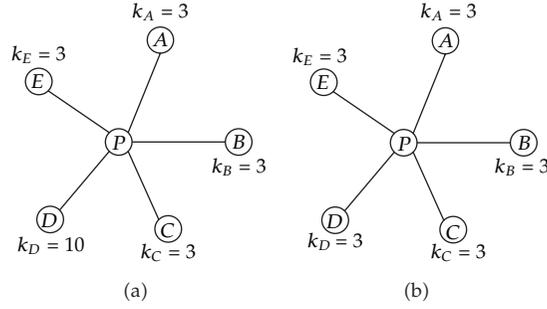


Figure 12: Illustration of uneven (a) and even (b) distributions of neighbors' degrees of a central node. Node P is the central node, and nodes $A, B, C, D,$ and E are five neighbors of node P . k_i is the degree of node i .

In [66], the authors pointed out the strategies of [64, 65] to identify the links to be removed in scale-free networks is far from optimal. On the basis of optimizing results which are obtained by using the SA (simulated annealing) algorithm [67, 68], the authors of [66] proposed a link-removal strategy. This new strategy, called variance-of-neighbor-degree-reduction (VNDR) strategy, not only considers the important role of hub nodes, but also balances the amount of packets from each node to its neighbors by reducing the variances of neighbors' degrees of some nodes. For node i , if the relative variance of neighbors' degrees (RVND), defined as

$$\text{rvar}(i) = \frac{\text{std}_{j \in \text{nei}(i)}(d(j))}{(1/|\text{nei}(i)|) \sum_{j \in \text{nei}(i)} d(j)}, \quad (4.1)$$

is large ($\text{nei}(i)$ is the neighborhood set of node i), then packets are more likely to be routed to hub nodes under the shortest path routing strategy. In (4.1),

$$\text{std}_{j \in \text{nei}(i)}(d(j)) = \sqrt{\frac{\left(d(j) - (1/|\text{nei}(i)|) \sum_{j \in \text{nei}(i)} d(j)\right)^2}{|\text{nei}(i)| - 1}}. \quad (4.2)$$

Hence, the hubs nodes are more likely to be congested. If the link between the hub node and the node with high RVND is removed, for example, the link between node P and node D is removed in the left figure of Figure 12, the distribution of neighbors' degree of the node with high RVND will become more even, and therefore, traffic amounts are balanced in the network. Finally, the traffic capacity is enhanced.

The VNDR link-removal strategy to enhance traffic capacity in scale-free networks is carried out as follows. For each node i in the initial network G , the authors in [66] defined

$$H(i) = \text{rvar}(i) \times \text{std}_{j \in \text{nei}(i)}(d(j)) \times d(i). \quad (4.3)$$

Choose the node p_1 with the maximal value of H . Then, choose the node q_1 with the highest degree from the neighborhood set of node p_1 . Remove the link connecting node p_1 and node

q_1 . Then a new network G_2 is generated. For the network G_2 , recompute H for all nodes and choose nodes p_2 and q_2 in the same way. The whole link-removal process is repeated until a predefined fraction f_r of the links is removed. All nodes have to be in one component in the whole process. If the removal of a link causes certain nodes to be disconnected from the original network, the process is no longer executed. Instead, go to deal with the link that connects the node p with the variable H ranked next and the node q with the highest degree in the neighborhood set of node p . The results using SA algorithm are also reported.

The performance of the VNDR strategy is evaluated by comparing it with the high-degree-first (HDF) strategy [64] and the high-betweenness-first (HBF) strategy [65]. In conducting the SA searching, the initial and final temperatures are set to 10 and 10^{-4} , respectively. The temperature decreasing coefficient is set to $\epsilon\% = 0.99$. For simplicity, the fraction of removed links is denoted as f_r . The results of traffic capacity ρ_c and average shortest path length L_{ave} versus f_r under the shortest path routing strategy are shown in Figure 13. It can be found that the VNDR strategy is superior to HDF and HBF link-removal strategies in enhancing the traffic capacity. The superiority of VNDR strategy over HDF and HBF strategies in enhancing traffic capacity under the shortest path routing strategy is rationalized in [66]. Furthermore, the authors in [66] also pointed out that under the shortest path routing strategy, the VNDR strategy can also reduce the average transport time as compared to the HDF and HBF strategies.

4.2. Adding Links

Link-removal strategy is operationally convenient and economical in real communication networks to enhance transport efficiency. However, with the rapid development of society, the sizes of real communication networks are consistently increasing, which brings new challenges to network administrators in charge of network communications. For instance, the numbers of roadways and cities keep on increasing in highway networks, and the number of computers is also increasing explosively in the Internet. But due to the cost incurred in adding new nodes and links to existing networks, real network designers have to be cautious when preparing to add new nodes and links to a given communication network. Otherwise, if new nodes and links are added to an existing network in an improper way, the new links may not be helpful for enhancing transport efficiency while packets are generated on both existing nodes and new nodes, which may aggravate congestion in the network. More links then need to be added into the networks to alleviate congestion with extra cost. Thus, it is necessary to investigate how to add nodes and links in an efficient way so that the traffic capacity can be enhanced maximally.

In [69], the authors proposed a strategy that can effectively enhance the traffic capacity via the process of adding nodes and links. They consider two cases. (1). The number of nodes is kept unchanged and only the number of links is allowed to be increased. (2) Both nodes and links are allowed to be added to the existing network. In the strategy of [69], shortcut links are added among nodes that have the longest shortest path lengths. The shortcut links are placed in proper positions to avoid packets flowing through hub nodes so that there are not too many packets accumulated on hub nodes.

In a scale-free network with N nodes and L links, for simplicity, the authors of [69] denote the fraction of new added links over the total L existing links as f_a . Figure 14 shows the traffic capacity R_c and the average shortest path length l_{ave} versus f_a under the shortest path routing strategy. From Figure 14(a), it is found that both the strategy of [69] and the

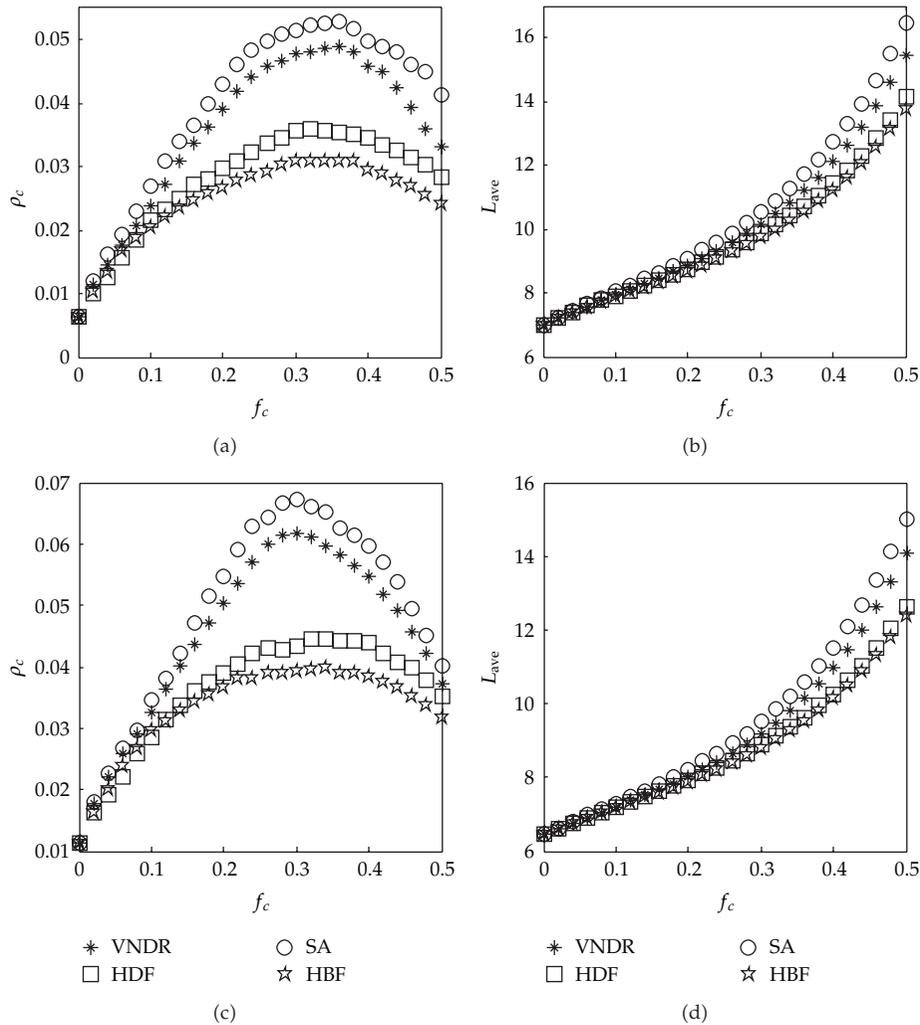


Figure 13: Illustration of ρ_c and L_{ave} versus f_r under the shortest path routing strategy. Network sizes are set to $N = 1000$ in (a), (b) and $N = 500$ in (c), (d) [66].

random strategy can enhance the traffic capacity R_c when links are added into networks. But the traffic capacity R_c is enhanced more using the strategy of [69] than using the random strategy.

For the case that both nodes and links are allowed to be added into networks under the shortest path routing strategy, the authors of [69] compared their strategy with the “degree preferential attachment” mechanism and the low-degree-first strategy. The results are shown in Figure 15. Results show that in the process of adding nodes and links to existing networks under the shortest path routing strategy, the strategy in [69] can effectively enhance the traffic capacity of networks at the cost of lengthening the average shortest path lengths.

It is worth noting that there is an underlying relation between cascading dynamic [70–73] and congestion of packet traffic [74]. Cascading dynamics means that breakdown on a global scale can be triggered by small failures on nodes through the mechanism

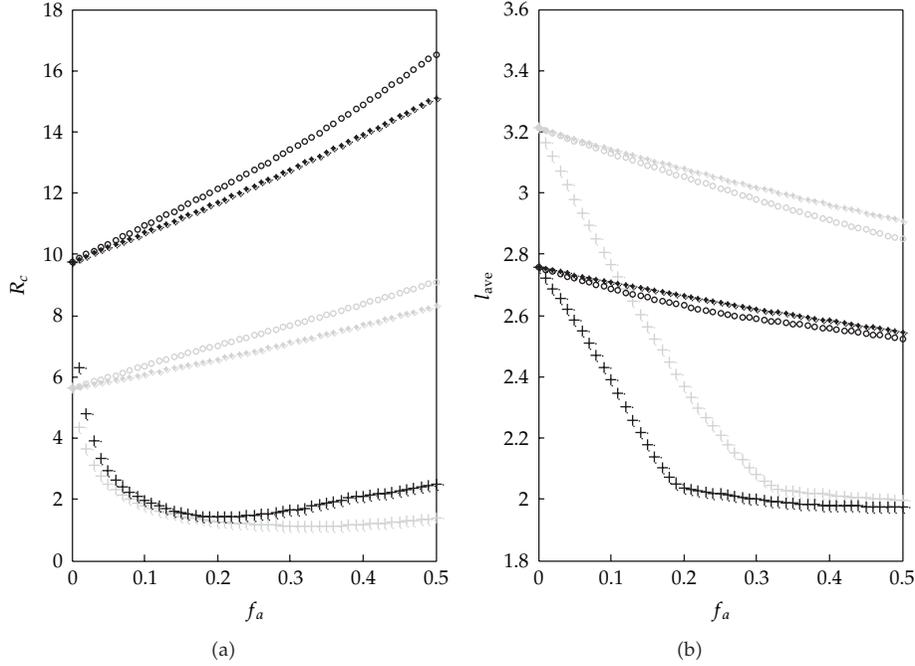


Figure 14: The traffic capacity R_c and the average shortest path length l_{ave} as a function of f_a under the shortest path routing strategy. In each figure, the gray and black colors are for the cases with $m_0 = m = 3$ and $m_0 = m = 5$, respectively. The network size is set to $N = 500$. The circles, asterisks, and plus signs are the results of the strategy in [69], the random strategy, and the strategy that adds links among hub nodes, respectively. Each result is obtained by averaging over fifty different network configurations and ten independent link-adding processes for each network realization when executing our strategy and the random strategy.

of cascading. Typical examples include large-scale blackouts of power grids and heavy congestion on the Internet and so forth. In [74], the authors pointed out that the smaller \overline{B}_m is, the more robust a network is against cascades. On the other hand, the traffic capacity, R_c , can be estimated as [26]

$$R_c = \frac{N(N-1)}{\overline{B}_m}. \quad (4.4)$$

This means that the maximal node betweenness can determine both the robustness against cascades and the traffic capacity of a network. Removing or adding links has been proved to be able to effectively enhance the traffic capacity, which means that \overline{B}_m can be effectively reduced. Therefore, the robustness against cascades of a network can also be effectively enhanced by using appropriate link-removal or link-adding strategies.

The key purpose of studying routing is to study performances of networks such as delay, admission control, and resource allocation [14, 15]. That is particularly true in real-time systems. Thus, this survey is only a conventional part of contents from a view of computer science. The recent contents are associated with network calculus. Moreover, the performances of networks for traffic transport are highly related to the scales observed. Interested reader can refer to [14–18] for their further studies.

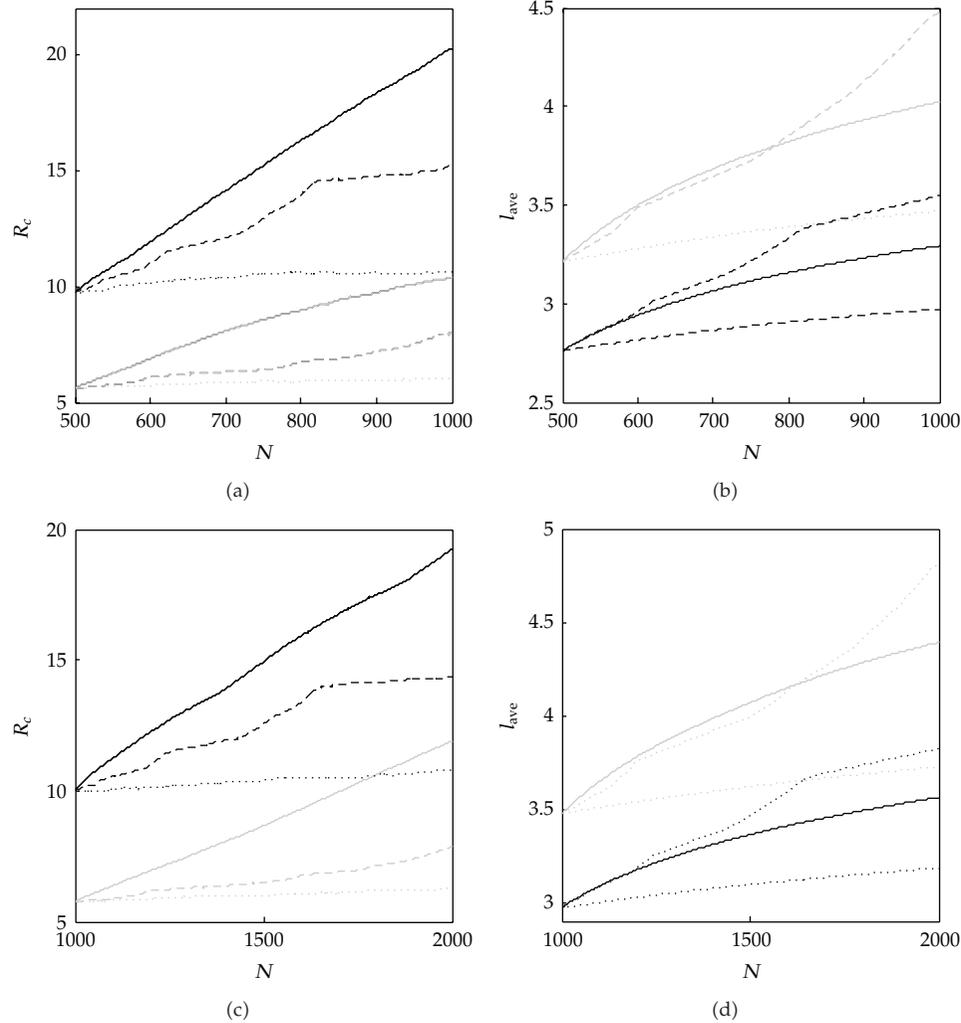


Figure 15: [69] The traffic capacity R_c and the average shortest path length l_{ave} as a function of the network size N under the shortest path routing strategy. Figures 15(a) and 15(b) start from scale-free networks with $N = 500$ nodes, while Figures 15(c) and 15(d) start from scale-free networks with $N = 1000$ nodes. Each result is obtained by averaging over fifty different network configurations. In each figure, the cases of $m_0 = m = 3$ and $m_0 = m = 5$ are marked with gray and black colors. Dotted, dashed, and solid lines are for the “degree preferential attachment” mechanism, the low-degree-first strategy, and our proposed strategy, respectively.

5. Concluding Remarks

The heavy traffic loads in real networked communication systems motivate the intense study of traffic dynamics on complex networks in recent years. The main goal is to enhance transport efficiency so that traffic congestions on complex networks can be alleviated. Generally speaking, the strategies which can enhance the traffic capacity can be categorized into two classes: designing efficient routing strategies and making appropriate adjustments to network structures. Many strategies have been proved to be able to enhance the transport

efficiency and are helpful for people in understanding and controlling traffic congestion in real networks.

Although the issues of enhancing transport efficiency have been extensively reported in the literatures, there are still many open questions to be studied. Real communication networks are usually much more complicated than the traffic routing model. Other dynamics such as the virus prevalence and cascading breakdown, may also be mixed with the traffic routing dynamics. A few open questions to be solved in future work are listed as follows.

- (1) The cascading breakdown dynamics can be incorporated with the traffic routing dynamics. A node or a link may lose its function in delivering packets due to traffic congestion on this node or link. Therefore, the node or the link will be removed from the network. As a result, the traffic capacity may be reduced. Under this circumstance, it is thus interesting to study the relationship between the two kinds of dynamics.
- (2) It is not always the case in real communication networks that each node in the network can generate packets. Some nodes (i.e., generators) may only generate and send packets to other nodes but do not receive packets from elsewhere. The rest of nodes (i.e., receivers) only receive packets from elsewhere but do not generate and send packets to other nodes. The positions of generators may have impacts on the transport efficiency of packets. It is necessary, therefore, to study how to place generators and receivers in networks so that the transport efficiency can be maximized.
- (3) Most current studies about traffic routing dynamics on complex networks is based on traffic routing models. However, due to complexity of real communication networks, traffic routing models cannot fully reflect the characteristics of real communication networks. More attention should be focused on revealing new characteristics of real communication networks rather than only study traffic routing models.

Acknowledgments

This work was supported by the National Natural Science Foundation of China and Microsoft Research Asia (NSFC-61173096, 60802087), Zhejiang Provincial S&T Department (2010R10006, 2010C33095), and Zhejiang Provincial Natural Science Foundation (R1110679).

References

- [1] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge University Press, Cambridge, UK, 2004.
- [2] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, "Dynamical and correlation properties of the internet," *Physical Review Letters*, vol. 87, no. 25, Article ID 258701, 4 pages, 2001.
- [3] H. Li and M. Maresca, "Polymorphic-torus network," *IEEE Transactions on Computers*, vol. 38, no. 9, pp. 1345–1351, 1989.
- [4] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic," *Computer Communication Review*, vol. 23, article 283, 1993.
- [5] T. Ohira and R. Sawatari, "Phase transition in a computer network traffic model," *Physical Review E*, vol. 58, no. 1, pp. 193–195, 1998.
- [6] A. Arenas, A. Díaz-Guilerà, and R. Guimerà, "Communication in networks with hierarchical branching," *Physical Review Letters*, vol. 86, no. 14, pp. 3196–3199, 2001.

- [7] R. Guimerà, A. Arenas, A. Díaz-Guilerà, and F. Giralt, "Dynamical properties of model communication networks," *Physical Review E*, vol. 66, no. 2, Article ID 026704, 8 pages, 2002.
- [8] T. M. Janaki and N. Gupte, "Connectivity strategies to enhance the capacity of weight-bearing networks," *Physical Review E*, vol. 67, no. 2, Article ID 021503, 6 pages, 2003.
- [9] R. V. Solé and S. Valverde, "Information transfer and phase transitions in a model of internet traffic," *Physica A*, vol. 289, no. 3-4, pp. 595–605, 2001.
- [10] S. Valverde and R. V. Solé, "Self-organized critical traffic in parallel computer networks," *Physica A*, vol. 312, no. 3-4, pp. 636–648, 2002.
- [11] D. K. Arrowsmith, R. J. Mondragon, J. M. Pittsy, and M. Woolf, Report 08 Institut Mittag-Leffler, Stockholm, Sweden, 2004.
- [12] M. S. Taqqu, W. Willinger, and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling," vol. 27, article 5.
- [13] M. E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, 1997.
- [14] X. Jia, W. Zhao, and J. Li, "An integrated routing and admission control mechanism for real-time multicast connections in ATM networks," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1515–1519, 2001.
- [15] M. Li and W. Zhao, "Asymptotic identity in min-plus algebra: a report on CPNS," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 154038, 11 pages, 2012.
- [16] M. Li and W. Zhao, "Visiting power laws in cyber-physical networking systems," *Mathematical Problems in Engineering*, vol. 2012, 2012.
- [17] 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.
- [18] 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.
- [19] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [20] A. L. Barabási, R. Albert, and A. Vespignani, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [21] A. L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A*, vol. 272, no. 1, pp. 173–187, 1999.
- [22] L. Zhao, Y. C. Lai, K. Park, and N. Ye, "Onset of traffic congestion in complex networks," *Physical Review E*, vol. 71, no. 2, Article ID 026125, 8 pages, 2005.
- [23] J. D. Noh and H. Rieger, "Random walks on complex networks," *Physical Review Letters*, vol. 92, no. 11, Article ID 118701, 4 pages, 2004.
- [24] Z. Eisler and J. Kertész, "Random walks on complex networks with inhomogeneous impact," *Physical Review E*, vol. 71, no. 5, Article ID 057104, pp. 1–4, 2005.
- [25] B. Tadić, S. Thurner, and G. J. Rodgers, "Traffic on complex networks: towards understanding global statistical properties from microscopic density fluctuations," *Physical Review E*, vol. 69, no. 3, Article ID 036102, 5 pages, 2004.
- [26] R. Guimerà, A. Díaz-Guilerà, F. Vega-Redondo, A. Cabrales, and A. Arenas, "Optimal network topologies for local search with congestion," *Physical Review Letters*, vol. 89, no. 24, Article ID 248701, 4 pages, 2002.
- [27] G. Mukherjee and S. S. Manna, "Phase transition in a directed traffic flow network," *Physical Review E*, vol. 71, no. 6, Article ID 066108, pp. 1–6, 2005.
- [28] T. Zhou, G. Yan, B. H. Wang et al., "Traffic dynamics on complex networks," *Dynamics of Continuous, Discrete and Impulsive Systems Series B*, vol. 13, no. 3-4, pp. 463–469, 2006.
- [29] Z. Wu, G. Peng, W. Wong, and K. Yeung, "Improved routing strategies for data traffic in scale-free networks," *Journal of Statistical Mechanics*, vol. 2008, no. 11, Article ID P11002, 2008.
- [30] H. Zhang, Z. H. Liu, M. Tang, and P. M. Hui, "An adaptive routing strategy for packet delivery in complex networks," *Physics Letters, Section A*, vol. 364, no. 3-4, pp. 177–182, 2007.
- [31] Z. Chen and X. Wang, "Effects of network structure and routing strategy on network capacity," *Physical Review E*, vol. 73, no. 3, Article ID 036107, pp. 1–5, 2006.
- [32] M. B. Hu, B. H. Wang, R. Jiang, Q. S. Wu, and Y. H. Wu, "The effect of bandwidth in scale-free network traffic," *Europhysics Letters*, vol. 79, no. 1, Article ID 14003, 2007.
- [33] M. B. Hu, R. Jiang, Y. H. Wu, and Q. S. Wu, "The effect of link and node capacity on traffic dynamics in weighted scale-free networks," in *Proceedings of the International Conference on Complex Sciences*,

- vol. 4, part 1 of *Lecture Notes of the Institute for Computer Sciences and Telecommunications*, pp. 580–588, Springer, Berlin Heidelberg, Germany, 2009.
- [34] R. Jiang, M. B. Hu, W. X. Wang, G. Yan, Q. S. Wu, and B. H. Wang, "Traffic dynamics of packets generated with none-homogeneously selected sources and destinations in scale-free networks," *Dynamics of Continuous, Discrete and Impulsive Systems B Supplement*, vol. 14, supplement 7, pp. 51–54, 2007.
- [35] M. E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Physical Review E*, vol. 64, no. 1, Article ID 016132, 2001.
- [36] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, Article ID 026113, 2004.
- [37] K. I. Goh, B. Kahng, and D. Kim, "Universal behavior of load distribution in scale-free networks," *Physical Review Letters*, vol. 98, article 278701, 2004.
- [38] M. Barthélemy, "Betweenness centrality in large complex networks," *European Physical Journal B*, vol. 38, no. 2, pp. 163–168, 2004.
- [39] W. Huang and T. W. S. Chow, "Investigation of both local and global topological ingredients on transport efficiency in scale-free networks," *Chaos*, vol. 19, no. 4, Article ID 043124, 2009.
- [40] D. Krioukov, K. Fall, and X. Yang, "Compact routing on internet-like graphs," in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies*, pp. 209–219, March 2004.
- [41] A. S. Tanenbaum, *Computer Network*, Prentice Hall Press, 1996.
- [42] C. Huitema, *Routing in the Internet*, Prentice Hall, Upper Saddle River, NJ, USA, 2000.
- [43] B. Danila, Y. Yu, J. A. Marsh, Z. Toroczkai, and K. E. Bassler, "Transport optimization on complex networks," *Chaos*, vol. 17, no. 2, Article ID 026102, 2007.
- [44] G. Yan, T. Zhou, B. Hu, Z. Q. Fu, and B. H. Wang, "Efficient routing on complex networks," *Physical Review E*, vol. 73, no. 4, Article ID 046108, pp. 1–5, 2006.
- [45] P. Echenique, J. Gomez-Gardenes, and Y. Moreno, "Improved routing strategies for Internet traffic delivery," *Physical Review E*, vol. 70, no. 5, Article ID 056105, 2004.
- [46] P. Echenique, J. Gomez-gardenes, and Y. Moreno, "Dynamics of jamming transitions in complex networks," *Europhysics Letters*, vol. 71, no. 2, pp. 325–331, 2005.
- [47] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in power-law networks," *Physical Review E*, vol. 64, no. 4, Article ID 046135, 2001.
- [48] R. Albert, H. Jeong, and A. L. Barabási, "Diameter of the world-wide web," *Nature*, vol. 401, no. 6749, pp. 130–131, 1999.
- [49] W. X. Wang, C. Y. Yin, G. Yan, and B. H. Wang, "Integrating local static and dynamic information for routing traffic," *Physical Review E*, vol. 74, no. 1, Article ID 016101, 2006.
- [50] M. B. Hu, R. Jiang, Y. H. Wu, W. X. Wang, and Q. S. Wu, "Urban traffic from the perspective of dual graph," *European Physical Journal B*, vol. 63, no. 1, pp. 127–133, 2008.
- [51] S. Scellato, L. Fortuna, M. Frasca, J. Gómez-Gardeñes, and V. Latora, "Traffic optimization in transport networks based on local routing," *European Physical Journal B*, vol. 73, no. 2, pp. 303–308, 2010.
- [52] J. M. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, article 845, 2000.
- [53] B. J. Kim, C. N. Yoon, S. K. Han, and H. Jeong, "Path finding strategies in scale-free networks," *Physical Review E*, vol. 65, no. 2, Article ID 027103, 2002.
- [54] C. P. Herrero, "Self-avoiding walks on scale-free networks," *Physical Review E*, vol. 71, no. 1, Article ID 016103, 2005.
- [55] S. J. Yang, "Exploring complex networks by walking on them," *Physical Review E*, vol. 71, no. 1, Article ID 016107, 2005.
- [56] W. X. Wang, B. H. Wang, C. Y. Yin, Y. B. Xie, and T. Zhou, "Traffic dynamics based on local routing protocol on a scale-free network," *Physical Review E*, vol. 73, no. 2, Article ID 026111, 2006.
- [57] C. Y. Yin, B. H. Wang, W. X. Wang, T. Zhou, and H. J. Yang, "Efficient routing on scale-free networks based on local information," *Physics Letters, Section A*, vol. 351, no. 4-5, pp. 220–224, 2006.
- [58] B. Tadić and G. J. Rodgers, "Packet transport on scale-free networks," *Advances in Complex Systems*, vol. 5, article 445, 2002.
- [59] B. Tadić and S. Thurner, "Information super-diffusion on structured networks," *Physica A*, vol. 332, no. 1–4, pp. 566–584, 2004.
- [60] X. Ling, M. B. Hu, R. Jiang, and Q. S. Wu, "Global dynamic routing for scale-free networks," *Physical Review E*, vol. 81, no. 1, Article ID 016113, 2010.
- [61] A. E. Motter, "Cascade control and defense in complex networks," *Physical Review Letters*, vol. 93, no. 9, Article ID 98701, 2004.

- [62] A. E. Motter, N. Gulbahce, E. Almaas, and A. L. Barabási, "Predicting synthetic rescues in metabolic networks," *Molecular Systems Biology*, vol. 4, article 168, 2008.
- [63] T. Nishikawa and A. E. Motter, "Network synchronization landscape reveals compensatory structures, quantization, and the positive effect of negative interactions," *The Proceedings of the National Academy of Sciences USA*, vol. 107, no. 23, Article ID 10342, 2010.
- [64] Z. Liu, M. B. Hu, R. Jiang, W. X. Wang, and Q. S. Wu, "Method to enhance traffic capacity for scale-free networks," *Physical Review E*, vol. 76, no. 3, Article ID 037101, 2007.
- [65] G. Q. Zhang, D. Wang, and G. J. Li, "Enhancing the transmission efficiency by edge deletion in scale-free networks," *Physical Review E*, vol. 76, no. 1, Article ID 017101, 2007.
- [66] W. Huang and T. W. S. Chow, "An efficient strategy for enhancing traffic capacity by removing links in scale-free networks," *Journal of Statistical Mechanics*, vol. 2010, no. 1, Article ID P01016, 2010.
- [67] T. J. P. Penna, "Traveling salesman problem and Tsallis statistics," *Physical Review E*, vol. 51, no. 1, pp. R1–R3, 1995.
- [68] D. A. Stariolo and C. Tsallis, *Annual Review of Computational Physics II*, Edited by D. Stauffer, World Scientific, Singapore, 1994.
- [69] W. Huang and T. W. S. Chow, "Effective strategy of adding nodes and links for maximizing the traffic capacity of scale-free network," *Chaos*, vol. 20, no. 3, Article ID 033123, 2010.
- [70] A. E. Motter and Y. C. Lai, "Cascade-based attacks on complex networks," *Physical Review E*, vol. 66, no. 6, Article ID 065102, 2002.
- [71] L. Zhao, K. Park, and Y. C. Lai, "Attack vulnerability of scale-free networks due to cascading breakdown," *Physical Review E*, vol. 70, no. 3, Article ID 035101, 2004.
- [72] A. E. Motter, "Cascade control and defense in complex networks," *Physical Review Letters*, vol. 93, no. 9, Article ID 098701, 2004.
- [73] L. Zhang, K. Park, Y. C. Lai, and N. Ye, "Tolerance of scale-free networks against attack-induced cascades," *Physical Review E*, vol. 72, no. 2, Article ID 025104, pp. 1–4, 2005.
- [74] R. Yang, W. X. Wang, Y. C. Lai, and G. Chen, "Optimal weighting scheme for suppressing cascades and traffic congestion in complex networks," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 79, no. 2, Article ID 026112, 2009.

Research Article

A Model to Partly but Reliably Distinguish DDOS Flood Traffic from Aggregated One

Ming Li¹ and Wei Zhao²

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, Avenue Padre Tomas Pereira, Taipa, Macau SAR, China

Correspondence should be addressed to Ming Li, ming_lihk@yahoo.com

Received 23 April 2011; Accepted 7 June 2011

Academic Editor: Shengyong Chen

Copyright © 2012 M. Li and W. Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Reliable distinguishing DDOS flood traffic from aggregated traffic is desperately desired by reliable prevention of DDOS attacks. By reliable distinguishing, we mean that flood traffic can be distinguished from aggregated one for a predetermined probability. The basis to reliably distinguish flood traffic from aggregated one is reliable detection of signs of DDOS flood attacks. As is known, reliably distinguishing DDOS flood traffic from aggregated traffic becomes a tough task mainly due to the effects of flash-crowd traffic. For this reason, this paper studies reliable detection in the underlying DiffServ network to use static-priority schedulers. In this network environment, we present a method for reliable detection of signs of DDOS flood attacks for a given class with a given priority. There are two assumptions introduced in this study. One is that flash-crowd traffic does not have all priorities but some. The other is that attack traffic has all priorities in all classes, otherwise an attacker cannot completely achieve its DDOS goal. Further, we suppose that the protected site is equipped with a sensor that has a signature library of the legitimate traffic with the priorities flash-crowd traffic does not have. Based on those, we are able to reliably distinguish attack traffic from aggregated traffic with the priorities that flash-crowd traffic does not have according to a given detection probability.

1. Introduction

Attackers may take the advantages of the principles [1] of distributed systems (i.e., the internet), such as openness, resources sharing, assessability, and so on, to launch distributed denial of service (DDOS) attacks. The threats of DDOS attacks to the individuals are severe. For instance, any denial of service of a bank server implies a loss of money, disgruntling or losing customers.

According to the classification of the CERT Coordination Center (CERT/CC), DDOS attacks are divided into three categories [2]: (1) flood (i.e., bandwidth) attacks, (2) protocol attacks, and (3) logical attacks. This paper considers flood attacks. DDOS flood attacks consume resources (e.g., bandwidth) by sending flood packets in order to shut down the target or significantly degrade its performance. The flood packets may be generated by hundreds or thousands of machines distributed all over the world.

A network-based intrusion detection system (IDS) monitors the traffic on its network as a data source [3]. In this regard, there are two main approaches. One is misuse detection and the other anomaly detection. Solutions given by misuse detection are primarily based on a library of known signatures to match against network traffic. Hence, unknown signatures from new variants of an attack mean 100% miss positives. As a matter of fact, the form in which an attack takes place is usually determined by a large number of details many of which are unknown. This is particularly true for DDOS attacks [4]. Hence, anomaly detectors play a role in DDOS detection [2, 3, 5–12]. Anomaly detectors cannot replace signature-based systems [2, 3]. From a practical view, therefore, the combination of a signature-based system and anomaly detector is worth noting [2].

A traffic series is a packet flow. A packet consists of a number of fields, such as protocol, source IP, destination IP, ports, flag setting (in the case of TCP or UDP), message type (in the case of ICPM), timestamp, and length (packet size). Each may serve as a feature of a packet for statistical detection purpose, see for example, [8, 13–15]. In addition, there are other available features of traffic, such as flow rate [16], the number of connections [17], and so on [6, 11, 12]. This paper takes traffic series in packet size (traffic series for short) as a monitored objective.

Usually, detections are expected to be adaptable to a wide range of network environments (e.g., [7, 8, 11–17]). Nevertheless, it is obviously worth studying detections that are environment dependent. This paper studies detecting signs of DDOS flood attacks in the underlying network to use static-priority schedulers.

As known, two tough issues in detecting DDOS flood attacks are (1) reliable detection as can be seen from [2, 3, 5, 7, 9, 10], and (2) distinguishing attack traffic from aggregated traffic [7, 9, 16]. The solution to the first issue is crucial to practical applications because false positives can lead to inappropriate responses that cause denial of service to legitimate traffic. In addition, it is the basis to find the solution to the second.

It is noted that flash-crowd traffic and DDOS flood traffic may have similar statistics from a network view. DDOS flood is malicious but flash crowds legitimate. Flash crowds happen when a huge number of users try to access the same server simultaneously for some specific events (e.g., the NASA Pathfinder mission) [16]. Because an attacker aims at attacking the target such that it denies services of all legitimate traffic, we assume DDOS flood traffic has all priorities in all classes. On the other hand, according to the nature of differentiated services, we assume that flash-crowd traffic does not have all priorities. Further, we suppose that the protected site is equipped with a sensor that has a signature library of the legitimate traffic with the priorities flood crowds do not have. In these cases, DDOS flood attack traffic can be distinguished, according to a given detection probability, from aggregated traffic with the priorities flash crowds do not have.

The rest of paper is organized as follows. Section 2 introduces the randomized traffic regulator for feature extraction of arrival traffic. Section 3 considers the principle. A case study is demonstrated in Section 4; discussions are given in Section 5 and conclusions in Section 6.

2. Traffic Regulator and Its Randomization

There are two major areas of traffic modeling. One is based on random processes, see for example, [6, 8, 18–30]. The other is deterministically modeling, for example, traffic regulator [18, 30–33]. We take traffic regulator to characterize traffic in this research.

Definition 2.1 (see [31, 33]). Let $y(t)$ be the instantaneous rate of arrival traffic at time t . Then, the amount of traffic generated in the interval $[t_1, t_2]$ is upper bounded by

$$\int_{t_1}^{t_2} y(t) dt \leq \sigma + \rho(t_2 - t_1), \quad (2.1)$$

where σ and ρ are constants and $t_2 > t_1$. This property is written as $y \sim (\sigma, \rho)$ that is called traffic regulator.

Practically, traffic is considered in the discrete case on an interval-by-interval basis. Thus, we generalize Definition 2.1 as follows.

Definition 2.2. Let $y(t)$ be the instantaneous rate of arrival traffic at t . Then, the amount of traffic generated in the n th interval $[(n-1)I, nI]$ ($n = 1, 2, \dots, N$) is upper bounded by

$$\sum_{t=(n-1)I}^{nI} y(t) \leq \sigma(n, I) + \rho(n, I)I, \quad (2.2)$$

where $(\sigma(I, n), \rho(I, n))$ represents the traffic regulator in the n th interval, and I is a positively real number.

For the simplicity, denote $F(I, n) = \sigma(I, n) + \rho(I, n)I$.

Definition 2.3. Let $y_{p,j,k}^i(t)$ be the instantaneous rate of all flows of class i with priority p going through server k from input link j at t . Then, the amount of $y_{p,j,k}^i(t)$ generated in the n th interval $[(n-1)I, nI]$ ($n = 1, 2, \dots, N$) is upper bounded by $F_{p,j,k}^i(I, n)$. That is, $\sum_{y=(n-1)I}^{nI} y_{p,j,k}^i(t) \leq F_{p,j,k}^i(I, n)$.

Definition 2.3 provides a feature of arrival traffic $y_{p,j,k}^i(t)$ on an interval-by-interval basis. Theoretically, I can be any positively real number. In practice, however, I is selected as a finite positive integer.

Usually, $F_{p,j,k}^i(I, n) \neq F_{p,j,k}^i(I, q)$ for $n \neq q$. Therefore, $\{F_{p,j,k}^i(I, n)\}$ ($n = 1, 2, \dots$) is a random process. Computing the sample mean of $F_{p,j,k}^i(I, n)$ in terms of I yields

$$\frac{1}{I} \sum_{m=1}^I F_{p,j,k}^i(m, n) = \bar{F}_{p,j,k}^i(n). \quad (2.3)$$

Usually, $\bar{F}_{p,j,k}^i(n_1) \neq \bar{F}_{p,j,k}^i(n_2)$ for $n_1 \neq n_2$. In practice, if $I \geq 10$, $\bar{F}_{p,j,k}^i(n)$ quite accurately follows Gaussian distribution regardless of the distribution of $F_{p,j,k}^i(I, n)$ [34]. Denote $A = \text{Var}[\bar{F}_{p,j,k}^i(n)]$ and $B = E[\bar{F}_{p,j,k}^i(n)]$, where Var and E are operators of variance and mean, respectively. Then, one can use the sample distribution of $\bar{F}_{p,j,k}^i(n)$ as follows: $[B - \bar{F}_{p,j,k}^i(n)]/\sqrt{A} = z$, where z follows the standard Gaussian distribution. Thus,

$$\bar{F}_{p,j,k}^i(n) \sim \frac{1}{\sqrt{2\pi A}} e^{-[\bar{F}_{p,j,k}^i(n) - B]^2/2A}. \quad (2.4)$$

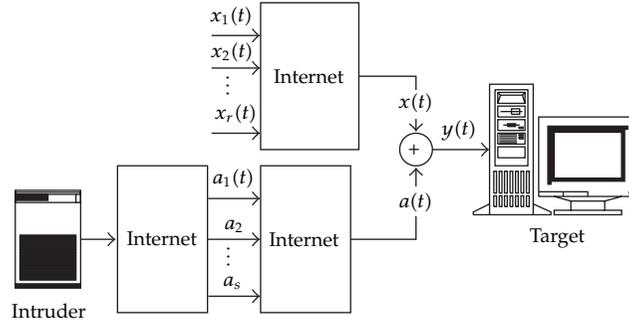


Figure 1: Illustration of DDoS flood attacks.

3. Principle

3.1. Detection Probability and Miss Probability

Normally, a server serves for a number of connections (clients) concurrently. Figure 1 illustrates a server that serves for r connections of normal traffic and s connections of attack traffic. Aggregated traffic $y(t)$ consists of normal traffic $x(t)$ and attack one $a(t)$.

In the case of $I \geq 10$, one has

$$\text{Prob} \left[z_{1-\alpha/2} < \frac{\bar{F}_{p,j,k}^i(n) - B}{\sqrt{A}} \leq z_{\alpha/2} \right] = 1 - \alpha, \quad (3.1)$$

where $(1 - \alpha)$ is called confidence coefficient. Let $C_{p,j,k}^i(\alpha)$ be the confidence interval with $(1 - \alpha)$ confidence coefficient. Then,

$$C_{p,j,k}^i(\alpha) = \left(B - \sqrt{A}z_{\alpha/2}, B + \sqrt{A}z_{\alpha/2} \right). \quad (3.2)$$

The above expression exhibits that B is a template of $\bar{F}_{p,j,k}^i(n)$. Thus, we have $(1 - \alpha)\%$ confidence to say that $\bar{F}_{p,j,k}^i(n)$ normally takes the value of B as its approximation with the variation less than or equal to $\sqrt{A}z_{\alpha/2}$.

Denote that $\xi(n) = \xi = \bar{F}_{p,j,k}^i(n)$. Then,

$$\text{Prob} \left(\xi > B + \sqrt{A}z_{\alpha/2} \right) = \frac{\alpha}{2}. \quad (3.3)$$

On the other hand,

$$\text{Prob} \left(\xi \leq B - \sqrt{A}z_{\alpha/2} \right) = \frac{\alpha}{2}. \quad (3.4)$$

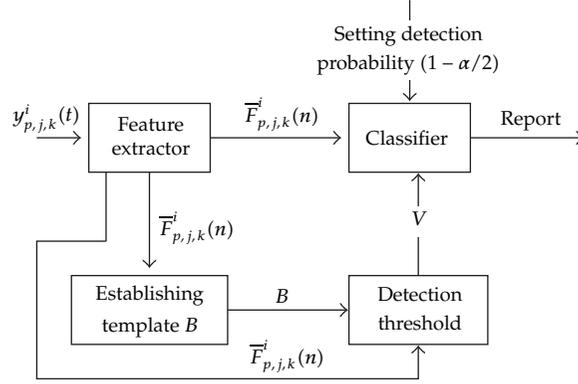


Figure 2: Diagram of detection model.

For facilitating the discussion, two terms are explained as follows. Correctly recognizing an abnormal sign means *detection* and failing to recognize it *miss*. We explain the detection probability and miss probability by the following theorem.

Theorem 3.1 (Detection probability). *Let*

$$V(\alpha) = V = B + \sqrt{A}z_{\alpha/2} \quad (3.5)$$

be the detection threshold. Denote $P_{\det} = P\{V < \xi < \infty\}$ as detection probability. Denote P_{miss} as miss probability. Then,

$$P_{\det} = P\{V < \xi < \infty\} = \left(1 - \frac{\alpha}{2}\right), \quad (3.6)$$

$$P_{\text{miss}} = \frac{\alpha}{2}. \quad (3.7)$$

Proof. The probability of $\xi \in C_{p,j,k}^i(\alpha)$ is $(1 - \alpha)$. Accordingly, the probability of $\xi \leq V$ is $(1 - \alpha/2)$. Therefore, the detection probability for $\xi > V$ is $(1 - \alpha/2)$. Hence, (3.6) holds. Since $P_{\det} + P_{\text{miss}} = 1$ [8], $P_{\text{miss}} = \alpha/2$. \square

In the case of $P_{\det} = 1$ and the computation precision being 4, one has

$$V = B + 4\sqrt{A}. \quad (3.8)$$

The diagram of our detection is indicated in Figure 2.

3.2. About False Alarm

False alarm means mistakenly recognizing a normal as abnormal. In this mechanism, detection criterion is $\bar{F}_{p,j,k}^i(n) > V(\alpha)$ with $P_{\det} = (1 - \alpha/2)$ and $P_{\text{miss}} = \alpha/2$. Therefore, if

$\bar{F}_{p,j,k}^i(n) > V(\alpha)$ happens in the case that $\bar{F}_{p,j,k}^i(n)$ comes from normal traffic and an alert is fired, then this alert will be a false alarm, which has the probability $\alpha/2$. Therefore,

$$P_{\text{false}} = P_{\text{miss}}. \quad (3.9)$$

In the case of $P_{\text{det}} = 1$, one has $P_{\text{false}} = P_{\text{miss}} = 0$.

3.3. Partly Distinguishing Attack Traffic

For the simplicity, suppose that traffic has two priorities p_1 and p_2 . We further suppose that flash-crowd traffic has the priority p_1 but does not have p_2 . Non-flash-crowd normal traffic has both p_1 and p_2 and DDOS flood traffic has both p_1 and p_2 . Then, $\bar{F}_{p_2,j,k}^i(n) > V(\alpha)$ implies a detection that the traffic $y_{p_2,j,k}^i(t)$ contains attack traffic of class i at the server k from the link j in the n th interval. The detection probability is $(1 - \alpha/2)$.

Denote $y_{p_2,j,k}^i(t) = x y_{p_2,j,k}^i(t) + a y_{p_2,j,k}^i(t)$, where $x y_{p_2,j,k}^i(t)$ and $a y_{p_2,j,k}^i(t)$ are normal traffic and attack traffic with p_2 , respectively. Note that $x y_{p_2,j,k}^i(t)$ does not have the components of flash-crowd traffic.

Usually, a signature-based sensor is designed such that it has a library that contains signatures of attack traffic. In the present mechanism, however, we use a signature-based sensor that has a library to contain signatures of legitimate traffic with the priorities that flash-crowd traffic does not have. In this way, traffic whose signatures cannot be matched by this signature-based sensor may be taken as flood traffic or suspicious. Thus, if $\bar{F}_{p_2,j,k}^i(n) > V(\alpha)$ occurs, the flows that are in $y_{p_2,j,k}^i(t)$ and cannot be matched by the signature-based sensor are flood traffic of class i with p_2 at the server k from the link j in the n th interval. The reason to use a signature library of legitimate traffic instead of attack one is that attackers make efforts to create new variants of signatures but legitimate users usually do not. Figure 3 indicates the process of distinguishing attack traffic $a y_{p_2,j,k}^i(t)$ from $y_{p_2,j,k}^i(t)$.

4. A Case Study

We consider fractional Gaussian noise (FGN), which is an approximation model of traffic time series [18, 19, 21, 22, 35, 36]. The autocorrelation function of discrete FGN is given by

$$R(l) = 0.5\sigma^2 \left[|l| + 1|^{2H} - 2|l|^{2H} + |l| - 1|^{2H} \right], \quad (4.1)$$

where $\sigma^2 = (\Gamma(2 - H) \cos(\pi H)) / \pi H(2H - 1)$ is the strength of FGN [37], l is an integer, $\Gamma(\cdot)$ is the Gamma function, and $H \in (0.5, 1)$ the Hurst parameter.

In Figures 4, 5, 6, and 7, subscripts and superscripts of y and F are omitted. Consider TCP traffic series $y(t)$ ($40 \leq y \leq 1500$ (Bytes)), indicating the number of bytes in a packet at t . By simulating FGN, we have a series with $H = 0.6$ as shown in Figure 4. According to Definition 2.2, we obtain $F(I, n)$ (Bytes) as shown in Figure 5 ($n, I = 1, 2, \dots, 16$). Figure 6 indicates $\xi(n)$ (Bytes). The histogram of ξ is given in Figure 7.

From Figure 7, we attain $\mu_\xi = 3,105$ and $\sigma_\xi = 344.402$. Under the condition of $P_{\text{det}} = 1$, one has the interval [1720, 4467] and the threshold $V = 4,467$.

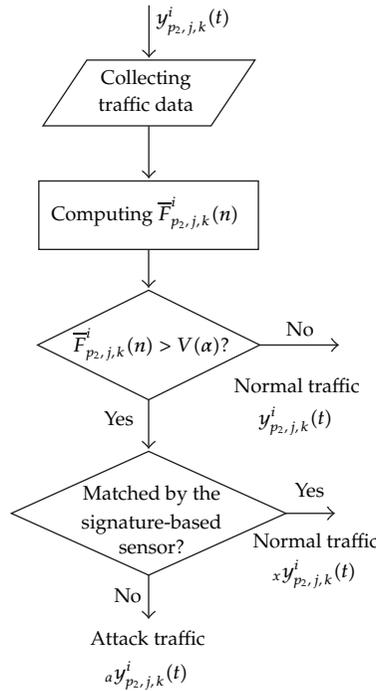


Figure 3: Distinguishing attack traffic.

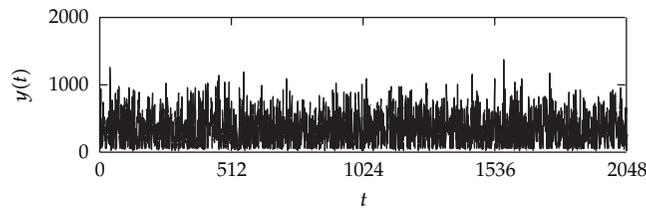


Figure 4: Synthesized FGN series.

5. Discussions

5.1. DiffServ Architecture: A Flexible Foundation

The above explanations only take the simple case of two priorities. In fact, there may be several priorities in a DiffServ domain, where applications are differentiated by their classes, and a certain portion of bandwidth is reserved for each class traffic [38]. Usually, all the flows in a class are assigned the same priority on each router. However, it is also available that the flows in a class may be assigned different priorities, and flows from different classes may have the same priority as can be seen from [32, Paragraph 5, Section 1, page 327]. This paper considers a class to be assigned different priorities. On the other side, the DiffServ architecture distinguishes two types of routers (edge routers and core routers) [32, Paragraph 2, Section 3, page 327]. Thus, a detector can be installed with either edge routers or core ones. Consequently, the DiffServ architecture provides a flexible foundation to design effective IDS to distinguish flood traffic from aggregated one. This paper is simply a beginning on this track.

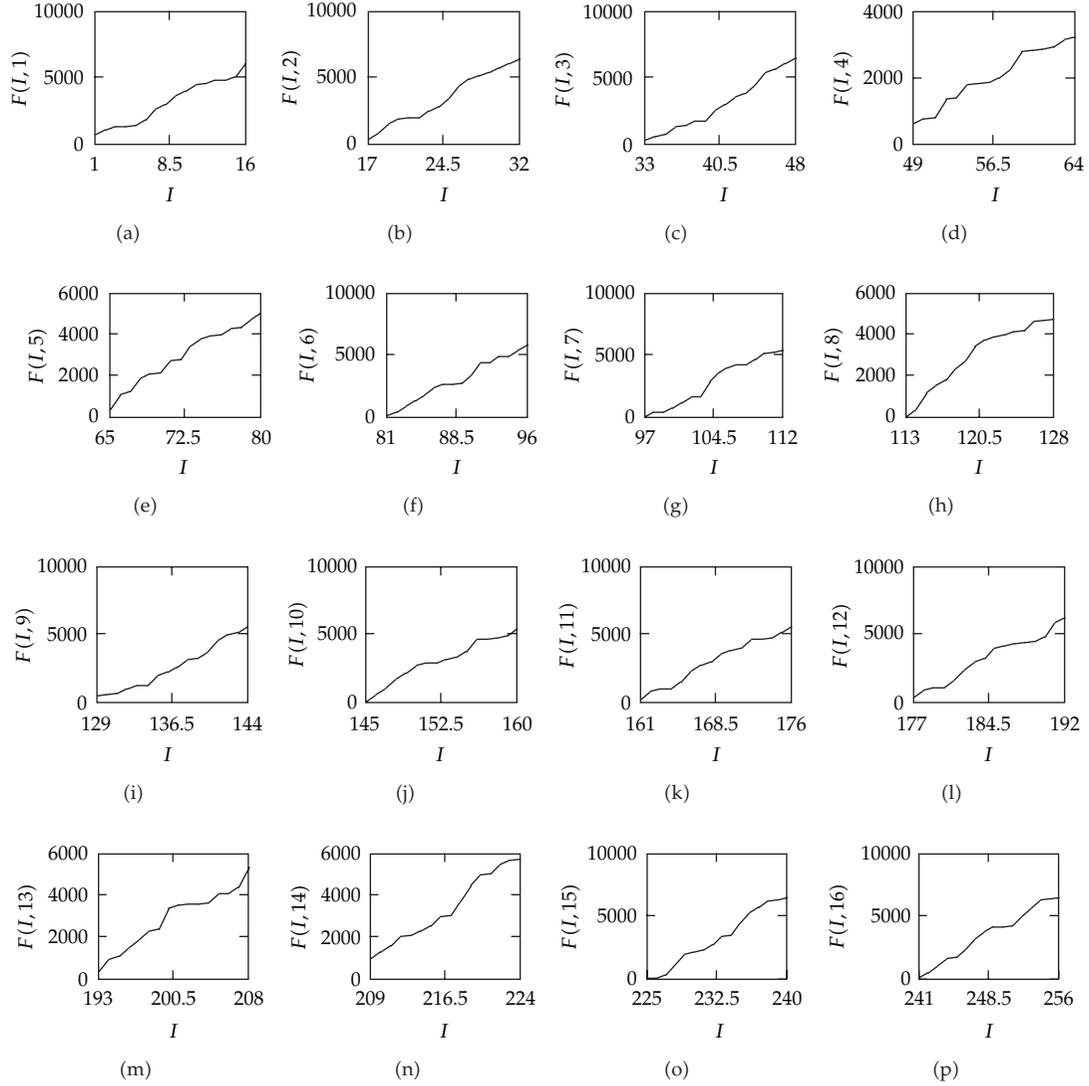


Figure 5: Illustrations of traffic regulators in different intervals.

5.2. Applicability

Mathematical properties of traditionally aggregated traffic time series have been studied deeply in a way, see for example, [18–22, 35]. However, math properties of aggregated traffic time series on a class-by-class basis for different priorities in the DiffServ domain are rarely seen. That is a main reason we use traffic regulator proposed by [33] because it is a tool particularly applicable in a flow-unaware environment. In addition to that, the traffic regulator is simple. Let T_m and T_c be the time for recording data and data processing, respectively. Suppose that we record a packet per 10 microsecond. Then, $T_m = 10^{-5}Q$ (second), where Q is the length of the series involved in computations. In the above case study, $Q = 16 \times 16 = 256$. Thus, $T_m = 2.56$ ms. On the other hand, T_c for a series of 256 length

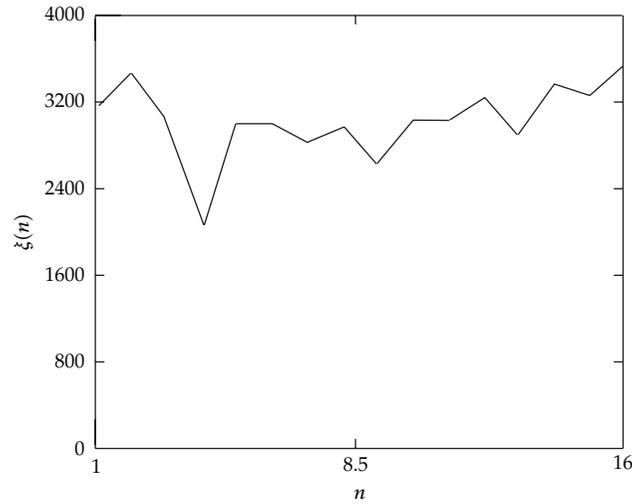


Figure 6: The sample mean of $\xi(n)$.

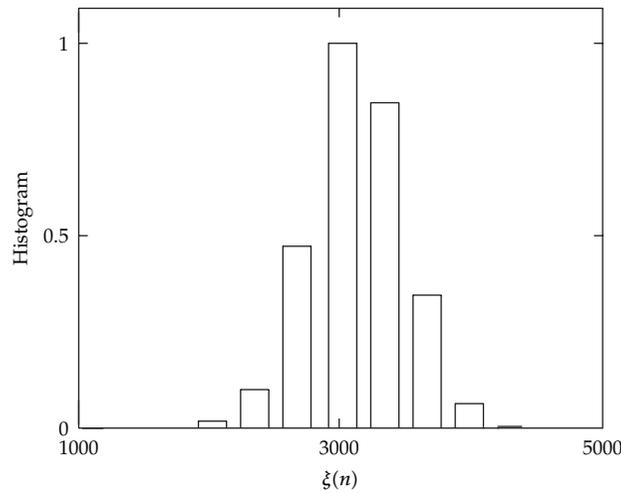


Figure 7: Histogram of ξ .

on an average Pentium IV PC is neglectable in comparison with T_m . This exhibits that the detection time is short enough to meet real-time use in practice.

It is worth noting that $\bar{F}_{p,j,k}^i(n)$ is a traffic pattern. In the present method, signs of DDOS flood attacks are identified by $\bar{F}_{p,j,k}^i(n) > V$, meaning traffic pattern under attacking must be significantly different from that of normal traffic. As a matter of fact, if an attacker were able to attack a target such that it would be overwhelmed by creating the floods that well mimic or be near to normal traffic, the target would be overwhelmed at its normal state even if there were no flood packets. This is obviously impossible even if the attacker knows normal traffic pattern exactly before attacking.

5.3. Future Work

The previous presentation is quite academic in the following senses. The detection mechanism previously exhibited was discussed based on postulated traffic models without analyzing real-traffic data. For this reason, we shall work on the traffic models in this paper with real-traffic data for anomaly detections. In addition, we will derive a general mechanism to reliably identify and distinguish attack traffic from aggregated traffic for the flows of class i with all priorities. In addition to that, we shall explore statistical learning methods discussed in other fields, see for example, [39–49].

6. Conclusions

This paper suggests a reliable method to detect signs of DDOS flood attacks in the DiffServ environment with static-priority schedulers. The present method can, with the combination of a signature-based sensor, partly but reliably distinguish attack traffic from aggregated traffic at a given server for a given link in a given time interval according to a predetermined detection probability. Given that static-priority schedulers are widely supported in current routers, it is our belief that this approach may be practical and effective in engineering.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (NSFC) under the project Grant nos. 60873264, 61070214, and the China national 973 plan under the project number 2011CB302801/2011CB302802.

References

- [1] G. Coulouris, J. Dollimore, and T. Kindberg, *Distributed Systems: Concepts and Design*, Addison-Wesley, 3rd edition, 2001.
- [2] K. Liston, "Intrusion Detection FAQ: Can You Explain Traffic Analysis and Anomaly Detection?" July 2004, http://www.sans.org/security-resources/idfaq/anomaly_detection.php.
- [3] E. Schultz, "Intrusion prevention," *Computers & Security*, vol. 23, no. 4, pp. 265–266, 2004.
- [4] W. W. Streilein, D. J. Fried, and R. K. Cunningham, "Detecting flood-based denial-of-service attacks with SNMP/RMON," in *Proceedings of the Workshop on Statistical and Machine Learning Techniques in Computer Intrusion Detection*, George Mason University, September 2003.
- [5] J. Leach, "TBSE—an engineering approach to the design of accurate and reliable security systems," *Computers & Security*, vol. 23, no. 1, pp. 22–28, 2004.
- [6] S. H. Oh and W. S. Lee, "An anomaly intrusion detection method by clustering normal user behavior," *Computers & Security*, vol. 22, no. 7, pp. 596–612, 2003.
- [7] F. Gong, *Deciphering Detection Techniques: Part III Denial of Service Detection, White Paper*, McAfee Network Security Technologies Group, 2003.
- [8] M. Li, "An approach to reliably identifying signs of DDOS flood attacks based on LRD traffic pattern recognition," *Computers & Security*, vol. 23, no. 7, pp. 549–558, 2004.
- [9] S. Sorensen, *Competitive Overview of Statistical Anomaly Detection, White Paper*, Juniper Networks, 2004, <http://www.juniper.net>.
- [10] R. A. Kemmerer and G. Vigna, "Intrusion detection: a brief history and overview, supplement to computer," *IEEE Security & Privacy*, vol. 35, no. 4, pp. 27–30, 2002.
- [11] S. B. Cho and H. J. Park, "Efficient anomaly detection by modeling privilege flows using hidden markov model," *Computers & Security*, vol. 22, no. 1, pp. 45–55, 2003.
- [12] S. Cho and S. Cha, "SAD: web session anomaly detection based on parameter estimation," *Computers & Security*, vol. 23, no. 4, pp. 312–319, 2004.
- [13] S. S. Kim, A. L. N. Reddy, and M. Vannucci, "Detecting traffic anomalies at the source though aggregate analysis of packet header data," in *Proceedings of the Networking*, vol. 3042 of *Lecture Notes in Computer Science*, pp. 1047–1059, Springer, Athens, Greece, May 2004.

- [14] B. Bencsath and I. Vajda, "Protection against DDoS attacks based on traffic level measurements," in *Proceedings of the International Symposium on Collaborative Technologies and Systems*, W. W. Smari, Ed., William McQuay, 2004.
- [15] L. Feinstein, D. Schnackenberg, R. Balupari, and D. Kindred, "Statistical approaches to DDoS attack detection and response," in *Proceedings of the DARPA Information Survivability Conference and Exposition*, vol. 1, pp. 303–314, Washington, DC, USA, April 2003.
- [16] R. Mahajan, S. Bellovin, S. Floyd, J. Ioannidis, V. Paxson, and S. Shenker, "Controlling high bandwidth aggregates in the network," *Computer Communication Review*, vol. 32, no. 3, pp. 62–73, July 2002.
- [17] J. B. D. Cabrera, B. Ravichandran, and R. K. Mehra, "Statistical modeling for network intrusion detection," in *Proceedings of the 8th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, San Francisco, Calif, USA, August-september 2000.
- [18] H. Michiel and K. Laevens, "Teletraffic engineering in a broad-band era," *Proceedings of the IEEE*, vol. 85, no. 12, pp. 2007–2033, 1997.
- [19] 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.
- [20] I. W. C. Lee and A. O. Fapojuwo, "Stochastic processes for computer network traffic modeling," *Computer Communications*, vol. 29, no. 1, pp. 1–23, 2005.
- [21] J. Beran, *Statistics for Long-Memory Processes*, Chapman and Hall, New York, NY, USA, 1994.
- [22] M. Garetto and D. Towsley, "An efficient technique to analyze the impact of bursty TCP traffic in wide-area networks," *Performance Evaluation*, vol. 65, no. 2, pp. 181–202, 2008.
- [23] M. Li, "Change trend of averaged hurst parameter of traffic under DDOS flood attacks," *Computers & Security*, vol. 25, no. 3, pp. 213–220, 2006.
- [24] H. G. Sun, Y. Q. Chen, and W. Chen, "Random-order fractional differential equation models," *Signal Processing*, vol. 91, no. 3, pp. 525–530, 2011.
- [25] 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.
- [26] 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.
- [27] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.
- [28] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.
- [29] E. G. Bakhoun 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, 2010.
- [30] 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.
- [31] C. S. Chang, "On deterministic traffic regulation and service guarantees: a systematic approach by filtering," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1097–1110, 1998.
- [32] S. Q. Wang, D. Xuan, R. Bettati, and W. Zhao, "Providing absolute differentiated services for real-time applications in static-priority scheduling networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 2, pp. 326–339, 2004.
- [33] R. Cruz, "A calculus for network delay. II. Network analysis," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 132–141, 1991.
- [34] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedure*, John Wiley & Sons, 2nd edition, 1991.
- [35] B. B. Mandelbrot, *Gaussian Self-Affinity and Fractals*, Springer, New York, NY, USA, 2002.
- [36] Y. Q. Chen, R. Sun, and A. Zhou, "An improved hurst parameter estimator based on fractional fourier transform," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 197–206, 2010.
- [37] H. Sheng, H. Sun, Y. Q. Chen, and T. Qiu, "Synthesis of multifractional gaussian noises based on variable-order fractional operators," *Signal Processing*, vol. 91, no. 7, pp. 1645–1650, 2011.
- [38] S. Black, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "Architecture for differentiated services," Tech. Rep. 2475, IETF, 1998.
- [39] J. Chen, C. Hu, and Z. Ji, "An improved ARED algorithm for congestion control of network transmission," *Mathematical Problems in Engineering*, vol. 2010, Article ID 329035, 2010.
- [40] M. Dong, "A tutorial on nonlinear time-series data mining in engineering asset health and reliability prediction: concepts, models, and algorithms," *Mathematical Problems in Engineering*, vol. 2010, Article ID 175936, 2010.

- [41] M. Dong, "A novel approach to equipment health management based on auto-regressive hidden semi-Markov model (AR-HSMM)," *Science in China*, vol. 51, no. 9, pp. 1291–1304, 2008.
- [42] 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.
- [43] 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.
- [44] W. Mikhael and T. Yang, "A gradient-based optimum block adaptation ICA technique for interference suppression in highly dynamic communication channels," *Eurasip Journal on Applied Signal Processing*, vol. 2006, Article ID 84057, 2006.
- [45] S. Y. Chen, Y. F. Li, and J. W. Zhang, "Vision processing for realtime 3D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
- [46] D. She and X. Yang, "A new adaptive local linear prediction method and its application in hydrological time series," *Mathematical Problems in Engineering*, vol. 2010, Article ID 205438, 2010.
- [47] J. Chen, C. Hu, and Z. Ji, "Self-tuning random early detection algorithm to improve performance of network transmission," *Mathematical Problems in Engineering*, vol. 2011, Article ID 872347, 17 pages, 2011.
- [48] H. Dong, Z. Wang, D. W. C. Ho, and H. Gao, "Variance-constrained H_∞ filtering for a class of nonlinear time-varying systems with multiple missing measurements: the finite-horizon case," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2534–2543, 2010.
- [49] B. Shen, Z. Wang, and X. Liu, "Bounded H_∞ synchronization and state estimation for discrete time-varying stochastic complex networks over a finite horizon," *IEEE Transactions on Neural Networks*, vol. 22, no. 1, pp. 145–157, 2010.

Research Article

Haar-Wavelet-Based Just Noticeable Distortion Model for Transparent Watermark

**Lu-Ting Ko,¹ Jwu-E Chen,¹ Hsi-Chin Hsin,²
Yaw-Shih Shieh,³ and Tze-Yun Sung³**

¹ Department of Electrical Engineering, National Central University, Chungli City 320-01, Taiwan

² Department of Computer Science and Information Engineering, National United University,
Miaoli 360-03, Taiwan

³ 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 3 June 2011; Accepted 5 July 2011

Academic Editor: Carlo Cattani

Copyright © 2012 Lu-Ting Ko et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Watermark transparency is required mainly for copyright protection. Based on the characteristics of human visual system, the just noticeable distortion (JND) can be used to verify the transparency requirement. More specifically, any watermarks whose intensities are less than the JND values of an image can be added without degrading the visual quality. It takes extensive experimentations for an appropriate JND model. Motivated by the texture masking effect and the spatial masking effect, which are key factors of JND, Chou and Li (1995) proposed the well-known full-band JND model for the transparent watermark applications. In this paper, we propose a novel JND model based on discrete wavelet transform. Experimental results show that the performance of the proposed JND model is comparable to that of the full-band JND model. However, it has the advantage of saving a lot of computation time; the speed is about 6 times faster than that of the full-band JND model.

1. Introduction

Watermarking is a process that hides information into a host image for the purpose of copyright protection, integrity checking, or captioning [1–3]. In order to achieve the transparency of watermark, many commonly used techniques are based on the characteristics of human visual system (HVS) [1–13]. Jayant et al. [14, 15] introduced a key concept known as the just noticeable distortion (JND), against which insignificant errors are not perceptible by human eyes. The JND of an image is in general dependent on background luminance, contrast of luminance, and dominant spatial frequency. It takes extensive experimentations to obtain an appropriate JND model.

Perceptual redundancies refer to the details of an image that are not perceivable by human eyes and therefore can be discarded without affecting the visual quality. As noted, human visual perception is sensitive to the contrast of luminance rather than their individual values [16–18]. In addition, the visibility of stimuli can be reduced by nonuniformly quantizing the background luminance [18–20]. The above known as the texture masking effect and the spatial masking effect are key factors that affect the JND of an image. Chou and Li proposed an effective model called the full-band JND model for the transparent watermark applications [21].

Wavelet transform provides an efficient multiresolution representation with various desirable properties such as subband decompositions with orientation selectivity and joint space-spatial frequency localization. In wavelet domain, the higher detailed information of a signal is projected onto the shorter basis function with higher spatial resolution; the lower detailed information is projected onto the larger basis function with higher spectral resolution. This matches the characteristics of HVS. Many wavelet-transform-based algorithms were proposed for various applications [22–34].

In this paper, we propose a wavelet-transform-based JND model for the watermark applications. It has the advantage of saving a lot of computation time. The remainder of the paper proceeds as follows. In Section 2, the full-band JND model is reviewed briefly. In Section 3, the discrete-wavelet-transform- (DWT-) based JND model is proposed. The modified DWT-based JND model and its evaluation are presented in Section 4. Conclusion can be found in Section 5.

2. Review of the Full-Band Just Noticeable Distortion (JND) Model

The full-band JND model [21] makes use of the properties of the HVS to measure the perceptual redundancies of an image. It produces the JND file for image pixels as follows:

$$\text{JND}(i, j) = \max\{f_1(bg(i, j), mg(i, j)), f_2(bg(i, j))\}, \quad (2.1)$$

where

$$f_1(bg(i, j), mg(i, j)) = mg(i, j) \times [\alpha(bg(i, j))] + \beta(bg(i, j)), \quad (2.2)$$

$$f_2(bg(i, j)) = \begin{cases} 17 \times \left(1 - \left(\frac{bg(i, j)}{127}\right)^{1/2}\right) + 3 & \text{for } bg(i, j) \leq 127, \\ \frac{3}{128} \times (bg(i, j) - 127) + 3 & \text{for } bg(i, j) > 127, \end{cases} \quad (2.3)$$

$$\alpha(bg(i, j)) = bg(i, j) \times 0.0001 + 0.115, \quad (2.4)$$

$$\beta(bg(i, j)) = \frac{1}{2} - bg(i, j) \times 0.01, \quad (2.5)$$

$$bg(i, j) = \frac{1}{32} \sum_{w=1}^5 \sum_{z=1}^5 p(i-3+w, j-3+z) \times B(w, z), \quad (2.6)$$

$$mg(i, j) = \max_{k=1,2,3,4} \{|\text{grad}_k(i, j)|\}, \quad (2.7)$$

$$\text{grad}_k(i, j) = \frac{1}{16} \sum_{w=1}^5 \sum_{z=1}^5 p(i-3+w, j-3+z) \times G_k(w, z), \quad (2.8)$$

where $f_1(i, j)$ and $f_2(i, j)$ are the texture mask and the spatial mask, respectively, as mentioned in Section 1, $bg(i, j)$ is the average background luminance obtained by using a low-pass filter, $B(w, z)$, is given in Figure 1, $mg(i, j)$ is the maximum gradient obtained by using a set of high-pass filters, $G_k(w, z)$, $k = 1, 2, 3, 4$, are given in Figure 2, functions $\alpha(\circ)$ and $\beta(\circ)$ are dependent on the average background luminance, and $p(i, j)$ is the luminance value at pixel position (i, j) .

3. Discrete-Wavelet-Transform-Based JND Model

In this section, we propose a novel JND model based on discrete wavelet transform. It has the advantage of reducing computational complexity significantly.

3.1. Discrete Wavelet Transform

Discrete wavelet transform (DWT) provides an efficient multiresolution analysis for signals, Specifically, any finite energy signal $f(x)$ can be written by

$$f(x) = \sum_n S_J(n) \phi_{Jn}(x) + \sum_{\ell \leq J} \sum_n D_\ell(n) \psi_{\ell n}(x), \quad (3.1)$$

where ℓ denotes the resolution index with larger values meaning coarser resolutions, n is the translation index, $\psi(x)$ is a mother wavelet, $\phi(x)$ is the corresponding scaling function, $\psi_{\ell n}(x) = 2^{-\ell/2} \psi(2^{-\ell}x - n)$, $\phi_{\ell n}(x) = 2^{-\ell/2} \phi(2^{-\ell}x - n)$, $S_J(n)$ is the scaling coefficient representing the approximation information of $f(x)$ at the coarsest resolution 2^J , and $D_\ell(n)$ is the wavelet coefficient representing the detail information of $f(x)$ at resolution 2^ℓ . Coefficients $S_\ell(n)$ and $D_\ell(n)$ can be obtained from the scaling coefficient $S_{\ell-1}(n)$ at the next finer resolution $2^{\ell-1}$ by using 1-level DWT, which is given by

$$\begin{aligned} S_\ell(n) &= \sum_k S_{\ell-1}(k) h(2n - k), \\ D_\ell(n) &= \sum_k S_{\ell-1}(k) g(2n - k), \end{aligned} \quad (3.2)$$

where $h(n) = \langle \phi, \phi_{-1, -n} \rangle$, $g(n) = \langle \psi, \phi_{-1, -n} \rangle$, and $\langle \cdot, \cdot \rangle$ denotes the inner product. It is noted that $h(n)$ and $g(n)$ are the corresponding low-pass filter and high-pass filter, respectively. Moreover, $S_{\ell-1}(n)$ can be reconstructed from $S_\ell(n)$ and $D_\ell(n)$ by using the inverse DWT, which is given by

$$S_{\ell-1}(n) = \sum_k S_\ell(k) \tilde{h}(n - 2k) + \sum_k D_\ell(k) \tilde{g}(n - 2k), \quad (3.3)$$

1	1	1	1	1
1	2	2	2	1
1	2	0	2	1
1	2	2	2	1
1	1	1	1	1

B

Figure 1: Low-pass filter used in (2.6).

where $\tilde{h}(n) = h(-n)$ and $\tilde{g}(n) = g(-n)$.

For image applications, 2D DWT can be obtained by using the tensor product of 1D DWT. Among wavelets, Haar wavelet [22] is the simplest one, which has been widely used for many applications. The low-pass filter and high-pass filter of Haar wavelet are as follows:

$$\begin{aligned} h(0) &= 0.5, & h(1) &= 0.5, \\ g(0) &= 0.5, & g(1) &= -0.5. \end{aligned} \tag{3.4}$$

Figures 3 and 4 show the row decomposition and the column decomposition using Haar wavelet, respectively. Notice that the column decomposition may follow the row decomposition, or vice versa, in 2D DWT:

$$\begin{aligned} LL &= \frac{A + B + C + D}{4}, \\ LH &= \frac{A + B - C - D}{4}, \\ HL &= \frac{A - B + C - D}{4}, \\ HH &= \frac{A - B - C + D}{4}, \end{aligned} \tag{3.5}$$

where $A, B, C,$ and D are pixel values, and $LL, LH, HL,$ and HH denote the approximation, detail information in the horizontal, vertical, and diagonal orientations, respectively, of the input image. Figure 5 shows 1-level, 2D DWT using Haar wavelet.

The LL subband of an image can be further decomposed into four subbands: $LLLL, LLLH, LLHL,$ and $LLHH$ at the next coarser resolution, which together with $LH, HL,$ and HH forms the 2-level DWT of the input image. Thus, higher level DWT can be obtained by decomposing the approximation subband in the recursive manner.

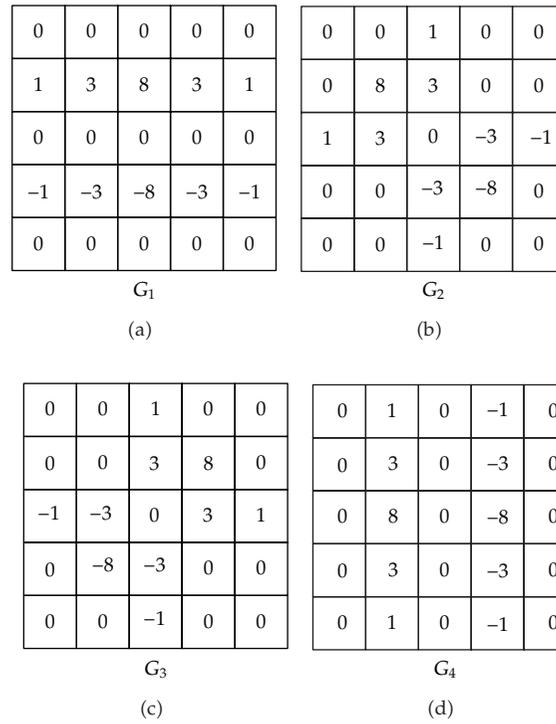


Figure 2: A set of high-pass filters used in (2.8).

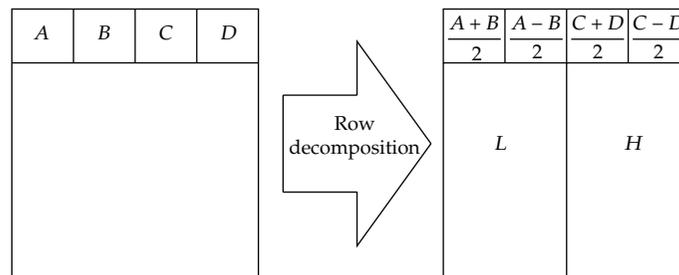


Figure 3: The row decomposition using Haar wavelet (A, B, C and D , are pixel values).

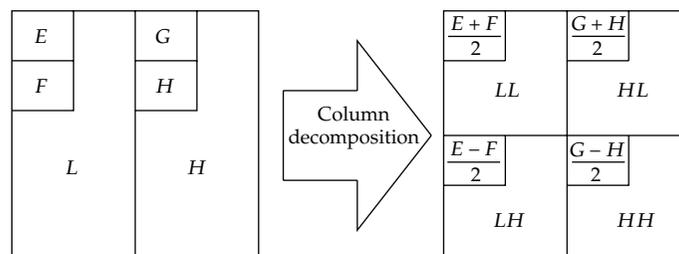


Figure 4: The column decomposition using Haar wavelet (E, F, G , and H are pixel values).

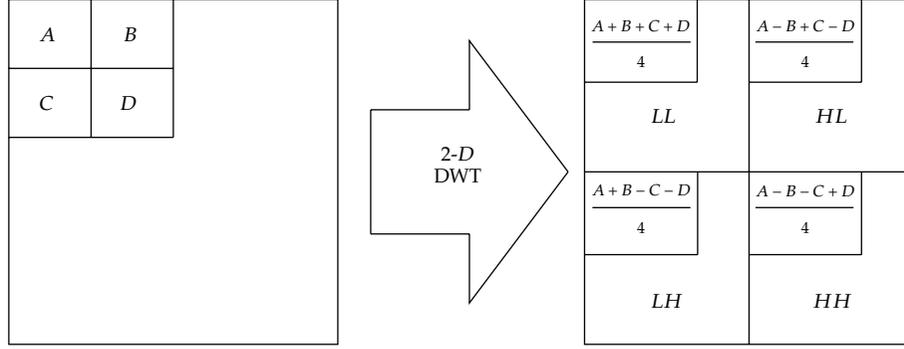


Figure 5: 1-level 2D DWT using Haar wavelet (A , B , C , and D are pixel values).

3.2. Proposed DWT-Based JND Model

As mentioned in Section 2, the full-band JND model [21] consists of (2.1)–(2.8), which is essentially computation consuming. In order to simplify computational complexity, a novel JND model based on Haar wavelet is proposed as follows.

$$\text{JND}_{\text{DWT}}(i, j) = \max\{f_{1,\text{DWT}}(i, j), f_{2,\text{DWT}}(i, j)\}, \quad (3.6)$$

where $f_{1,\text{DWT}}(i, j)$ and $f_{2,\text{DWT}}(i, j)$ are the proposed texture mask and spatial mask, respectively, based on DWT, which are given by

$$f_{1,\text{DWT}}(i, j) = |D_i(i, j) \cdot \alpha[\text{LLLL}'(i, j)] + \beta[\text{LLLL}'(i, j)]|, \quad (3.7)$$

$$f_{2,\text{DWT}}(i, j) = \begin{cases} 17 \times \left(1 - \left[\frac{\text{LLLL}'(i, j)}{127}\right]^{1/2}\right) + 3 & \text{for } \text{LLLL}'(i, j) \leq 127, \\ \frac{3}{128} \times [\text{LLLL}'(i, j) - 127] + 3 & \text{for } \text{LLLL}'(i, j) > 127, \end{cases} \quad (3.8)$$

where

$$\alpha[\text{LLLL}'(i, j)] = \text{LLLL}'(i, j) \times 0.0001 + 0.115, \quad (3.9)$$

$$\beta[\text{LLLL}'(i, j)] = \frac{1}{2} - \text{LLLL}'(i, j) \times 0.01, \quad (3.10)$$

$$\text{LLLL}' = \mathbf{T}_2 \mathbf{T}_1 [\text{LLLL}] \mathbf{T}_1^T \mathbf{T}_2^T, \quad (3.11)$$

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}_{4 \times 2}, \quad (3.12)$$

$$\mathbf{T}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{8 \times 4}. \quad (3.13)$$

$LLLL'$ is the modified $LLLL$, which replaces the low-pass filter $B(w, z)$ of the full-band JND model. $D_i(i, j)$ replaces the maximum gradient, $mg(i, j)$, which is given by

$$D_i(i, j) = \max\{|LH'(i, j)|, |HL'(i, j)|, |SL'_R(i, j)|, |SL'_L(i, j)|\}, \quad (3.14)$$

where

$$LH' = \mathbf{T}_2[LH]\mathbf{T}_2^T, \quad (3.15)$$

$$HL' = \mathbf{T}_2[HL]\mathbf{T}_2^T, \quad (3.16)$$

$$SL'_R = \mathbf{T}_2[SL_R]\mathbf{T}_2^T, \quad (3.17)$$

$$SL'_L = \mathbf{T}_2[SL_L]\mathbf{T}_2^T, \quad (3.18)$$

$$SL_R = \frac{LH + HL}{2}, \quad (3.19)$$

$$SL_L = \frac{LH - HL}{2}. \quad (3.20)$$

LH' , HL' , SL'_R , and SL'_L replace $G_1(w, z)$, $G_2(w, z)$, $G_3(w, z)$, and $G_4(w, z)$, respectively.

4. Modification of the DWT-Based JND Model

In this section, we introduce an adjustable parameter to modify the DWT-based JND model such that the computation time can be reduced significantly while the performance is comparable to that of the benchmark full-band JND model. The test images, namely, Lena, Cameraman, Baboon, Board, and Peppers are shown in the first row of Figure 12.

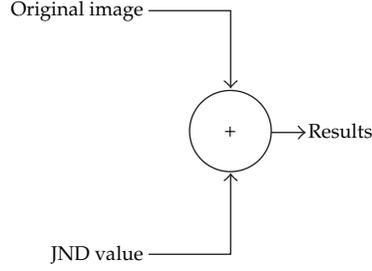


Figure 6: Distortion-tolerant evaluation model for the proposed JND model.

Table 1: PSNR comparisons of the benchmark full-band JND model, the proposed DWT-based JND model, and the modified DWT-based JND model.

JND model	Lena	Cameraman	Baboon	Board	Peppers
Full-band JND	32.7041	29.9122	34.0845	25.3486	30.3052
DWT-based JND	34.3301	31.7556	35.8156	30.8003	32.1375
The modified DWT-based JND	33.7117	30.9614	34.0931	25.7192	31.7148

4.1. Evaluation of JND Models

Figure 6 shows the distortion-tolerant model, which can be used to evaluate JND models. It takes the JND value as noise, adds to the original image, and computes the peak-signal-to-noise ratio (PSNR) defined as

$$\text{PSNR} = 20 \log \left(\frac{255}{\sqrt{\text{MSE}}} \right), \quad (4.1)$$

where MSE is the mean squared error, which is defined as

$$\text{MSE} = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [\text{JND}(i, j)]^2, \quad (4.2)$$

where image size is $m \times n$. As shown in Table 1, the proposed DWT-based JND model is somewhat different from the benchmark full-band JND model in terms of the PSNR values.

4.2. Modified DWT-Based JND Model

Based on (2.1)–(2.3) and (3.6)–(3.8), one can examine the influences of the dominant mask, texture mask, and spatial mask for the full-band JND model and the DWT-based JND model, respectively. Their respective MSE values are shown in Table 2. As one can see, the proposed texture mask for the DWT-based JND model is less significant than that of the full-band JND model. Thus, we propose an adjustable parameter, K , to modify (3.17) and (3.18) as follows:

$$\begin{aligned} SL''_R &= K \cdot SL'_R, \\ SL''_L &= K \cdot SL'_L, \end{aligned} \quad (4.3)$$

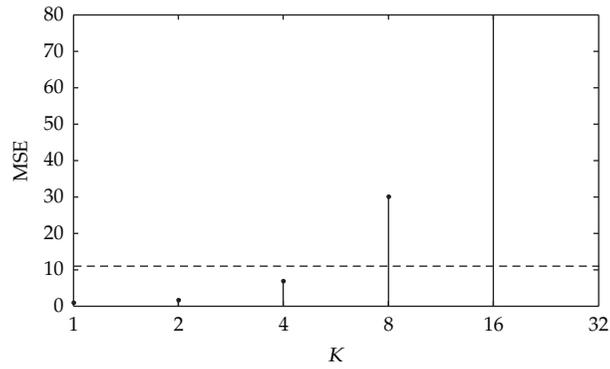


Figure 7: MSE values obtained by modifying the DWT-based texture mask using (4.3) with various K for Lena image; dashed line: MSE value obtained by using the texture mask of the full-band JND model.

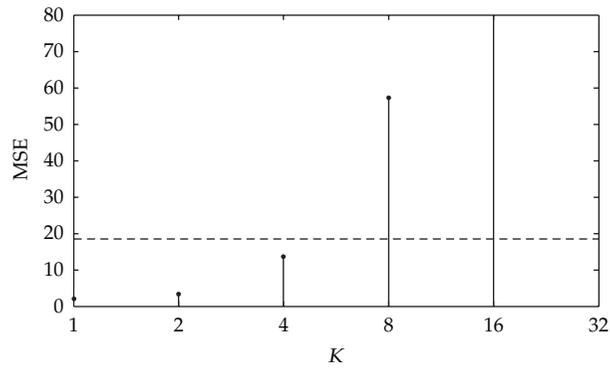


Figure 8: MSE values obtained by modifying the DWT-based texture mask using (4.3) with various K for Cameraman image; dashed line: MSE value obtained by using the texture mask of the full-band JND model.

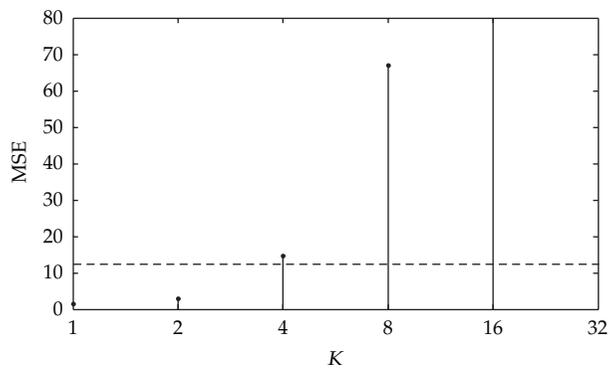


Figure 9: MSE values obtained by modifying the DWT-based texture mask using (4.3) with various K for Baboon image; dashed line: MSE value obtained by using the texture mask of the full-band JND model.

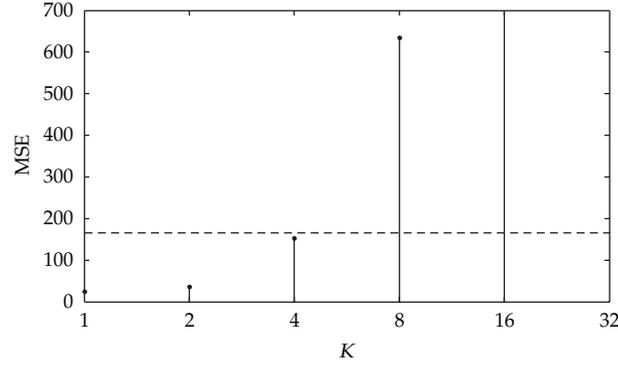


Figure 10: MSE values obtained by modifying the DWT-based texture mask using (4.3) with various K for Board image; dashed line: MSE value obtained by using the texture mask of the full-band JND model.

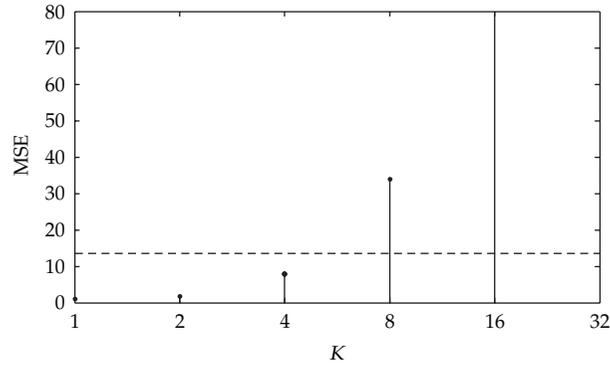


Figure 11: MSE values obtained by modifying the DWT-based texture mask using (4.3) with various K for Peppers image; dashed line: MSE value obtained by using the texture mask of the full-band JND model.

where SL''_R and SL''_L replace SL'_R and SL'_L , respectively. Figures 7, 8, 9, 10, and 11 show the MSE values obtained by modifying the DWT-based texture mask with various K in (4.3). In this paper, the adjustable parameter K is set to 4 after extensive simulations. The performance of the modified DWT-based JND model with $K = 4$ is comparable to that of the full-band JND model in terms of the PSNR and MSE values as shown in Tables 1 and 2, respectively.

Figure 12 shows the noisy images obtained by adding the JND values to the original images (Figure 12(a)) using the full-band JND model (Figure 12(b)), the DWT-based JND model (Figure 12(c)), and the modified DWT-based JND model (Figure 12(d)). It is noted that the images in the second and fourth rows are almost indistinguishable from the original images. As a result, the modified DWT-based JND model is visually comparable to the full-band JND model.

4.3. Computation Complexity of the Proposed JND Models

In the full-band JND model, the computation of $bg(i, j)$ requires 9 multiplications per pixel, the computation of $mg(i, j)$ requires 28 multiplications per pixel, and the computation required for (2.2)–(2.5) is 6 multiplications per pixel. Thus, for an $n \times n$ image, it requires



Figure 12: (a) The original images, namely, Lena, Cameraman, Baboon, Board, and Peppers; (b), (c), and (d) the noisy images obtained by adding the full-band JND values, the DWT-based JND values, and the modified DWT-based JND values, respectively.

Table 2: The MSE values due to the dominant mask (case 1), the spatial mask (case 2), and the texture mask (case 3) using the full-band JND model, the DWT-based JND model, and the modified DWT-based JND model.

JND model	Case	Lena	Cameraman	Baboon	Board	Peppers
Full-band JND	1	34.8876	66.3534	25.3878	189.7653	60.6121
	2	28.5202	53.0297	19.6219	41.7222	52.5825
	3	10.9574	18.4772	12.373	165.6492	12.583
DWT-based JND	1	23.9924	43.4027	17.0419	54.0822	39.7492
	2	23.9269	43.0406	16.942	38.4702	39.642
	3	0.972	2.1076	1.4603	24.4885	1.1393
The modified DWT-based JND	1	27.6254	52.1122	25.3378	174.244	43.8131
	2	23.9269	43.0406	16.942	38.4702	39.642
	3	6.9379	13.7233	14.7008	153.4052	7.8595

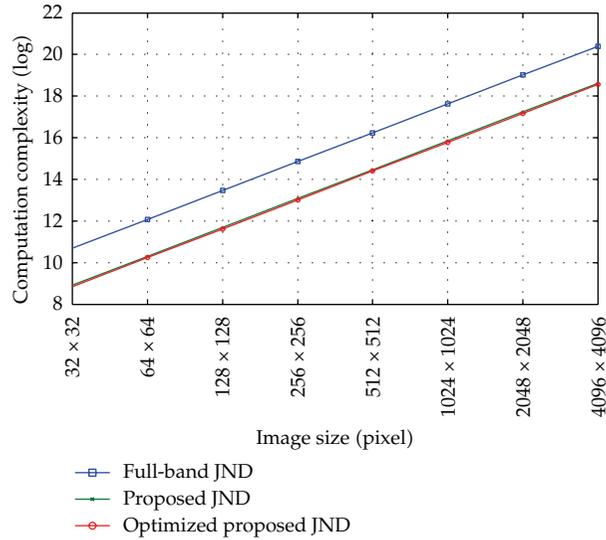


Figure 13: Log plot of numbers of multiplications required for the three JND models versus different image sizes.

$43n^2$ multiplications. In the proposed DWT-based JND model, the computations of LH' , HL' , LLL' , SL'_R , and SL'_L require $(1/4)n^2$, $(1/4)n^2$, $(5/16)n^2$, $(1/4)n^2$, and $(1/4)n^2$ multiplications, respectively, for an $n \times n$ image, and the computation required for (3.7)–(3.10) is also 6 multiplications per pixel. Thus, for an $n \times n$ image, it requires $7.3125n^2$ multiplications. In the modified DWT-based JND model, as the computations of SL''_R and SL''_L require no multiplication, it only needs $6.8125n^2$ multiplications for an $n \times n$ image. Figure 13 shows the log plot of numbers of multiplications required for the three JND models versus different image sizes. As a result, the speed of the DWT-based JND model is about 6 times faster than that of the full-band JND model, which is the main advantage.

5. Conclusion

In this paper, an efficient DWT-based JND model is presented. It has the advantage of saving a lot of computation time while the performance is comparable to the benchmark full-band JND model. More specifically, the computation complexity of the proposed DWT-based JND model is only one sixth of that of the full-band JND model. As a result, it is suitable for the real-time applications.

Acknowledgment

The National Science Council of Taiwan, under Grants NSC98-2221-E-216-037 and NSC99-2221-E-239-034, supported this work.

References

- [1] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoan, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

- [2] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1064–1087, 1998.
- [3] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 783–791, 2001.
- [4] M. Kutter and S. Winkler, "A vision-based masking model for spread-spectrum image watermarking," *IEEE Transactions on Image Processing*, vol. 11, no. 1, pp. 16–25, 2002.
- [5] C. De Vleeschouwer, J. F. Delaigle, and B. Macq, "Invisibility and application functionalities in perceptual watermarking - An overview," *Proceedings of the IEEE*, vol. 90, no. 1, pp. 64–77, 2002.
- [6] Q. Li, C. Yuan, and Y. Z. Zhong, "Adaptive DWT-SVD domain image watermarking using human visual model," in *Proceedings of the 9th International Conference on Advanced Communication Technology (ICACT '07)*, pp. 1947–1951, February 2007.
- [7] F. Atrousseau and P. L. Callet, "A robust image watermarking technique based on quantization noise visibility thresholds," *Signal Processing*, vol. 87, no. 6, pp. 1363–1383, 2007.
- [8] H. S. Moon, T. You, M. H. Sohn, H. S. Kim, and D. S. Jang, "Expert system for low frequency adaptive image watermarking: using psychological experiments on human image perception," *Expert Systems with Applications*, vol. 32, no. 2, pp. 674–686, 2007.
- [9] H. Qi, D. Zheng, and J. Zhao, "Human visual system based adaptive digital image watermarking," *Signal Processing*, vol. 88, no. 1, pp. 174–188, 2008.
- [10] A. Koz and A. A. Alatan, "Oblivious spatio-temporal watermarking of digital video by exploiting the human visual system," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 3, pp. 326–337, 2008.
- [11] S. Y. Chen, Y. F. Li, and J. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
- [12] 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.
- [13] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 480–487, 2011.
- [14] N. Jayant, "Signal compression: technology targets and research directions," *IEEE Journal on Selected Areas in Communications*, vol. 10, pp. 314–323, 1992.
- [15] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385–1422, 1993.
- [16] R. F. Boyer and R. S. Spencer, "Thermal expansion and second-order transition effects in high polymers: part II. Theory," *Journal of Applied Physics*, vol. 16, no. 10, pp. 594–607, 1945.
- [17] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.
- [18] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 6, pp. 742–751, 2005.
- [19] J. Pandel, "Variable bit-rate image sequence coding with adaptive quantization," *Signal Processing*, vol. 3, no. 2-3, pp. 123–128, 1991.
- [20] B. Girod, "Psychovisual aspects of image communication," *Signal Processing*, vol. 28, no. 3, pp. 239–251, 1992.
- [21] C. H. Chou and Y. C. Li, "Perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 6, pp. 467–476, 1995.
- [22] B. S. Kim, I. J. Shim, M. T. Lim, and Y. J. Kim, "Combined preorder and postorder traversal algorithm for the analysis of singular systems by Haar wavelets," *Mathematical Problems in Engineering*, vol. 2008, Article ID 323080, 16 pages, 2008.
- [23] G. Mattioli, M. Scalia, and C. Cattani, "Analysis of large-amplitude pulses in short time intervals: application to neuron interactions," *Mathematical Problems in Engineering*, vol. 2010, Article ID 895785, 15 pages, 2010.
- [24] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.
- [25] C. Cattani, "Shannon wavelets theory," *Mathematical Problems in Engineering*, vol. 2008, Article ID 164808, 24 pages, 2008.

- [26] A. Kudreyko and C. Cattani, "Application of periodized harmonic wavelets towards solution of eigenvalue problems for integral equations," *Mathematical Problems in Engineering*, vol. 2010, Article ID 570136, 8 pages, 2010.
- [27] M. Li and W. Zhao, "Representation of a stochastic traffic bound," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 9, Article ID 5342414, pp. 1368–1372, 2010.
- [28] M. Li, "Generation of teletraffic of generalized Cauchy type," *Physica Scripta*, vol. 81, no. 2, Article ID 025007, 2010.
- [29] M. Li, "Fractal time series-a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [30] E. G. Bakhoun and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.
- [31] E. G. Bakhoun 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.
- [32] E. G. Bakhoun 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.
- [33] T. Y. Sung and H. C. Hsin, "A hybrid image coder based on SPIHT algorithm with embedded block coding," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E90-A, no. 12, pp. 2979–2984, 2007.
- [34] 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.

Research Article

A Unified Algorithm for Subband-Based Discrete Cosine Transform

Lu-Ting Ko,¹ Jwu-E Chen,¹ Hsi-Chin Hsin,²
Yaw-Shih Shieh,³ and Tze-Yun Sung³

¹ Department of Electrical Engineering, National Central University, Chungli City 320-01, Taiwan

² Department of Computer Science and Information Engineering, National United University, Miaoli City 360-03, Taiwan

³ 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 13 April 2011; Revised 22 May 2011; Accepted 4 June 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Lu-Ting Ko et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Discrete cosine transform (DCT) and inverse DCT (IDCT) have been widely used in many image processing systems and real-time computation of nonlinear time series. In this paper, the unified DCT/IDCT algorithm based on the subband decompositions of a signal is proposed. It is derived from the data flow of subband decompositions with factorized coefficient matrices in a recursive manner. The proposed algorithm only requires $(4^{(\log_2 n)-1} - 1)$ and $(4^{(\log_2 n)-1} - 1)/3$ multiplication time for n -point DCT and IDCT, with a single multiplier and a single processor, respectively. Moreover, the peak signal-to-noise ratio (PSNR) of the proposed algorithm outperforms the conventional DCT/IDCT. As a result, the subband-based approach to DCT/IDCT is preferable to the conventional approach in terms of computational complexity and system performance. The proposed reconfigurable architecture of linear array DCT/IDCT processor has been implemented by FPGA.

1. Introduction

The discrete cosine transform (DCT) first proposed by Ahemd et al. [1] is a Fourier-like transform. While the Fourier transform decomposes a signal into sine and cosine functions, DCT only makes use of cosine functions with the property of high energy compaction. As DCT is preferable for a trade-off between the optimal decorrelation known as the Karhunen-Loève transform and computational simplicity [2], it has been extensively used in many applications [3–14]. In particular, two-dimensional (2D) DCT, such as 8×8 DCT, has been adopted in some international standards such as JPEG, MPEG, and H.264 [15]. In MP3

audio codec, the subband analysis and synthesis filter banks requires the use of 32-point DCT/integer DCT to expedite computation [16]. Other audio compression standards, for example, the Dolby Digital AC-3 codec, utilize a modified DCT with 256 or 512 data points.

Many algorithms have been proposed for DCT/IDCT [17–21]. In which, the transportation matrix is factorized into products of simpler matrices. It is noted that, however, the factorized matrices are no longer as regular as the fast Fourier transform (FFT); thus, these algorithms can only achieve moderate computational speed. Specifically, the dedicated data paths deduced from the signal flow graphs (SFGs) of the above algorithms need to be optimized for performance enhancement, which is computationally intensive, and the custom-designed DCT is often complicated and cannot be easily scalable for variable data points.

In this paper, we propose a novel linear-array architecture based on the subband decomposition of a signal for scalable DCT/IDCT. The remainder of this paper proceeds as follows. First, the subband-based 8-point DCT/IDCT algorithm [22] is reviewed in Section 2. Its extension to n -point DCT/IDCT called the unified subband-based algorithm is proposed in Section 3. Section 4 presents the analysis of system complexity. The reconfigurable architecture of linear-array DCT/IDCT processor implemented by FPGA (field programmable gate array) is proposed in Section 5, and the conclusion can be found in Section 6.

2. The Subband-Based 8-Point DCT/IDCT Algorithm

The discrete cosine transform (DCT) of an 8-point signal, [8], is defined as

$$c[k] = \alpha[k] \sum_{n=0}^7 x[n] \cos \left[\frac{(2n+1)k\pi}{16} \right], \quad k = 0, 1, \dots, 7, \quad (2.1)$$

where $\alpha[0] = 1/(2\sqrt{2})$, and $\alpha[k] = 1/2$ for $k > 0$. It can be rewritten in the following matrix form:

$$\mathbf{c}_8 = \mathbf{D}_8 \cdot \mathbf{x}_8, \quad (2.2)$$

where $\mathbf{c}_8 = [c[0] \cdots c[7]]^T$, $\mathbf{x}_8 = [x[0] \cdots x[7]]^T$, and the transformation matrix \mathbf{D}_8 is as follows:

$$\mathbf{D}_8 = \frac{1}{\sqrt{8}} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ a & c & d & f & -f & -d & -c & -a \\ b & e & -e & -b & -b & -e & e & b \\ c & -f & -a & -d & d & a & f & -c \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ d & -a & f & c & -c & -f & a & -d \\ e & -b & b & -e & -e & b & -b & e \\ f & -d & c & -a & a & -c & d & -f \end{bmatrix}, \quad (2.3)$$

where $a = \sqrt{2} \cos(\pi/16)$, $b = \sqrt{2} \cos(2\pi/16)$, $c = \sqrt{2} \cos(3\pi/16)$, $d = \sqrt{2} \cos(5\pi/16)$, $e = \sqrt{2} \cos(6\pi/16)$, and $f = \sqrt{2} \cos(7\pi/16)$.

Let $x_L[n]$ and $x_H[n]$ denote the low-frequency and high-frequency subband signals of $x[n]$, respectively [22], which can be obtained by

$$\begin{aligned} x_L[n] &= \frac{1}{2} \{x[2n] + x[2n+1]\}, \\ x_H[n] &= \frac{1}{2} \{x[2n] - x[2n+1]\}, \end{aligned} \quad (2.4)$$

where $n = 0, 1, 2, 3$. As one can see, the DCT of $x[8]$ can be rewritten as

$$\begin{aligned} c[k] &= \sum_{n=0}^3 \alpha[k] x[2n] \cos\left(\frac{(4n+1)k\pi}{16}\right) + \sum_{n=0}^3 \alpha[k] x[2n+1] \cos\left(\frac{(4n+3)k\pi}{16}\right) \\ &= 2 \cos\left(\frac{\pi k}{16}\right) \underbrace{\sum_{n=0}^3 \alpha[k] x_L[n] \cos\left(\frac{(2n+1)k\pi}{8}\right)}_{c_L[k]} \\ &\quad + 2 \sin\left(\frac{\pi k}{16}\right) \underbrace{\sum_{n=0}^3 \alpha[k] x_H[n] \sin\left(\frac{(2n+1)k\pi}{8}\right)}_{s_H[k]}, \end{aligned} \quad (2.5)$$

where $c_L[k]$ and $s_H[k]$ are the subbands DCT and DST (discrete sine transform) of $x[n]$, respectively. Its vector form is as follows:

$$\begin{aligned} \mathbf{c}_8 &= [\mathbf{T}_{\text{SB,DCT},8} \quad \mathbf{T}_{\text{SB,DCT},8}]_{8 \times 8} \cdot \mathbf{M}_8 \cdot \mathbf{x}_8 \\ &= [\mathbf{T}_{\text{SB,DCT},8} \quad \mathbf{T}_{\text{SB,DCT},8}]_{8 \times 8} \cdot \begin{bmatrix} \mathbf{x}_L \\ \mathbf{x}_H \end{bmatrix}_{8 \times 1} \\ &= \underbrace{\mathbf{T}_{\text{SB,DCT},8} \cdot \mathbf{x}_L}_{\hat{\mathbf{c}}_{L,8}} + \underbrace{\mathbf{T}_{\text{SB,DST},8} \cdot \mathbf{x}_H}_{\hat{\mathbf{s}}_{H,8}}, \end{aligned} \quad (2.6)$$

where $\mathbf{T}_{\text{SB,DCT},8}$ and $\mathbf{T}_{\text{SB,DST},8}$ denote the 8×4 matrices of the subband DCT and DST, respectively, $\mathbf{x}_L = [x_L[0] \cdots x_L[3]]^T$, $\mathbf{x}_H = [x_H[0] \cdots x_H[3]]^T$ and $\mathbf{c}_8 = [c[0] \cdots c[7]]^T$.

According to (2.4), the 8-point matrix \mathbf{M}_8 can be written as

$$\mathbf{M}_8 = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & -0.5 \end{bmatrix}. \quad (2.7)$$

Due to the orthogonality between $\mathbf{T}_{\text{SB-DCT},8}$ and $\mathbf{T}_{\text{SB-DST},8}$, $x_L[n]$ and $x_H[n]$ can be obtained from $c[k]$ by

$$\begin{aligned} x_L[n] &= \sum_{k=0}^7 \alpha[k] \cos\left(\frac{\pi k}{16}\right) c[k] \cos\left(\frac{(2n+1)k\pi}{8}\right), \quad n = 0, \dots, 7, \\ x_H[n] &= \sum_{k=0}^7 \alpha[k] \sin\left(\frac{\pi k}{16}\right) c[k] \sin\left(\frac{(2n+1)k\pi}{8}\right), \quad n = 0, \dots, 7. \end{aligned} \quad (2.8)$$

In [22], the multistage subband decomposition is as follows. $\mathbf{c}_{L,4}$ and $\mathbf{c}_{H,4}$ denote the DCTs of \mathbf{x}_L and \mathbf{x}_H , respectively, which can be obtained by $\mathbf{c}_{L,4} = \mathbf{D}_4 \cdot \mathbf{x}_L$ and $\mathbf{c}_{H,4} = \mathbf{D}_4 \cdot \mathbf{x}_H$, where \mathbf{D}_4 is the transformation matrix

$$\mathbf{D}_4 = \frac{1}{\sqrt{4}} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ \sqrt{2} \cos \frac{\pi}{8} & \sqrt{2} \cos \frac{3\pi}{8} & -\sqrt{2} \cos \frac{3\pi}{8} & -\sqrt{2} \cos \frac{\pi}{8} \\ 1 & -1 & -1 & 1 \\ \sqrt{2} \cos \frac{3\pi}{8} & -\sqrt{2} \cos \frac{\pi}{8} & \sqrt{2} \cos \frac{\pi}{8} & -\sqrt{2} \cos \frac{3\pi}{8} \end{bmatrix}. \quad (2.9)$$

We have

$$\begin{aligned} \hat{\mathbf{c}}_{L,8} &= \mathbf{T}_{\text{SB.DCT},8} \cdot \mathbf{D}_4^{-1} \cdot \mathbf{c}_{L,4} = \begin{bmatrix} 1.412 & 0 & 0 & 0 \\ 0 & 1.3870 & 0 & 0 \\ 0 & 0 & 1.3066 & 0 \\ 0 & 0 & 0 & 1.1759 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.7857 \\ 0 & 0 & -0.5412 & 0 \\ 0 & -0.2759 & 0 & 0 \end{bmatrix} \cdot \mathbf{c}_{L,4}, \\ \hat{\mathbf{s}}_{H,8} &= \mathbf{T}_{\text{SB.DST},8} \cdot \mathbf{D}_4^{-1} \cdot \mathbf{c}_{H,4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.2549 & 0 & -0.1056 & 0 \\ 0 & 0.5 & 0 & -0.2071 \\ 0.3007 & 0 & 0.7259 & 0 \\ 0 & 0.5412 & 0 & 1.3066 \\ 0.4500 & 0 & 1.0864 & 0 \\ 0 & 1.2071 & 0 & -0.5 \\ 1.2815 & 0 & -0.5308 & 0 \end{bmatrix} \cdot \mathbf{c}_{H,4}. \end{aligned} \quad (2.10)$$

Similarly, the 4-point DCT computations of $\mathbf{c}_{L,4}$ and $\mathbf{c}_{H,4}$ can be obtained by

$$\begin{aligned} \mathbf{c}_{L,4} &= [\mathbf{T}_{\text{SB.DCT},4} \quad \mathbf{T}_{\text{SB.DCT},4}]_{4 \times 4} \cdot \mathbf{M}_4 \cdot \mathbf{x}_L \\ &= \underbrace{\mathbf{T}_{\text{SB.DCT},4} \cdot \mathbf{x}_{LL}}_{\hat{\mathbf{c}}_{LL,4}} + \underbrace{\mathbf{T}_{\text{SB.DST},4} \cdot \mathbf{x}_{LH}}_{\hat{\mathbf{s}}_{LH,4}}, \end{aligned} \quad (2.11)$$

$$\begin{aligned} \mathbf{c}_{H,4} &= [\mathbf{T}_{\text{SB.DCT},4} \quad \mathbf{T}_{\text{SB.DCT},4}]_{4 \times 4} \cdot \mathbf{M}_4 \cdot \mathbf{x}_H \\ &= \underbrace{\mathbf{T}_{\text{SB.DCT},4} \cdot \mathbf{x}_{HL}}_{\hat{\mathbf{c}}_{HL,4}} + \underbrace{\mathbf{T}_{\text{SB.DST},4} \cdot \mathbf{x}_{HH}}_{\hat{\mathbf{s}}_{HH,4}}, \end{aligned} \quad (2.12)$$

where $\mathbf{x}_{LL} = [x_{LL}[0] \ x_{LL}[1]]^T$, $\mathbf{x}_{LH} = [x_{LH}[0] \ x_{LH}[1]]^T$, $\mathbf{x}_{HL} = [x_{HL}[0] \ x_{HL}[1]]^T$, $\mathbf{x}_{HH} = [x_{HH}[0] \ x_{HH}[1]]^T$, and

$$x_{LL}[n] = \frac{1}{2}\{x_L[2n] + x_L[2n+1]\}, \quad (2.13)$$

$$x_{LH}[n] = \frac{1}{2}\{x_L[2n] - x_L[2n+1]\}, \quad (2.14)$$

$$x_{HL}[n] = \frac{1}{2} \{x_H[2n] + x_H[2n+1]\}, \quad (2.15)$$

$$x_{HH}[n] = \frac{1}{2} \{x_H[2n] - x_H[2n+1]\} \quad (2.16)$$

for $n = 0, 1$.

The 4-point distributed matrix \mathbf{M}_4 , can be defined as

$$\mathbf{M}_4 = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 \end{bmatrix}. \quad (2.17)$$

Let $\mathbf{c}_{LL,2}$, $\mathbf{c}_{LH,2}$, $\mathbf{c}_{HL,2}$, and $\mathbf{c}_{HH,2}$ be the 2-point DCT of \mathbf{x}_{LL} , \mathbf{x}_{LH} , \mathbf{x}_{HL} , and \mathbf{x}_{HH} , respectively, which can be computed by using the following 2-point transformation matrix, \mathbf{D}_2 :

$$\mathbf{D}_2 = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (2.18)$$

We have

$$\hat{\mathbf{c}}_{LL,A} = \mathbf{T}_{\text{SB.DCT},A} \cdot \mathbf{x}_{LL} = \mathbf{T}_{\text{SB.DCT},A} \cdot \mathbf{D}_2^{-1} \cdot \mathbf{c}_{LL,2} = \begin{bmatrix} 1.4142 & 0 \\ 0 & 1.3066 \\ 0 & 0 \\ 0 & -0.5412 \end{bmatrix} \cdot \mathbf{c}_{LL,2}, \quad (2.19)$$

$$\hat{\mathbf{c}}_{HL,A} = \mathbf{T}_{\text{SB.DCT},A} \cdot \mathbf{x}_{HL} = \mathbf{T}_{\text{SB.DCT},A} \cdot \mathbf{D}_2^{-1} \cdot \mathbf{c}_{HL,2} = \begin{bmatrix} 1.4142 & 0 \\ 0 & 1.3066 \\ 0 & 0 \\ 0 & -0.5412 \end{bmatrix} \cdot \mathbf{c}_{HL,2}, \quad (2.20)$$

$$\hat{\mathbf{c}}_{LH,A} = \mathbf{T}_{\text{SB.DST},A} \cdot \mathbf{x}_{LH} = \mathbf{T}_{\text{SB.DST},A} \cdot \mathbf{D}_2^{-1} \cdot \mathbf{c}_{LH,2} = \begin{bmatrix} 0 & 0 \\ 0.5412 & 0 \\ 0 & 1.4142 \\ 1.3066 & 0 \end{bmatrix} \cdot \mathbf{c}_{LH,2}, \quad (2.21)$$

$$\hat{\mathbf{s}}_{HH,4} = \mathbf{T}_{\text{SB.DST},4} \cdot \mathbf{x}_{HH} = \mathbf{T}_{\text{SB.DST},4} \cdot \mathbf{D}_2^{-1} \cdot \mathbf{c}_{HH,2} = \begin{bmatrix} 0 & 0 \\ 0.5412 & 0 \\ 0 & 1.4142 \\ 1.3066 & 0 \end{bmatrix} \cdot \mathbf{c}_{HH,2}. \quad (2.22)$$

For the 2-point DCT computations of $\mathbf{c}_{LL,2}$, $\mathbf{c}_{LH,2}$, $\mathbf{c}_{HL,2}$, and $\mathbf{c}_{HH,2}$, let

$$x_{LLL}[n] = \frac{1}{2} \{x_{LL}[2n] + x_{LL}[2n+1]\}, \quad (2.23)$$

$$x_{LLH}[n] = \frac{1}{2} \{x_{LL}[2n] - x_{LL}[2n+1]\}, \quad (2.24)$$

$$x_{LHL}[n] = \frac{1}{2} \{x_{LH}[2n] + x_{LH}[2n+1]\}, \quad (2.25)$$

$$x_{LHH}[n] = \frac{1}{2} \{x_{LH}[2n] - x_{LH}[2n+1]\}, \quad (2.26)$$

$$x_{HLL}[n] = \frac{1}{2} \{x_{HL}[2n] + x_{HL}[2n+1]\}, \quad (2.27)$$

$$x_{HLH}[n] = \frac{1}{2} \{x_{HL}[2n] - x_{HL}[2n+1]\}, \quad (2.28)$$

$$x_{HHL}[n] = \frac{1}{2} \{x_{HH}[2n] + x_{HH}[2n+1]\}, \quad (2.29)$$

$$x_{HHH}[n] = \frac{1}{2} \{x_{HH}[2n] - x_{HH}[2n+1]\}, \quad (2.30)$$

where $n = 0$, we then have

$$\mathbf{c}_{LL,2} = [\mathbf{T}_{\text{SB.DCT},2} \quad \mathbf{T}_{\text{SB.DCT},2}]_{2 \times 2} \cdot \mathbf{M}_2 \cdot \mathbf{x}_{LL} = \underbrace{\mathbf{T}_{\text{SB.DCT},2} \cdot \mathbf{x}_{LLL}}_{\hat{\mathbf{c}}_{LLL,2}} + \underbrace{\mathbf{T}_{\text{SB.DST},2} \cdot \mathbf{x}_{LLH}}_{\hat{\mathbf{s}}_{LLH,2}}, \quad (2.31)$$

$$\mathbf{c}_{LH,2} = [\mathbf{T}_{\text{SB.DCT},2} \quad \mathbf{T}_{\text{SB.DCT},2}]_{2 \times 2} \cdot \mathbf{M}_2 \cdot \mathbf{x}_{LH} = \underbrace{\mathbf{T}_{\text{SB.DCT},2} \cdot \mathbf{x}_{LHL}}_{\hat{\mathbf{c}}_{LHL,2}} + \underbrace{\mathbf{T}_{\text{SB.DST},2} \cdot \mathbf{x}_{LHH}}_{\hat{\mathbf{s}}_{LHH,2}}, \quad (2.32)$$

$$\mathbf{c}_{HL,2} = [\mathbf{T}_{\text{SB.DCT},2} \quad \mathbf{T}_{\text{SB.DCT},2}]_{2 \times 2} \cdot \mathbf{M}_2 \cdot \mathbf{x}_{HL} = \underbrace{\mathbf{T}_{\text{SB.DCT},2} \cdot \mathbf{x}_{HLL}}_{\hat{\mathbf{c}}_{HLL,2}} + \underbrace{\mathbf{T}_{\text{SB.DST},2} \cdot \mathbf{x}_{HLH}}_{\hat{\mathbf{s}}_{HLH,2}}, \quad (2.33)$$

$$\mathbf{c}_{HH,2} = [\mathbf{T}_{\text{SB.DCT},2} \quad \mathbf{T}_{\text{SB.DCT},2}]_{2 \times 2} \cdot \mathbf{M}_2 \cdot \mathbf{x}_{HH} = \underbrace{\mathbf{T}_{\text{SB.DCT},2} \cdot \mathbf{x}_{HHL}}_{\hat{\mathbf{c}}_{HHL,2}} + \underbrace{\mathbf{T}_{\text{SB.DST},2} \cdot \mathbf{x}_{HHH}}_{\hat{\mathbf{s}}_{HHH,2}}. \quad (2.34)$$

\mathbf{M}_2 is a 2-point matrix defined as

$$\mathbf{M}_2 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}. \quad (2.35)$$

Finally, according to equations (2.6)~(2.33) together with (2.4), we have

$$\mathbf{c}_8 = \tilde{\mathbf{K}} \cdot \hat{\mathbf{K}} \cdot \mathbf{M}_8 \cdot \widehat{\mathbf{M}}_8 \cdot \widetilde{\mathbf{M}}_8 \cdot \mathbf{x}_8, \quad (2.36)$$

where

$$\begin{aligned} \tilde{\mathbf{K}} &= \left[(\mathbf{T}_{\text{SB.DCT},8} \cdot \mathbf{D}_4^{-1})_{8 \times 4} \quad (\mathbf{T}_{\text{SB.DCT},8} \cdot \mathbf{D}_4^{-1})_{8 \times 4} \right]_{8 \times 8}, \\ \hat{\mathbf{K}} &= \left[\begin{array}{cc} (\mathbf{T}_{\text{SB.DCT},4} \cdot \mathbf{D}_2^{-1})_{4 \times 2} & (\mathbf{T}_{\text{SB.DCT},4} \cdot \mathbf{D}_2^{-1})_{4 \times 2} \\ \mathbf{0}_{4 \times 2} & \mathbf{0}_{4 \times 2} \end{array} \quad \begin{array}{cc} \mathbf{0}_{4 \times 2} & \mathbf{0}_{4 \times 2} \\ (\mathbf{T}_{\text{SB.DCT},4} \cdot \mathbf{D}_2^{-1})_{4 \times 2} & (\mathbf{T}_{\text{SB.DCT},4} \cdot \mathbf{D}_2^{-1})_{4 \times 2} \end{array} \right]_{8 \times 8}. \end{aligned} \quad (2.37)$$

The following matrix, $\widehat{\mathbf{M}}_8$, can derived from (2.12)~(2.15).

$$\widehat{\mathbf{M}}_8 = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & -0.5 \end{bmatrix}. \quad (2.38)$$

Similarly, the following matrix, $\widetilde{\mathbf{M}}_8$, can be derived from (2.21)~(2.28)

$$\widetilde{\mathbf{M}}_8 = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & -0.5 \end{bmatrix}. \quad (2.39)$$

The decomposition matrix \mathbf{R}_8 can be defined as

$$\mathbf{R}_8 = 8 \cdot \mathbf{M}_8 \cdot \widehat{\mathbf{M}}_8 \cdot \widetilde{\mathbf{M}}_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix} \quad (2.40)$$

and the coefficient matrix \mathbf{F}_8 can be defined as

$$\mathbf{F}_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.9061 & 0.3753 & 0 & 0.1802 & 0 & 0 & -0.0747 \\ 0 & 0 & 0 & 0.9239 & 0 & 0.3827 & 0 & 0 \\ 0 & -0.3182 & 0.7682 & 0 & 0.2126 & 0 & 0 & 0.5133 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0.2126 & -0.5133 & 0 & 0.3182 & 0 & 0 & 0.7682 \\ 0 & 0 & 0 & -0.3827 & 0 & 0.9239 & 0 & 0 \\ 0 & -0.1802 & -0.0747 & 1 & 0.9061 & 0 & 0 & -0.3753 \end{bmatrix}. \quad (2.41)$$

According to (2.34), (2.38), and (2.39), we have

$$\mathbf{c}_8 = \frac{\sqrt{2}}{4} \cdot \mathbf{F}_8 \cdot \mathbf{R}_8 \cdot \mathbf{x}_8. \quad (2.42)$$

The coefficient matrix \mathbf{F}_8 can be represented by the reordered coefficient matrix $\widehat{\mathbf{F}}_8$, prepermutation matrix $\widehat{\mathbf{T}}_8$ and post-permutation matrix $\widetilde{\mathbf{T}}_8$, and can be written as

$$\mathbf{F}_8 = \widehat{\mathbf{T}}_8 \cdot \widehat{\mathbf{F}}_8 \cdot \widetilde{\mathbf{T}}_8, \quad (2.43)$$

where the matrices $\hat{\mathbf{F}}_8$, $\hat{\mathbf{T}}_8$, and $\tilde{\mathbf{T}}_8$ can be defined as

$$\hat{\mathbf{F}}_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9238 & 0.3826 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.3826 & 0.9238 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9061 & 0.1802 & 0.3753 & -0.0746 \\ 0 & 0 & 0 & 0 & -0.1802 & 0.9061 & -0.0746 & -0.3753 \\ 0 & 0 & 0 & 0 & 0.2126 & 0.3181 & -0.5132 & 0.7682 \\ 0 & 0 & 0 & 0 & -0.3181 & 0.2126 & 0.7682 & 0.5132 \end{bmatrix},$$

$$\hat{\mathbf{T}}_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad (2.44)$$

$$\tilde{\mathbf{T}}_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The reordered coefficient matrix $\hat{\mathbf{F}}_8$ can be represented as

$$\hat{\mathbf{F}}_8 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{C} \\ \mathbf{0} & \mathbf{0} & \mathbf{D} & \mathbf{E} \end{bmatrix}. \quad (2.45)$$

The computation of sub-coefficient matrix \mathbf{A} can be written as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{A} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad (2.46)$$

where $a = 0.9239$ and $b = 0.3827$. The above can be rewritten as [20]

$$y_1 = (b - a) \cdot x_2 + a \cdot (x_1 + x_2), \quad (2.47)$$

$$y_2 = -(a + b) \cdot x_1 + a \cdot (x_1 + x_2). \quad (2.48)$$

Thus, the number of multiplications can be reduced to 3 for matrix \mathbf{A} ; this technique can also be applied to matrices \mathbf{B} , \mathbf{C} , \mathbf{D} , and \mathbf{E} . As a result, the total number of multiplications of the subband-based 8-point DCT is only 15.

Based on (2.40) and (2.41), the corresponding subband-based IDCT can be obtained by

$$\mathbf{x}_8 = 2\sqrt{2} \cdot \mathbf{R}_8^{-1} \cdot \tilde{\mathbf{T}}_8^{-1} \cdot \hat{\mathbf{F}}_8^{-1} \cdot \hat{\mathbf{T}}_8^{-1} \cdot \mathbf{c}_8, \quad (2.49)$$

where \mathbf{R}_8^{-1} is the inverse decomposition matrix. As the decomposition matrix \mathbf{R}_8 is orthonormal, \mathbf{R}_8^{-1} can be derived from the transportation of \mathbf{R}_8

$$\mathbf{R}_8^{-1} = \mathbf{R}_8^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}. \quad (2.50)$$

$\hat{\mathbf{F}}_8^{-1}$ is the inverse reordered coefficient matrix, $\hat{\mathbf{T}}_8^{-1}$ is the inverse prepermutation matrix and $\tilde{\mathbf{T}}_8^{-1}$ is the inverse post-permutation matrix. Since the reordered coefficient matrix $\hat{\mathbf{F}}_8$, prepermutation matrix $\hat{\mathbf{T}}_8$ and post-permutation matrix $\tilde{\mathbf{T}}_8$ are all orthonormal, three matrices $\hat{\mathbf{F}}_8^{-1}$, $\hat{\mathbf{T}}_8^{-1}$, and $\tilde{\mathbf{T}}_8^{-1}$ can be written as

$$\hat{\mathbf{F}}_8^{-1} = \hat{\mathbf{F}}_8^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9238 & -0.3826 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.3826 & 0.9238 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9061 & -0.1802 & 0.2126 & -0.3181 \\ 0 & 0 & 0 & 0 & 0.1802 & 0.9061 & 0.3181 & 0.2126 \\ 0 & 0 & 0 & 0 & 0.3753 & -0.0746 & -0.5132 & 0.7682 \\ 0 & 0 & 0 & 0 & -0.0746 & -0.3753 & 0.7682 & 0.5132 \end{bmatrix}, \quad (2.51)$$

$$\hat{\mathbf{T}}_8^{-1} = \hat{\mathbf{T}}_8^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (2.52)$$

$$\tilde{\mathbf{T}}_8^{-1} = \tilde{\mathbf{T}}_8^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2.53)$$

The inverse reordered coefficient matrix $\hat{\mathbf{F}}_8^{-1}$ can be represented as

$$\hat{\mathbf{F}}_8^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}^T & \mathbf{D}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{C}^T & \mathbf{E}^T \end{bmatrix}. \quad (2.54)$$

As a result, the total number of multiplications of the subband based 8-point IDCT is only 15.

3. The Unified Subband-Based n -Point DCT/IDCT Algorithm

The subband-based DCT algorithm [22] can be unified for n -point DCT/IDCT due to the inherent regular pattern. For an n -point signal, \mathbf{x}_n , the unified subband-based discrete cosine transform can be defined as

$$\mathbf{c}_n = \frac{\sqrt{n}}{n} \cdot \hat{\mathbf{T}}_n \cdot \hat{\mathbf{F}}_n \cdot \tilde{\mathbf{T}}_n \cdot \mathbf{R}_n \cdot \mathbf{x}_n, \quad (3.1)$$

where $n = \{2^m | m = 3, 4, 5, \dots\}$, \mathbf{R}_n is the decomposition matrix, $\hat{\mathbf{F}}_n$ is the reordered coefficient matrix, $\hat{\mathbf{T}}_n$ is the pre-permutation matrix, and $\tilde{\mathbf{T}}_n$ is the post-permutation matrix. The unified decomposition matrix \mathbf{R}_n can be written as

$$\mathbf{R}_n = n \cdot \mathbf{M}_{n,n \times n} \cdot \begin{bmatrix} \mathbf{M}_{n/2,(n/2) \times (n/2)} & \mathbf{0}_{(n/2) \times (n/2)} \\ \mathbf{0}_{(n/2) \times (n/2)} & \mathbf{M}_{n/2,(n/2) \times (n/2)} \end{bmatrix}_{n \times n} \cdot \begin{bmatrix} \mathbf{M}_{n/4,(n/4) \times (n/4)} & \mathbf{0}_{(n/4) \times (n/4)} & \mathbf{0}_{(n/4) \times (n/4)} & \mathbf{0}_{(n/4) \times (n/4)} \\ \mathbf{0}_{(n/4) \times (n/4)} & \mathbf{M}_{n/4,(n/4) \times (n/4)} & \mathbf{0}_{(n/4) \times (n/4)} & \mathbf{0}_{(n/4) \times (n/4)} \\ \mathbf{0}_{(n/4) \times (n/4)} & \mathbf{0}_{(n/4) \times (n/4)} & \mathbf{M}_{n/4,[(n/4) \times (n/4)]} & \mathbf{0}_{(n/4) \times (n/4)} \\ \mathbf{0}_{(n/4) \times (n/4)} & \mathbf{0}_{(n/4) \times (n/4)} & \mathbf{0}_{(n/4) \times (n/4)} & \mathbf{M}_{n/4,(n/4) \times (n/4)} \end{bmatrix}_{n \times n} \cdots \begin{bmatrix} \mathbf{M}_{2,2 \times 2} & \mathbf{0}_{2 \times 2} & \cdots & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{M}_{2,2 \times 2} & & \vdots \\ \vdots & & \ddots & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \cdots & \mathbf{0}_{2 \times 2} & \mathbf{M}_{2,2 \times 2} \end{bmatrix}_{n \times n}, \quad (3.2)$$

where the basic 2×2 matrix, \mathbf{M}_2 , consists of two submatrices, $\mathbf{M}'_{2,1 \times 2}$ and $\mathbf{M}''_{2,1 \times 2}$

$$\mathbf{M}_2 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix} = \begin{bmatrix} \mathbf{M}'_{2,1 \times 2} \\ \mathbf{M}''_{2,1 \times 2} \end{bmatrix}. \quad (3.3)$$

As noted, \mathbf{M}_4 can be represented by the sub-matrices of \mathbf{M}_2 , or sub-matrices, $\mathbf{M}'_{4,2 \times 4}$ and $\mathbf{M}''_{4,2 \times 4}$ as follows:

$$\mathbf{M}_4 = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 \end{bmatrix} = \begin{bmatrix} \mathbf{M}'_{2,1 \times 2} & \mathbf{0}_{1 \times 2} \\ \mathbf{0}_{1 \times 2} & \mathbf{M}'_{2,1 \times 2} \\ \mathbf{M}''_{2,1 \times 2} & \mathbf{0}_{1 \times 2} \\ \mathbf{0}_{1 \times 2} & \mathbf{M}''_{2,1 \times 2} \end{bmatrix} = \begin{bmatrix} \mathbf{M}'_{4,2 \times 4} \\ \mathbf{M}''_{4,2 \times 4} \end{bmatrix}. \quad (3.4)$$

According to (3.1) and (3.2), the unified distributed matrix \mathbf{M}_n can be derived as

$$\mathbf{M}_n = \begin{bmatrix} \mathbf{M}'_{n/2,(n/4) \times (n/2)} & \mathbf{0}_{(n/4) \times (n/2)} \\ \mathbf{0}_{(n/4) \times (n/2)} & \mathbf{M}'_{n/2,(n/4) \times (n/2)} \\ \mathbf{M}''_{n/2,(n/4) \times (n/2)} & \mathbf{0}_{(n/4) \times (n/2)} \\ \mathbf{0}_{(n/4) \times (n/2)} & \mathbf{M}''_{n/2,(n/4) \times (n/2)} \end{bmatrix} = \begin{bmatrix} \mathbf{M}'_{n,(n/2) \times n} \\ \mathbf{M}''_{n,(n/2) \times n} \end{bmatrix}. \quad (3.5)$$

According to (2.2) and (2.40), we have

$$\mathbf{F}_n = \mathbf{D}_n \cdot \mathbf{R}_n^{-1}. \quad (3.6)$$

The 4-point reordered coefficient matrix $\widehat{\mathbf{F}}_4$ can be derived as

$$\widehat{\mathbf{F}}_{4,4 \times 4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.9239 & 0.3827 \\ 0 & 0 & -0.3827 & 0.9239 \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{F}}_{2,2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \widehat{\mathbf{F}}'_{4,2 \times 2} \end{bmatrix}. \quad (3.7)$$

According to (3.5), the 8-point reordered coefficient matrix $\widehat{\mathbf{F}}_8$ can be derived as

$$\widehat{\mathbf{F}}_{8,8 \times 8} = \begin{bmatrix} \widehat{\mathbf{F}}_{2,2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{2 \times 2} & \widehat{\mathbf{F}}'_{4,2 \times 2} & \mathbf{0}_{2 \times 4} \\ \mathbf{0}_{4 \times 2} & \mathbf{0}_{4 \times 2} & \widehat{\mathbf{F}}'_{8,4 \times 4} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{F}}_{4,4 \times 4} & \mathbf{0}_{4 \times 4} \\ \mathbf{0}_{4 \times 4} & \widehat{\mathbf{F}}'_{8,4 \times 4} \end{bmatrix}. \quad (3.8)$$

Hence, the unified reordered coefficient matrix can be written as

$$\widehat{\mathbf{F}}_{n,n \times n} = \begin{bmatrix} \widehat{\mathbf{F}}_{2,2 \times 2} & \mathbf{0}_{2 \times 2} & \cdots & \mathbf{0}_{2 \times n} \\ \mathbf{0}_{2 \times 2} & \widehat{\mathbf{F}}'_{4,2 \times 2} & & \vdots \\ \vdots & & \ddots & \mathbf{0}_{(n/2) \times n} \\ \mathbf{0}_{n \times 2} & \cdots & \mathbf{0}_{n \times (n/2)} & \widehat{\mathbf{F}}'_{n,(n/2) \times (n/2)} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{F}}_{n/2,(n/2) \times (n/2)} & \mathbf{0}_{n/2,(n/2) \times (n/2)} \\ \mathbf{0}_{n/2,(n/2) \times (n/2)} & \widehat{\mathbf{F}}'_{n,(n/2) \times (n/2)} \end{bmatrix}. \quad (3.9)$$

The 4-point pre-permutation matrix $\hat{\mathbf{T}}_4$ and post-permutation matrix $\tilde{\mathbf{T}}_4$ can be written as

$$\hat{\mathbf{T}}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{T}}_{41,4 \times 1} & \hat{\mathbf{T}}_{42,4 \times 1} & \hat{\mathbf{T}}_{43,4 \times 1} & \hat{\mathbf{T}}_{44,4 \times 1} \end{bmatrix}, \quad (3.10)$$

$$\tilde{\mathbf{T}}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{T}}_{41,4 \times 1} & \tilde{\mathbf{T}}_{42,4 \times 1} & \tilde{\mathbf{T}}_{43,4 \times 1} & \tilde{\mathbf{T}}_{44,4 \times 1} \end{bmatrix}.$$

According to (2.50), (2.51), (3.8), and (3.9), the 8-point pre- and post-permutation matrices can be written as

$$\hat{\mathbf{T}}_8 = \begin{bmatrix} \hat{\mathbf{T}}_{41,4 \times 1} & \mathbf{0}_{4 \times 1} & \hat{\mathbf{T}}_{42,4 \times 1} & \mathbf{0}_{4 \times 1} & \hat{\mathbf{T}}_{43,4 \times 1} & \mathbf{0}_{4 \times 1} & \hat{\mathbf{T}}_{44,4 \times 1} & \mathbf{0}_{4 \times 1} \\ \mathbf{0}_{4 \times 1} & \hat{\mathbf{T}}'_{81,4 \times 1} & \mathbf{0}_{4 \times 1} & \hat{\mathbf{T}}'_{82,4 \times 1} & \mathbf{0}_{4 \times 1} & \hat{\mathbf{T}}'_{83,4 \times 1} & \mathbf{0}_{4 \times 1} & \hat{\mathbf{T}}'_{84,4 \times 1} \end{bmatrix}, \quad (3.11)$$

$$\tilde{\mathbf{T}}_8 = \begin{bmatrix} \tilde{\mathbf{T}}_{41,4 \times 1} & \mathbf{0}_{4 \times 1} & \tilde{\mathbf{T}}_{42,4 \times 1} & \mathbf{0}_{4 \times 1} & \tilde{\mathbf{T}}_{43,4 \times 1} & \mathbf{0}_{4 \times 1} & \tilde{\mathbf{T}}_{44,4 \times 1} & \mathbf{0}_{4 \times 1} \\ \mathbf{0}_{4 \times 1} & \tilde{\mathbf{T}}'_{81,4 \times 1} & \mathbf{0}_{4 \times 1} & \tilde{\mathbf{T}}'_{82,4 \times 1} & \mathbf{0}_{4 \times 1} & \tilde{\mathbf{T}}'_{83,4 \times 1} & \mathbf{0}_{4 \times 1} & \tilde{\mathbf{T}}'_{84,4 \times 1} \end{bmatrix}.$$

Hence, the unified pre- and post-permutation matrices can be represented as

$$\hat{\mathbf{T}}_n = \begin{bmatrix} \hat{\mathbf{T}}_{(n/2)1, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \hat{\mathbf{T}}_{(n/2)2, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \hat{\mathbf{T}}_{(n/2)3, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \hat{\mathbf{T}}_{(n/2)4, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} \\ \mathbf{0}_{\mathcal{A}} & \hat{\mathbf{T}}'_{n1, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \hat{\mathbf{T}}'_{n2, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \hat{\mathbf{T}}'_{n3, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \hat{\mathbf{T}}'_{n4, \mathcal{A}} \end{bmatrix},$$

$$\tilde{\mathbf{T}}_n = \begin{bmatrix} \tilde{\mathbf{T}}_{(n/2)1, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \tilde{\mathbf{T}}_{(n/2)2, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \tilde{\mathbf{T}}_{(n/2)3, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \tilde{\mathbf{T}}_{(n/2)4, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} \\ \mathbf{0}_{(n/2) \times (n/8)} & \tilde{\mathbf{T}}'_{(n/2)1, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \tilde{\mathbf{T}}'_{(n/2)2, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \tilde{\mathbf{T}}'_{(n/2)3, \mathcal{A}} & \mathbf{0}_{\mathcal{A}} & \tilde{\mathbf{T}}'_{(n/2)4, \mathcal{A}} \end{bmatrix}. \quad (3.12)$$

where \mathcal{A} denotes $(n/2) \times (n/8)$.

According to (2.47) and (2.53), the unified subband-based IDCT can be obtained.

$$\mathbf{x}_n = \frac{n}{\sqrt{n}} \cdot \mathbf{R}_n^{-1} \cdot \tilde{\mathbf{T}}_n^{-1} \cdot \hat{\mathbf{F}}_n^{-1} \cdot \hat{\mathbf{T}}_n^{-1} \cdot \mathbf{c}_n, \quad (3.13)$$

where \mathbf{R}_n^{-1} is the inverse decomposition matrix, $\hat{\mathbf{F}}_n^{-1}$ is the inverse coefficient matrix, $\hat{\mathbf{T}}_n^{-1}$ is the inverse pre-transportation matrix, and $\tilde{\mathbf{T}}_n^{-1}$ is the inverse post-transportation matrix. Note that

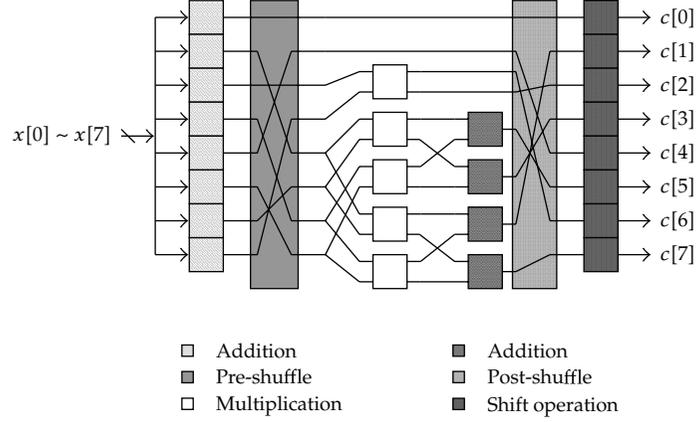


Figure 1: Data flow of the subband-based 8-point DCT with six pipelined stages.

the decomposition matrix, reordered coefficient matrix, pre- and post-permutation matrices are all orthonormal. Hence, we have

$$\begin{aligned}
 \mathbf{R}_n^{-1} &= \mathbf{R}_n^T, \\
 \hat{\mathbf{F}}_n^{-1} &= \hat{\mathbf{F}}_n^T, \\
 \hat{\mathbf{T}}_n^{-1} &= \hat{\mathbf{T}}_n^T, \\
 \tilde{\mathbf{T}}_n^{-1} &= \tilde{\mathbf{T}}_n^T.
 \end{aligned} \tag{3.14}$$

4. Analysis of Computation Complexity and System Performance

Based on the 8-point subband-based DCT and IDCT algorithm, the data flow of parallel-pipelined processing for 8-point DCT and IDCT are described as follows. The data flow of the subband-based 8-point DCT with six pipelined stages is shown in Figure 1. In which, $\mathbf{y}_8 = \mathbf{R}_8 \cdot \mathbf{x}_8$, $\mathbf{z}_8 = \hat{\mathbf{T}}_8 \cdot \hat{\mathbf{F}}_8 \cdot \tilde{\mathbf{T}}_8 \cdot \mathbf{y}_8$ and $\mathbf{c}_8 = (\sqrt{2}/4) \cdot \mathbf{z}_8$, the matrix-vector multiplication of $\mathbf{R}_8 \cdot \mathbf{x}_8$ in the first stage, takes one simple-addition time for each element of \mathbf{y}_8 . The preshuffle performs the prepermutation matrix $\tilde{\mathbf{T}}_8$ operation in the second stage. The matrix-vector multiplication is used to compute $\hat{\mathbf{F}}_8 \cdot \mathbf{y}_8$ in the third and fourth stages. In the fifth stage, the postshuffle is used for the post-permutation matrix $\hat{\mathbf{T}}_8$. The final stage is to compute $(\sqrt{2}/4) \cdot \mathbf{z}_8$ by using simple shift operation with the Booth recoded algorithm.

Similarly, Figure 2 shows the data flow of the subband-based 8-point IDCT with seven pipelined stages.

In which, $\mathbf{u}_8 = (\sqrt{2}/4) \cdot \mathbf{c}_8$, $\mathbf{w}_8 = \hat{\mathbf{F}}_8^{-1} \cdot \hat{\mathbf{T}}_8^{-1} \cdot \mathbf{u}_8$, $\mathbf{v}_8 = \tilde{\mathbf{T}}_8^{-1} \cdot \mathbf{w}_8$, and $\mathbf{x}_8 = \mathbf{R}_8^{-1} \cdot \mathbf{v}_8$, it performs $\mathbf{u}_8 = (\sqrt{2}/4) \cdot \mathbf{c}_8$ by using shift operation with the Booth recoded algorithm in the first stage. The preshuffle performs the pre-permutation matrix $\hat{\mathbf{T}}_8^{-1}$ operation in the second stage. The matrix-vector multiplication is used to compute \mathbf{w}_8 in the third and fourth stages.

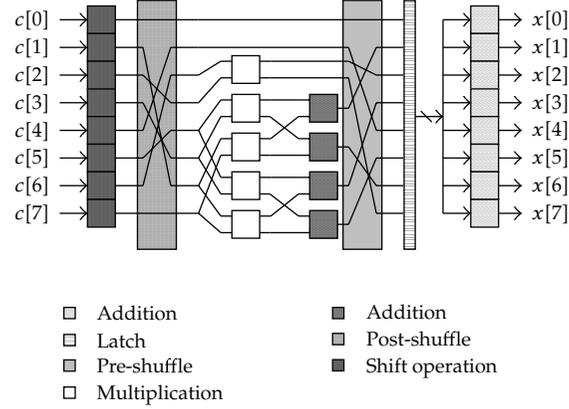


Figure 2: Data flow of the subband-based 8-point IDCT with seven pipelined stages.

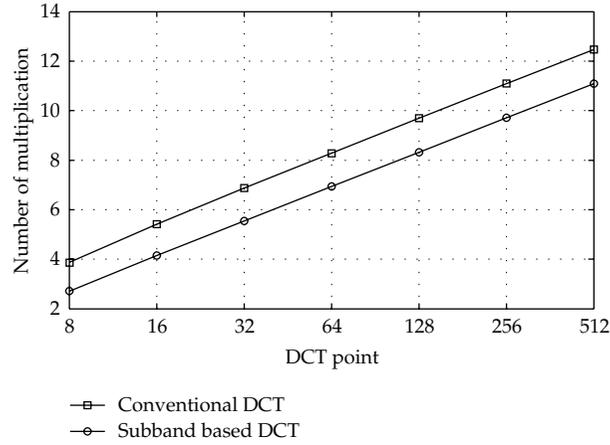


Figure 3: Log plot of the number of multiplications versus the number of DCT points.

In the fifth stage, the post-shuffle performs the post-permutation matrix $\tilde{\mathbf{T}}_8^{-1}$. The sixth and seventh stages are to perform $\mathbf{R}_8^{-1} \cdot \mathbf{v}_8$ with simple addition. Both of the subband-based DCT and IDCT algorithms need one multiplication operation with parallel-pipelined processing, in comparison to [23] using linear array, which needs five multiplication operations.

Recall that the DCT of a signal, \mathbf{x}_n , can be represented as $\mathbf{c}_n = (\sqrt{n}/n) \cdot \hat{\mathbf{T}}_n \cdot \hat{\mathbf{F}}_n \cdot \tilde{\mathbf{T}}_n \cdot \mathbf{R}_n \cdot \mathbf{x}_n$. The multiplication time of the unified subband-based algorithm can be derived as

$$T_M = 3 \cdot \left(4^0 + 4^1 + 4^2 + \dots + 4^{(\log_2 n) - 2} \right) = 4^{(\log_2 n) - 1} - 1, \quad (4.1)$$

where $n = \{2^m \mid m = 3, 4, 5, \dots\}$.

The log plot of the subband-based DCT computations versus the number of DCT points is shown in Figure 3.

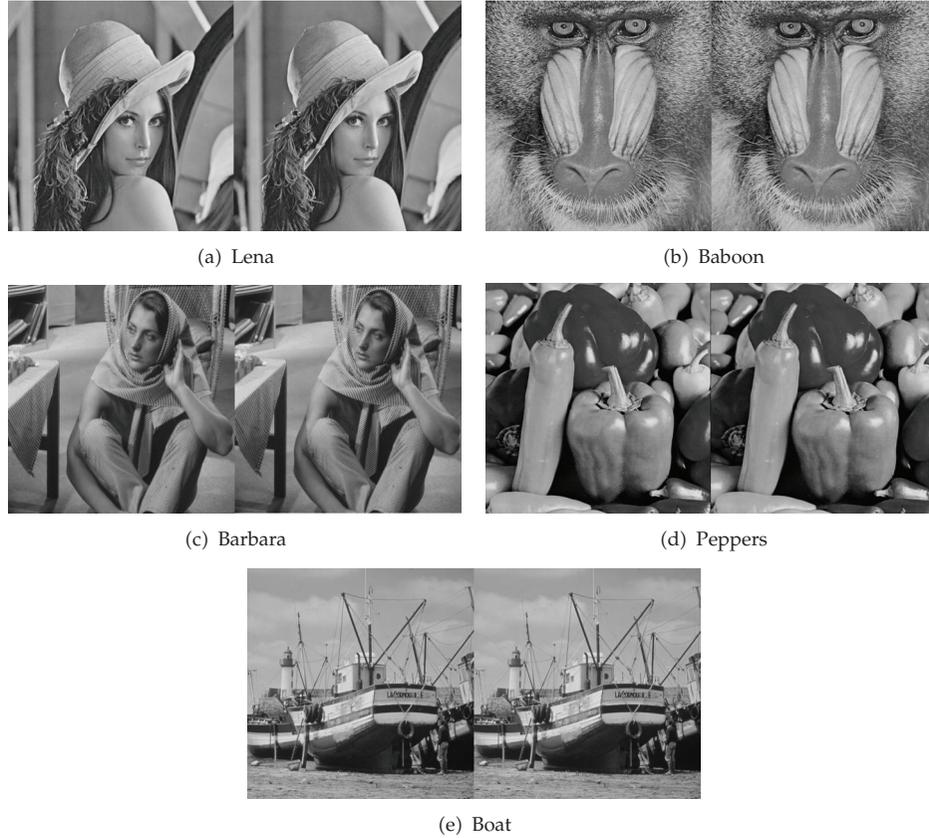


Figure 4: The original image (left side) and the reconstructed image (right side).

The multiplication time of the unified subband-based DCT with single processor can be derived as follows [23]:

$$T_M = 4^0 + 4^1 + 4^2 + \dots + 4^{(\log_2 n)-2} = \frac{(4^{(\log_2 n)-1} - 1)}{3}. \quad (4.2)$$

The left side of Figures 4(a), 4(b), 4(c), 4(d), and 4(e) show the original 512×512 Lena, Baboon, Barbara, Peppers, and Boat images, respectively. The reconstructed Lena, Baboon, Barbara, Peppers and boat image shown in the right side of Figures 4(a), 4(b), 4(c), 4(d), and 4(e), respectively, are obtained by using the proposed subband based 8-point DCT/IDCT algorithm with 32-bit fixed-point operands; the peak signal-to-noise ratios (PSNRs) of Lena, Baboon, Barbara, Peppers, and Boat images are 149.67 dB, 142.12 dB, 143.08 dB, 143.36 dB, and 140.79 dB, respectively.

The PSNR curves of Lena, Baboon, Barbara, Peppers, and Boat images obtained by using the conventional 8-point DCT and the proposed subband-based 8-point DCT at various word lengths are shown in Figure 5. Figures 6(a), 6(b), 6(c), 6(d), and 6(e) show the PSNR curves of Lena, Baboon, Barbara, Peppers, and Boat images obtained by using

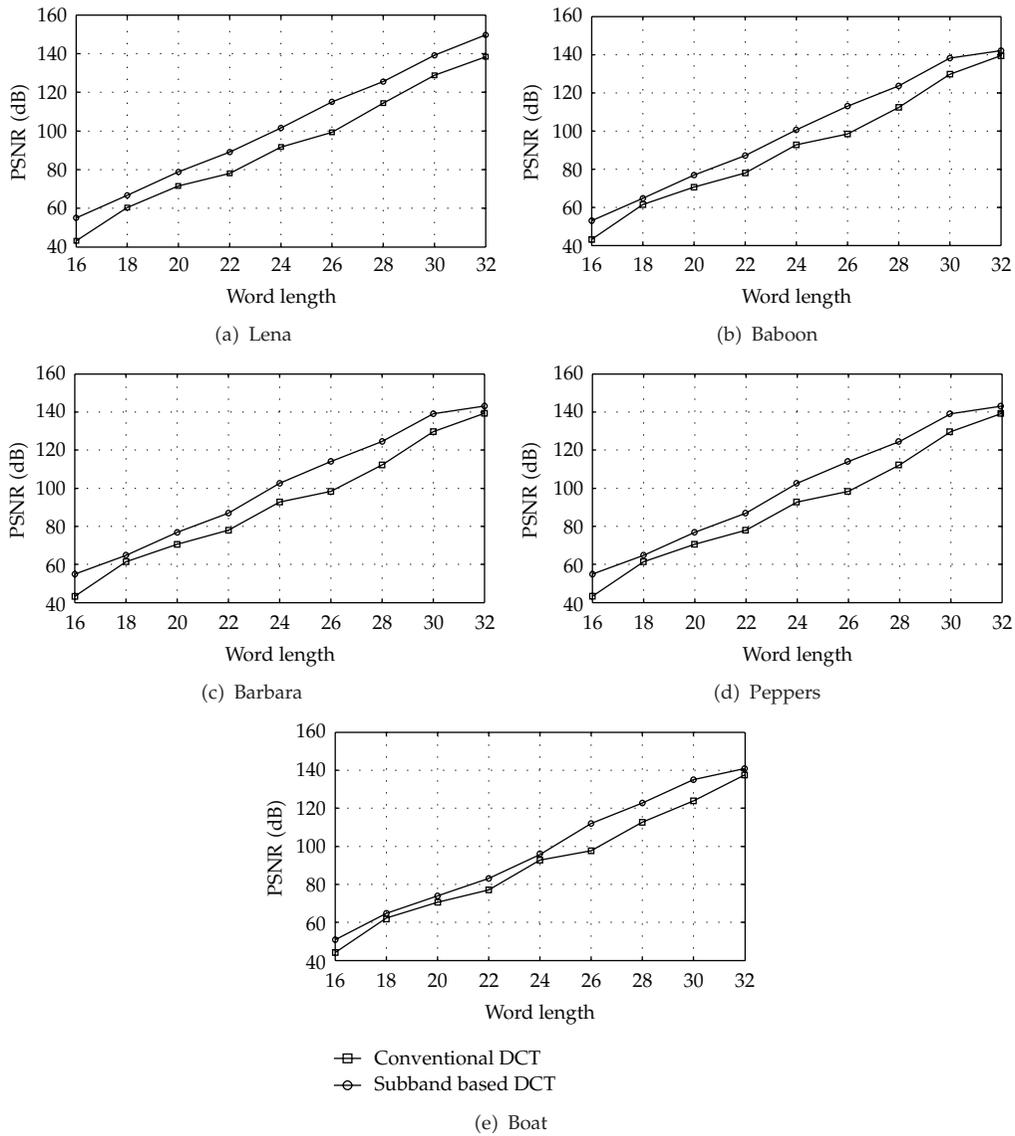


Figure 5: The PSNR curves of (a) Lena, (b) Baboon, (c) Barbara, (d) Peppers, and (e) Boat images obtained by using the conventional 8-point DCT and the proposed subband-based 8-point DCT at various word lengths.

the conventional DCT and the proposed subband-based DCT with 32-bit operand at various DCT points. As one can see, the subband-based DCT is preferable.

5. FPGA Implementation of the Reconfigurable Linear-Array DCT/IDCT Processor

The reconfigurable architecture of the fast 8-, 16-, 32- and 64-point DCT and IDCT processors based on the subband-based 8-point DCT is proposed in this section.

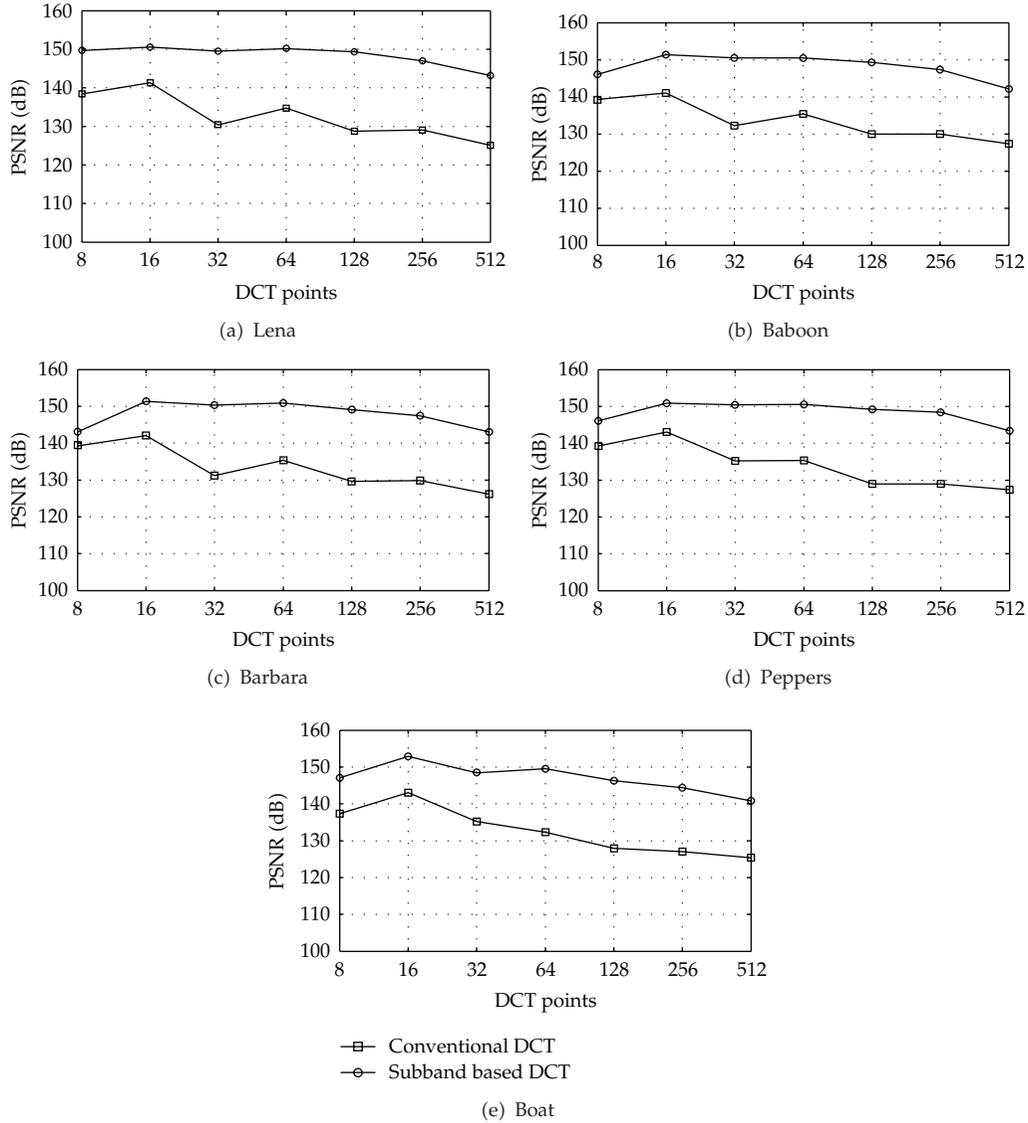


Figure 6: PSNR comparisons of (a) Lena, (b) Baboon, (c) Barbara, (d) Peppers, and (e) Boat images using the conventional DCT and the subband-based DCT with 32-bit operand at various DCT points.

5.1. The Proposed 8-Point DCT/IDCT Processor

According to the data flow of the subband-based 8-point DCT with six pipelined stages (Figure 1), the architecture of the proposed 8-point DCT processor is shown in Figure 7. In which, the adder array (AA) with three CSA(4,2)s performing the matrix-vector multiplication of $\hat{\mathbf{R}}_8 \cdot \mathbf{x}_8$ is shown in Figure 8. Figure 9 shows the multiplier array (MA) performing three types of operation, which are needed to compute the subcoefficient matrix computation of $\hat{\mathbf{F}}_8$. The control signals of *swap* and *inv* determine the types of operation. The functions determined by *swap* and *inv* are shown in Table 1. Figure 10 shows the hardwired

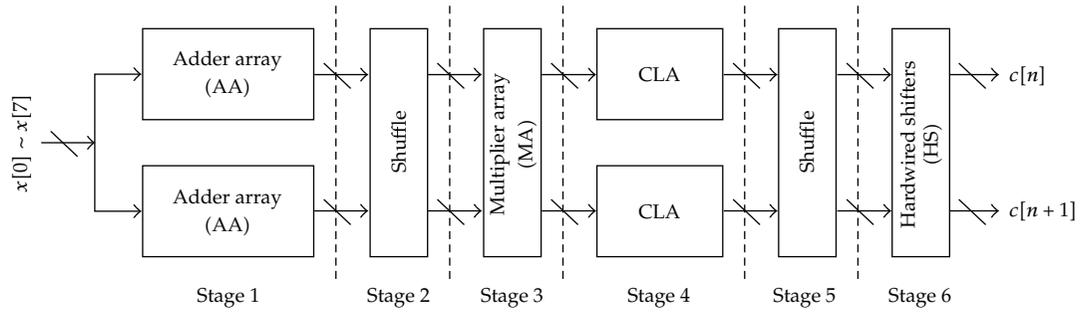


Figure 7: The proposed 8-point DCT Processor with six pipelined stages.

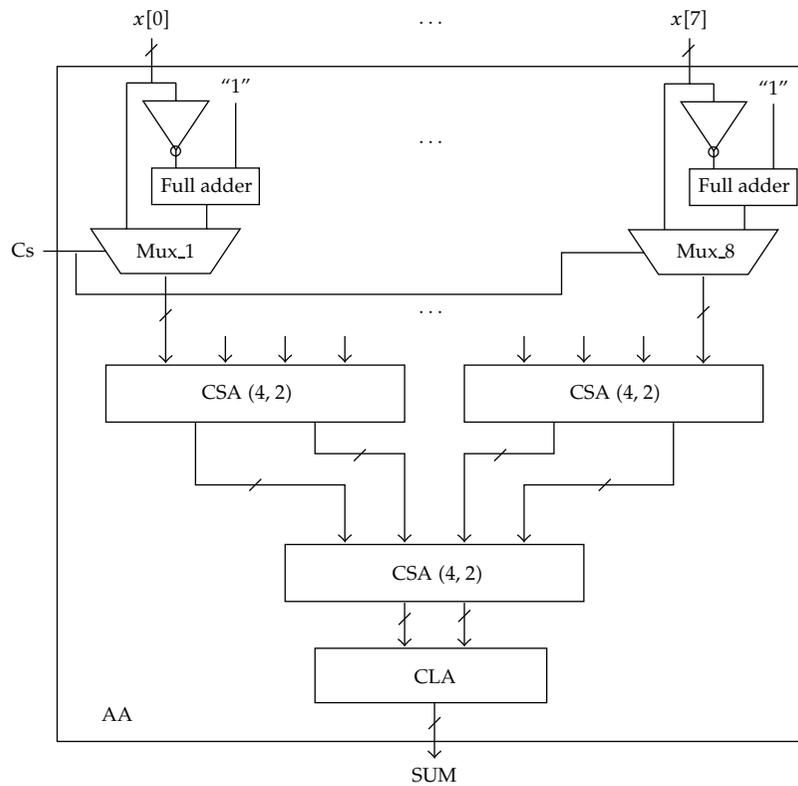


Figure 8: Adder array (AA) for the matrix-vector multiplication of $R_8 \cdot x_8$.

shifters used for performing $(\sqrt{2}/4) \cdot z_8$ by the Booth recoded algorithm [23]. Figure 11 shows the proposed 8-point IDCT processor with seven pipelined stages. In which, the fast adder arrays, shuffle, multiplier array, CLA, and hardwired shifters for DCT architecture can also be used for performing IDCT. The latch array for retiming the input data is shown in Figure 12.

The hardware complexity of the proposed subband-based IDCT architecture is the same as that of the proposed subband-based DCT architecture. Figure 13 shows the proposed integrated 8-point DCT/IDCT processor.

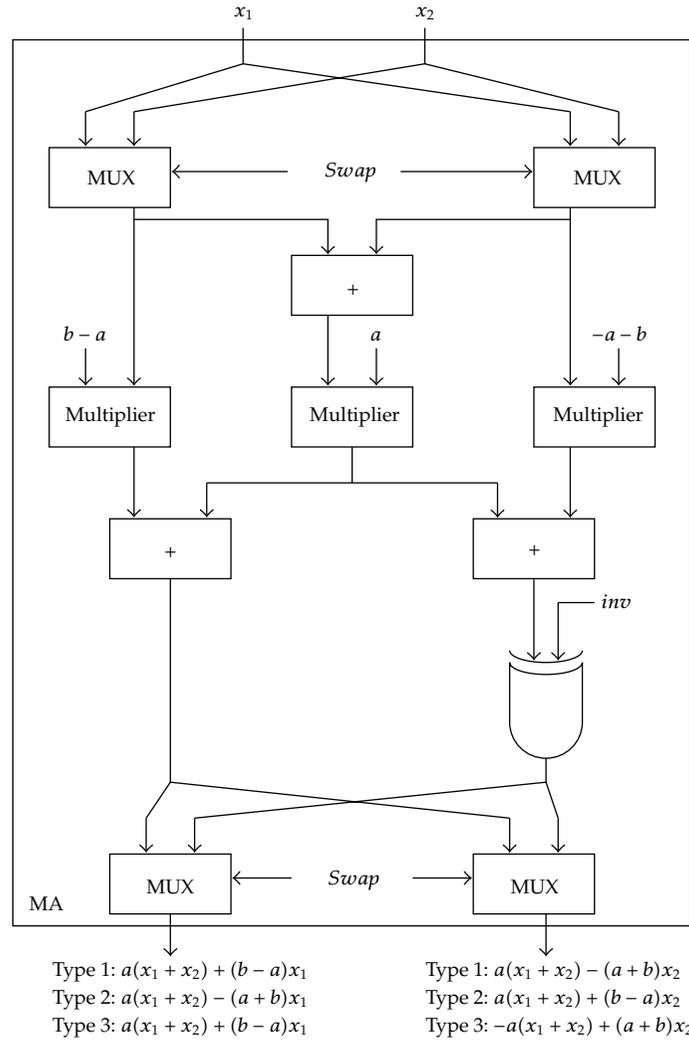


Figure 9: Multiplier array (MA) for computing the subcoefficient matrix of \hat{F}_8 .

Table 1: The functions determined by *swap* and *inv*.

	Type 1	Type 2	Type 3
<i>swap</i>	0	1	0
<i>inv</i>	0	0	1

5.2. The Proposed Reconfigurable DCT/IDCT Processor

According to the integrated 8-point DCT/IDCT processor (Figure 13), the proposed reconfigurable 8-, 16-, 32-, and 64-point DCT/IDCT processor is shown in Figure 14. In which, the integrated adder array (IAA) for the fast computation of 8-, 16-, 32-, and 64-point DCT/IDCT is shown in Figure 15. The modified hardwired shifter (MHS) for multiplication by \sqrt{n}/n (where $n = 8, 16, 32, 64$) using the Booth recoded algorithm is shown in Figure 16.

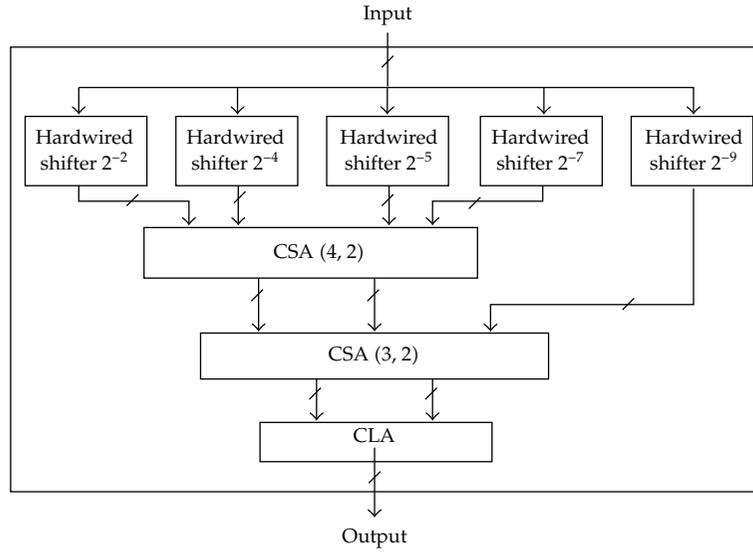


Figure 10: The proposed hardwired shifters used for performing $(\sqrt{2}/4) \cdot z_8$.

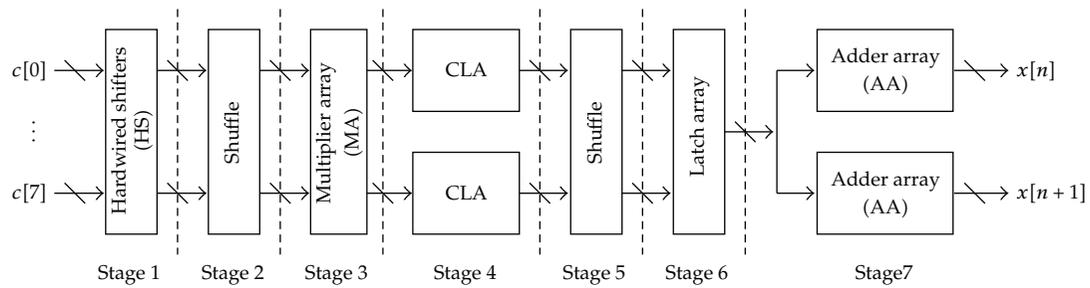


Figure 11: The proposed 8-point IDCT processor with seven pipelined stages.

In order to improve the computation efficiency, the number of multiplier arrays should be increased. The log plot of computation cycles versus number of multiplier arrays is shown in Figure 17.

5.3. FPGA Implementation of the Reconfigurable 2D DCT/IDCT Processor

The $N \times N$ DCT is defined as [29]

$$Z[u, v] = \frac{2 \cdot \alpha[u] \cdot \alpha[v]}{N} \cdot \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x[m, n] \cdot \cos\left(\frac{2m+1}{2N} u\pi\right) \cdot \cos\left(\frac{2n+1}{2N} v\pi\right), \quad (5.1)$$

$$0 \leq u, v \leq N-1,$$

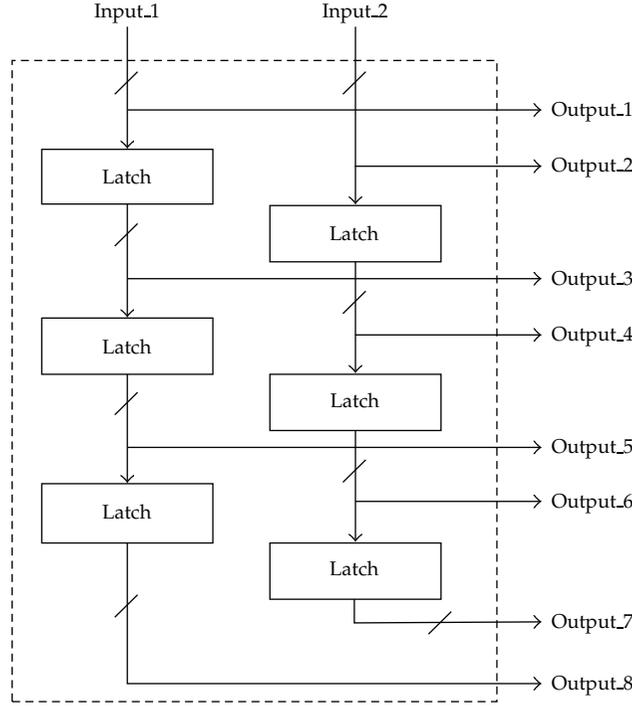


Figure 12: The latch array for retiming the input data.

where $\alpha[k] = 1/\sqrt{2}$ for $k = 0$, and $\alpha[k] = 1$ for $k > 0$. It can be rewritten as

$$Z[u, v] = \sqrt{\frac{2}{N}} \alpha[u] \sum_{m=0}^{N-1} \left[\sqrt{\frac{2}{N}} \alpha[v] \sum_{n=0}^{N-1} x(m, n) \cdot \cos\left(\frac{2n+1}{2N} v\pi\right) \right] \cdot \cos\left(\frac{2m+1}{2N} u\pi\right),$$

$$0 \leq u, v \leq N-1. \quad (5.2)$$

Thus, the separable 2-D DCT can be obtained by using 1-D DCT as follows:

$$2\text{-D DCT}(\mathbf{X}) = 1\text{-D DCT}\left(\left(1\text{-D DCT}(\mathbf{X})\right)^T\right). \quad (5.3)$$

Similarly, the separable 2-D IDCT can be obtained by using 1-D IDCT as follows:

$$2\text{-D IDCT}(\mathbf{Z}) = 1\text{-D IDCT}\left(\left(1\text{-D IDCT}(\mathbf{Z})\right)^T\right). \quad (5.4)$$

As a result, the architecture of 2D DCT/IDCT can be implemented by using two successive 1D DCT/IDCT processors with only one transpose memory [29]. The proposed architecture of 2-D DCT and IDCT is shown in Figure 18. In which, the control signals provided by the finite state machine (FSM) controller are used to manage the data flow and the

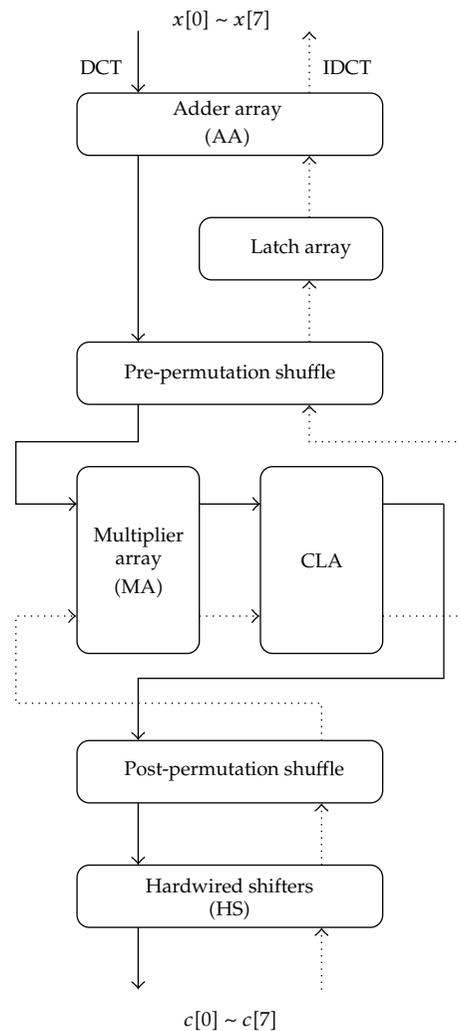


Figure 13: The integrated 8-point DCT/IDCT processor.

operation timing for the DCT/IDCT and transpose memory; the transpose memory allows simultaneous read and write operations between the two processors while performing matrix transposition. The data read and written timing diagram for 8×8 DCT/IDCT system is shown in Figure 19. In comparison with the conventional two transpose memories based 2-D DCT/IDCT architectures, the proposed architecture utilizes only one transpose memory.

The platform for architecture development and verification has been designed as well as implemented in order to evaluate the development cost. The architecture has been implemented on the Xilinx FPGA emulation board [30]. The Xilinx Spartan-3 FPGA has been integrated with the microcontroller (MCU) and I/O interface circuit (USB 2.0) to form the architecture development and verification platform. Figure 20 depicts block diagram and circuit board of the architecture development and evaluation platform. In which, the

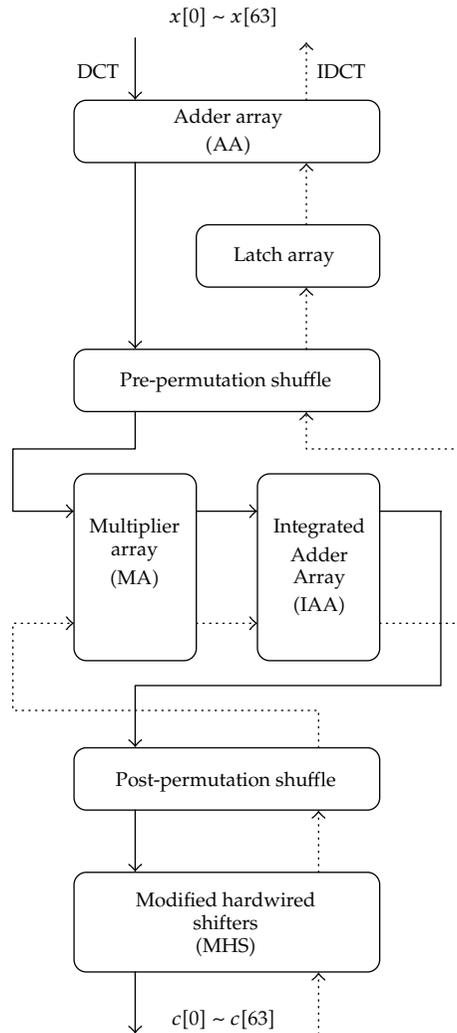


Figure 14: The reconfigurable 8-, 16-, 32-, and 64-point DCT/IDCT processor.

microcontroller reads data and commands from PC and writes the results back to PC by USB 2.0; the Xilinx Spartan-3 FPGA implements the proposed 2-D DCT/IDCT processor. The hardware code written in Verilog is for PC with the ModelSim simulation tool [31] and Xilinx ISE smart compiler [32]. It is noted that the throughput can be improved by using the proposed architecture while the computation accuracy is the same as that obtained by using the conventional one with the same word length. Thus, the proposed programmable DCT/IDCT architecture is able to improve the power consumption and computation speed significantly. The proposed processor for 8-, 16-, 32-, and 64-point DCT/IDCT is an extension of the 8-point DCT/IDCT processor. Moreover, the reusable intellectual property (IP) DCT/IDCT core has also been implemented in Verilog for the hardware realization. All the control signals are internally generated on chip. The proposed DCT/IDCT processor provides both high throughput and low gate count.

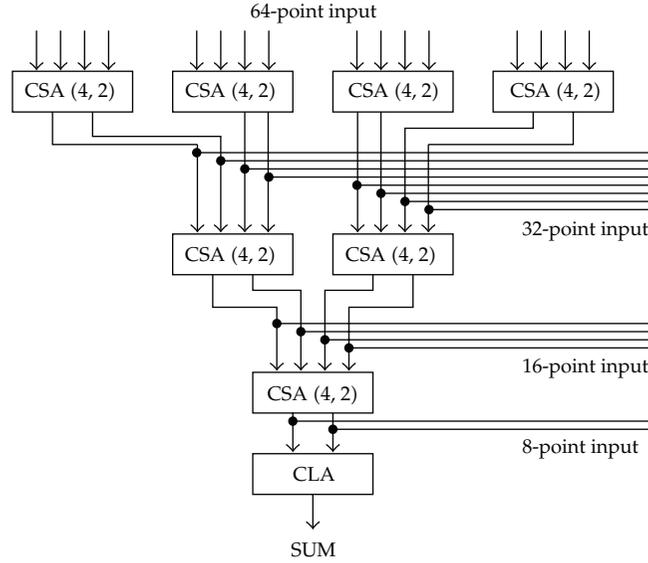


Figure 15: Integrated adder array (IAA) for 8-, 16-, 32-, and 64-point DCT/IDCT processor.

Table 2: Comparisons between the proposed algorithm and architecture and other commonly used algorithms and architectures.

DCT/IDCT	Jeong et al. [24]	Gong et al. [25]	Dimitrov et al. [26]	Alam et al. [27]	Hsiao and Tseng [28]	This Work 2011
Computation complexity	$O(N - \log_2 N + 1)$	$O(\log N)$	$O(1)$			
Hardware complexity	$O(2N)$	$O(2N)$	$O(2N)$	$O(2N)$	$O(N \log N)$	$O(\frac{4^{\log_2 N - 1} - 1}{3})$
Pipelability	no	no	no	no	no	good
Scalability	poor	poor	poor	poor	good	better

6. Conclusion

With the advantages of the subband decomposition of a signal, a high-efficiency algorithm with pipelined stages has been proposed for fast DCT/IDCT computations. It is noted that the proposed DCT/IDCT algorithm not only simplifies computation complexity but also improves system performance. The PSNR and system complexity of the proposed algorithm is better than those of the previous algorithms [33–36]. Table 2 shows comparisons between the proposed algorithm and architecture and other commonly used algorithms and architectures [24–28]. Thus, the proposed subband-based DCT/IDCT algorithm is suitable for the real-time signal processing applications. The proposed DCT/IDCT processor provides both high throughput and low gate count and has been applied to various images with great satisfactions.

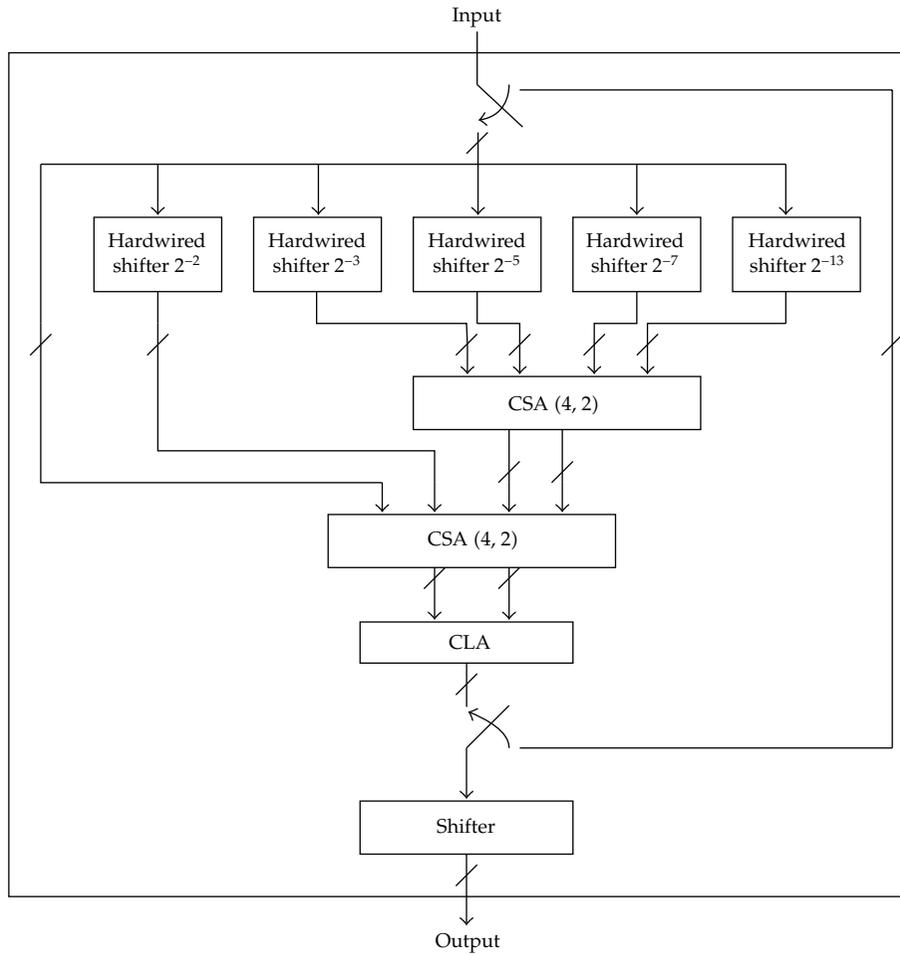


Figure 16: The modified hardwired shifter (MHS) for multiplication by \sqrt{n}/n (where $n = 8, 16, 32, 64$) using the Booth recoded algorithm.

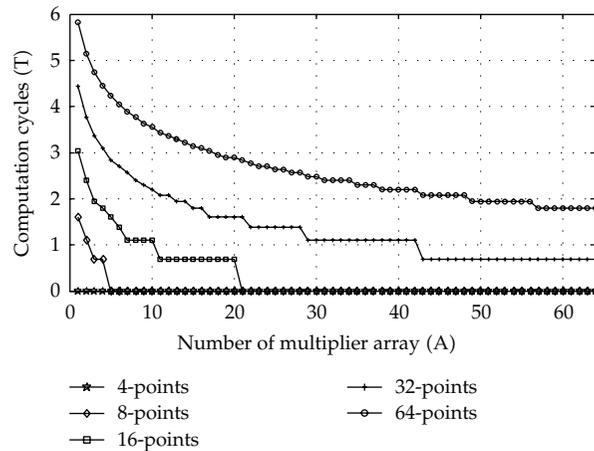


Figure 17: Log plot of number of multiplier array versus computation cycles.

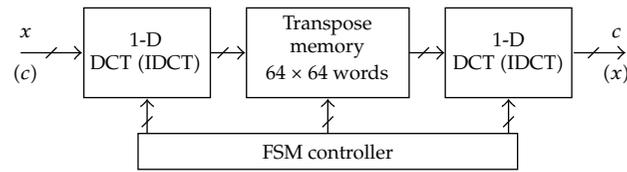


Figure 18: The proposed architecture of 2-D DCT and IDCT (FSM: finite state machine).

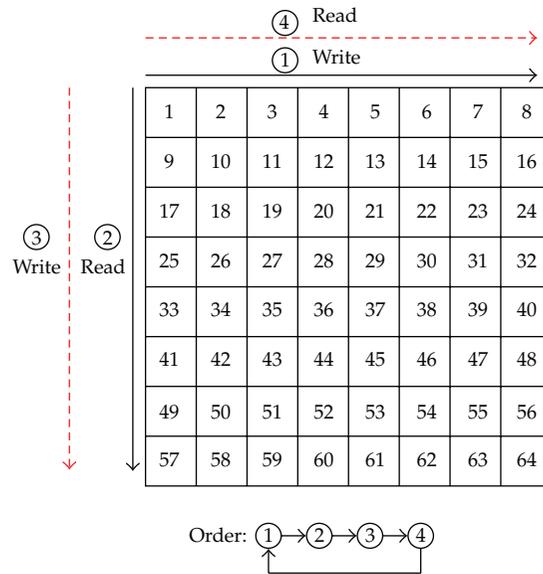


Figure 19: Data read and written timing diagram for 8 x 8 DCT/IDCT system.

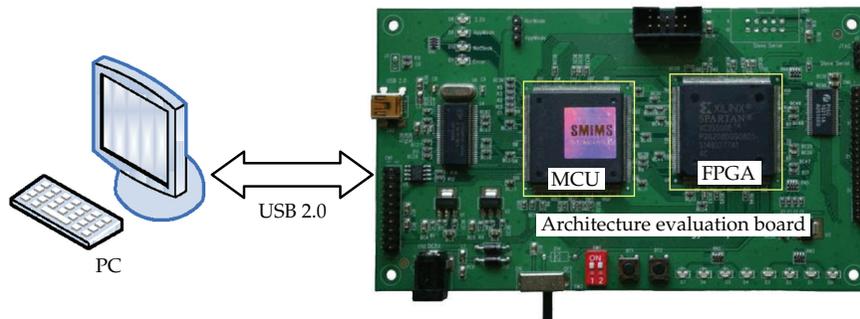


Figure 20: Block diagram and circuit board of the architecture development and verification platform.

Acknowledgment

The National Science Council of Taiwan, under Grants NSC98-2221-E-216-037 and NSC99-2221-E-239-034 supported this work.

References

- [1] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. 23, pp. 90–93, 1974.
- [2] A. K. Jain, "A sinusoidal family of unitary transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 4, pp. 356–365, 1979.
- [3] S. Y. Chen and Y. F. Li, "Determination of stripe edge blurring for depth sensing," *IEEE Sensors Journal*, vol. 11, no. 2, Article ID 5585653, pp. 389–390, 2011.
- [4] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3 part 1, pp. 480–487, 2011.
- [5] S. Y. Chen, Y. F. Li, and J. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
- [6] 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, 2009.
- [7] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [8] M. Li and S. C. Lim, "Modeling network traffic using generalized Cauchy process," *Physica A*, vol. 387, no. 11, pp. 2584–2594, 2008.
- [9] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.
- [10] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.
- [11] M. Li, "Generation of teletraffic of generalized Cauchy type," *Physica Scripta*, vol. 81, no. 2, Article ID 025007, p. 10, 2010.
- [12] 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.
- [13] E. G. Bakhoun and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.
- [14] E. G. Bakhoun 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.
- [15] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press, New York, NY, USA, 1990.
- [16] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*, kluwer academic publishers, Norwell, Mass, USA, 2nd edition, 1997.
- [17] W. H. Chen, C. H. Smith, and S. C. Fralick, "A fast computational algorithm for the discrete cosine transform," *IEEE Transactions on Communications*, vol. 25, no. 9, pp. 1004–1009, 1977.
- [18] B. G. Lee, "A new algorithm to compute the discrete cosine transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1243–1245, 1984.
- [19] H. S. Hou, "A fast recursive algorithm for computing the discrete cosine transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1455–1461, 1987.
- [20] C. Loeffler, A. Ligtenberg, and G. S. Moschytz, "Practical fast 1-D DCT algorithms with 11 multiplications," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '89)*, vol. 2, pp. 988–991, February 1989.
- [21] E. Feig and S. Winograd, "Fast algorithms for the discrete cosine transform," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2174–2193, 1992.
- [22] T. Y. Sung, Y. S. Shieh, and H. C. Hsin, "An efficient VLSI linear array for DCT/IDCT using subband decomposition algorithm," *Mathematical Problems in Engineering*, vol. 2010, Article ID 185398, 21 pages, 2010.
- [23] H. Huang, T. Y. Sung, and Y. S. Shieh, "A novel VLSI linear array for 2-D DCT/IDCT," in *Proceedings of the 3rd International Congress on Image and Signal Processing (CISP '10)*, vol. 8, pp. 3686–3690, Yantai, China, October 2010.
- [24] H. Jeong, J. Kim, and W. K. Cho, "Low-power multiplierless DCT architecture using image data correlation," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 262–267, 2004.
- [25] D. Gong, Y. He, and Z. Cao, "New cost-effective VLSI implementation of a 2-D discrete cosine transform and its inverse," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 4, pp. 405–415, 2004.

- [26] V. Dimitrov, K. Wahid, and G. Jullien, "Multiplication-free 8×8 2D DCT architecture using algebraic integer encoding," *Electronics Letters*, vol. 40, no. 20, pp. 1310–1311, 2004.
- [27] M. Alam, W. Badawy, and G. Jullien, "A new time distributed DCT architecture for MPEG-4 hardware reference model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 5, pp. 726–730, 2005.
- [28] S. F. Hsiao and J. M. Tseng, "New matrix formulation for two-dimensional DCT/IDCT computation and its distributed-memory VLSI implementation," *IEE Proceedings: Vision, Image and Signal Processing*, vol. 149, no. 2, pp. 97–107, 2002.
- [29] T. Y. Sung, "Memory-efficient and high-performance 2-D DCT and IDCT processors based on CORDIC rotation," *WSEAS Transactions on Electronics*, vol. 3, no. 12, pp. 565–574, 2006.
- [30] SMIMS[®] Technology Corp., <http://www.smims.com>.
- [31] ModelSim, <http://model.com/content/modelsim-de-simulation-and-verification>.
- [32] Xilinx[®] FPGA products, <http://www.xilinx.com/products/>.
- [33] A. Buemi, A. Bruna, M. Mancuso, A. Capra, and G. Spampinato, "Chroma noise reduction in DCT domain using soft-thresholding," *EURASIP Journal on Image and Video Processing*, vol. 2010, Article ID 323180, 13 pages, 2010.
- [34] 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.
- [35] T. Bianchi, A. Piva, and M. Barni, "Encrypted domain DCT based on homomorphic cryptosystems," *EURASIP Journal on Information Security*, vol. 2009, Article ID 716357, 12 pages, 2009.
- [36] F. S. Al-Kamali, M. I. Dessouky, B. M. Sallam, F. Shawki, and F. E. A. El-Samie, "Carrier frequency offsets problem in DCT-SC-FDMA system: investigation and compensation," *ISRN Communications and Networking*, vol. 2011, Article ID 842093, 7 pages, 2011.

Research Article

A Novel SFP Extracting/Tracking Framework for Rigid Microstructure Measurement

Sheng Liu,^{1,2} Ting Fang,¹ and Zichen Chen²

¹ College of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China

² Department of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China

Correspondence should be addressed to Sheng Liu, edliu@hotmail.com

Received 4 May 2011; Accepted 15 June 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Sheng Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

3D measurement and reconstruction of rigid microstructures rely on the precise matches of structural feature points (SFPs) between microscopic images. However, most of the existing algorithms fail in extracting and tracking at microscale due to the poor quality of the microscopic images. This paper presents a novel framework for extracting and matching SFPs in microscopic images under two stereo microscopic imaging modes in our system, that is, fixed-positioning stereo and continuous-rotational stereo modes, respectively. A 4-DOF (degree of freedom) micro visual measurement system is developed for 3D projective structural measurement of rigid microstructures using the SFPs obtained from microscopic images by the proposed framework. Under the fixed-positioning stereo mode, a similarity-pictorial structure algorithm is designed to preserve the illumination invariance in SFPs matching, while a method based on particle filter with affine transformation is developed for accurate tracking of multiple SFPs in image sequence under the continuous-rotational stereo mode. The experimental results demonstrate that the problems of visual distortion, illumination variability, and irregular motion estimation in micro visual measurement process can be effectively resolved by the proposed framework.

1. Introduction

Micro stereovision technology makes it possible to achieve 2D/3D information extraction and 3D reconstruction. It has been extensively used in micromanipulation, microassembly, microrobot navigation and bioengineering, and so forth. 3D reconstruction of the small objects in microscopic image is a challenging work according to the small sizes. Existing methods as structured-light-based 3D reconstruction [1, 2], and binocular stereo methods [3]. are not suitable directly for micro applications and then definitely need to be modified. For rigid 3D microstructures, the accurate matching of structural feature points between microscopic images is one of the most important steps in micro stereovision computation. In this paper we define the structural feature points (SFPs) as those key points that can

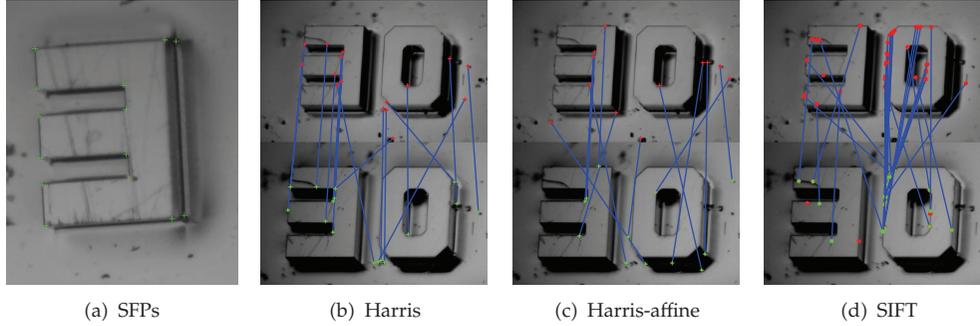


Figure 1: Structural feature points on a rigid 3D microstructure (a) and the failed detecting and matching results of some popular methods (b), (c), and (d). These results failed to locate precisely on the positions of SFPs.

fully represent the rigid microstructures, which are those intersection points of 3 planes on the micro 3D structure (Figure 1(a)). Although the vision method based on tracking of feature points plays an important role in normal-scale computer vision applications as 3D reconstruction, segmentation, and recognition, there is little research focused on the microscale applications except for medical applications. This is due to poor contrast and worse quality of microscopic image mainly resulting from the imaging complexity of optical microscope. The complex structure of microscopic lens system [4] requires the variety of optical elements that can introduce a wide array of image distortions.

In this paper we present a framework for extracting and matching SFPs in microscopic images for dealing with *fixed-positioning and continuous-rotational* stereo imaging modes, respectively. We believe that the proposed framework perfectly plays an important role in 3D reconstruction based on microscopic image.

First, since there exist some specific drawbacks in microscopic images as blurred edges, geometrical distortions, serious dispersion, and disturbance by noises (especially coming from illumination changes), which result in more troubles for SFPs' detection and matching, many popular key point detecting and matching methods [5–9] failed to achieve similar accuracy in microscopic images (Figures 1(b)–1(d)) in comparison with those results obtained when employed on normal-scale images. New methods that are suitable for SFPs' extracting and matching in microscopic images are highly required. We develop a novel illumination-invariant method based on similarity-pictorial structure algorithm (similarity-PS) to solve the SFPs' extracting and matching problem in the *fixed-positioning stereo mode*.

Second, existing feature point tracking algorithms for continuous image sequences are usually classified as methods based on template, motion parameters, and color patch [10], none of them meet the needs of SFPs' tracking in microscale images. The representative research achievements include: continuously adaptive mean shift (Camshift) [11] taking color histogram as object mode in rich-color images for the tracking tasks, which performs poorly when tracking the feature points in complicated background with areas of similar color. Yao and Chellappa first designed a probabilistic data association filter with extended Kalman filter (EKF) [12] to estimate the rotational motion and continuously track the feature points in frames, which effectively resolved occlusion problem. However, the real location must be predicted by probability analysis during arbitrary moving process of the object.

Buchanan proposed a combining local and global motion models with Kanade-Lucas-Tomasi (KLT) tracker to accurately track multiple feature points of nonrigid moving objects

[13]. But if motion predictions cannot be made for the subsequences consisting of the initial frames, this strategy must fail. Kwon et al. presented a template tracking approach based on applying the particle filtering (PF) algorithm on the affine group [14, 15], which can accurately track a large single object, but it performs well only in single template tracking. We extended their method to the multiple-points tracking case in the proposed monocular microvision system in the *continuous-rotational stereo mode*. Since the images of the measuring rigid micro structure keep fixed spatial relationship of global structures and local features affine invariabilities during the rotating process, the tracking problem is greatly simplified with affine transformation and covariance descriptor in our framework.

2. Proposed Framework

2.1. Imaging Modes of Proposed Micro Stereoscopic Measurement System

We developed a 4-DOF micro stereoscopic measurement system for 3D micro objects measurement. The system consists of a 4-DOF stage, a fixed-mounted illumination and a SLM-based optical imaging system. A monocular tube microscope is used in the system for reducing the imaging complexity. In our system the stereoscopic imaging relationship can be realized in two modes.

(i) Fixed-Positioning Stereo Mode (Mode 1)

The rotational transform by a fixed rotation angle is performed, followed by a tilting motion. Images are captured at the end of each movement. *Problem:* The changes of illumination direction bring huge contrast and intensity changes to the microscopic images captured at different position.

(ii) Continuous-Rotational Stereo Mode (Mode 2)

Tilting movement is performed firstly, followed by a continuously rotational movement. Image sequences are captured during the rotation process. *Problem:* The motion blur caused by continuous rotational motion will decrease the quality of microscopic image sequences.

2.2. Similarity-PS Method for SFPs Matching of Mode 1

2.2.1. Pictorial Structure Method

A PS method is represented by a collection of parts which has spatial relationships between certain pairs. The PS model can be expressed by a graph $G = (V, E)$, where the vertices $V = \{v_1, v_2, v_3, \dots, v_n\}$ correspond to the parts and $\{v_i, v_j\} \in E$ present the edge for each pair of connected parts v_i and v_j . An object is expressed by a configuration $L = (l_1, l_2, \dots, l_n)$ where l_i represents the location for each part v_i . For each part v_i , the appearance match cost function $a_i(I, l)$ represents how well the part matches the image I when placed at location l . A simple pixel template matching is used for this cost function in [16]. The connections between the locations of parts present the structure match cost. The cost function $t_{ij}(v_i, v_j)$ represents how well the locations l_i of v_i and l_j of v_j agree with the object model. Therefore,

the cost function L^* for PS includes 2 parts (the appearance cost function and structure cost function).

$$L^* = \arg \min \left(\sum_{v_i \in V} a_i(l_i, l_j) + \sum_{(v_i, v_j) \in E} t_{ij}(l_i, l_j) \right). \quad (2.1)$$

The best match of SFPs can be obtained by minimizing L^* .

2.2.2. The Local Self-Similarity Descriptor

Self-similarity descriptor is proposed by Buchanan and Fitzgibbon [13].

Figure 2 illustrates the procedure for generating the self-similarity descriptor d_q centered at q with a large image. q is a pixel in the input image.

The green square region is a small image patch (typically $5 * 5$, $3 * 3$) centered at q . The larger blue square region is a big image region (typically $30 * 30$, $40 * 40$) centered with q , too. First, the small image patch is compared with a larger image region using sum of square differences (SSDs). The CIE $L * a * b$ color space transforming is needed for color images. Second, the correlation surface is normalized to eliminate illumination influences. Finally, the normalized correlation surface is transformed into a "correlation surface" $S_q(x, y)$

$$S_q(x, y) = \exp \left(- \frac{SSD_q(x, y)}{\max(\text{var}_{\text{noise}}, \text{var}_{\text{auto}}(q))} \right), \quad (2.2)$$

where $SSD_q(x, y)$ is the normalized correlation surface and $\text{var}_{\text{noise}}$ is a constant number that corresponds to acceptable photometric variations (in color, illumination or due to noise, which is 150 in the paper). $\text{var}_{\text{auto}}(q)$ is the maximal variance of the difference of all patches within a very small neighborhood of q (of radius 1) relative to the patch centered at q .

The correlation surface $SSD_q(x, y)$ is then transformed into log-polar coordinates centered at q and portioned into $20 * 4$ bins ($m = 20$ angles, $n = 4$ radius). We choose the maximal value in every bin (it can help the descriptor to adapt to nonrigid deformation). We choose all the maximal values to form an $m * n$ vector as a self-similarity descriptor centered at q . Finally, this descriptor vector is normalized to the range $[0 \cdots 1]$ by linearly stretching its values.

2.2.3. Similarity-PS Algorithm

We introduce the local "self-similarity" descriptor-matching approach for SFPs' detection into a simplified PS model. This subsection covers 3 main steps as follows

Extraction of the "Template" Description T

Manually marked SFPs' coordinates on the micro structure are used for the training process. We describe the marked points by self-similarity descriptors. For every point, the average values of all the examples of descriptor as a trained descriptor are calculated, acting as the appearance description for the model.

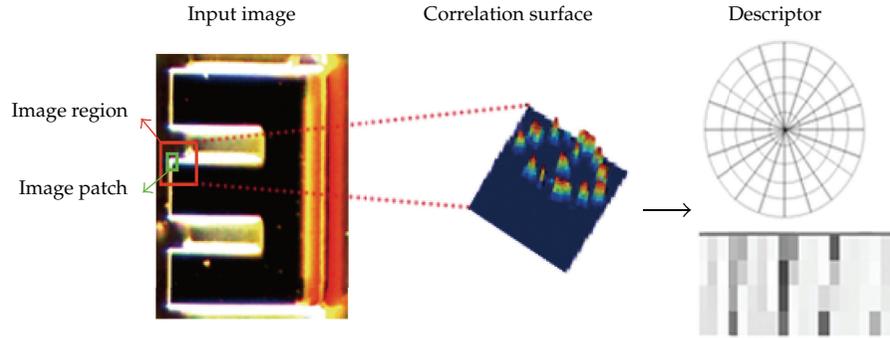


Figure 2: Extract the local self-similarity.

Dense Computation of the Local Self-Similarity Descriptor

These descriptors d_q are computed throughout the tested image I with 2 pixels apart from each other in this paper. The higher precision will be obtained if the searching process covers every pixel.

Detection of Similar Descriptors of T within I

The region centered at SFPs is chosen as the interesting region in I , which has the smallest weighted Euclidean distance between self-similarity descriptor vector within I and the one from the training descriptor. The coordinates of all interest points in I are recorded as the locations for the candidate key points. We usually find out many candidate key points for one marked point. In our experiment there are 200 points being chosen (but it varies for different templates), and their Euclidean distances between the candidate key points descriptors and the trained descriptors are recorded as $a_i(I, l_j)$ which are linearly normalized to the range of $[0 \cdots 1]$. Therefore the appearance cost function $a_i(I, l_j)$ is determined. We substituted the appearance model in (2.1) with the obtained self-similarity model; then, the best-matching SFPs for the PS model can be obtained by minimizing L^* .

Since the microscopic images are not always of the same size, to satisfy the scale invariance, we calculate the self-similarity descriptors at multiple scales both on the patterns and testing images. Moreover, the scale-invariant structures are obtained by multiplying a scale factor to all the mean and variance value of structural distances of SFPs.

2.3. PF-Based Multiple SFPs Tracking of Mode 2

2.3.1. The Affine Motion of Tracking Points

A single tracked point $b_t^{(i)} = (x_t^{(i)}, y_t^{(i)})^T$ ($i = 1, \dots, k, t = 1, \dots, N$, representing the number of tracked points and frames, resp.) cannot satisfy the accurate tracking in micro sequences

therefore it is necessary to take advantage of templates. Initializing multiple points with a set of given center coordinates $b_t = (x_t^{(1)}, y_t^{(1)}, \dots, x_t^{(i)}, y_t^{(i)})^T$, a set of small regions

point templates $T(X_t^{(i)})$ are denoted as a group of windows which surround each tracked point for describing their main characteristics. The coordinates of each $T(X_t^{(i)})$ are obtained via multiplication, the homogeneous coordinates with affine transformation matrix $X_t^{(i)}$, so the most important step in tracking process is to calculate the state $X_t^{(i)}$, for all point templates in each frame. An affine motion can be divided into a translation $M_t^{(i)}$ and a rotation R_t all the points of a rigid object are rotating with same angular velocity. The translation vectors are denoted by $M_t^{(i)} = (Mx_t^{(1)}, My_t^{(1)}, \dots, Mx_t^{(i)}, My_t^{(i)})^T$, and the affine motion action of a point can be written as $\ell(b_t^{(i)}) = R_t b_t^{(i)} + M_t^{(i)}$ and

$$\begin{bmatrix} R_t & M_t^{(i)} \\ 0 & 1 \end{bmatrix} = \exp \begin{pmatrix} \Gamma & \gamma \\ 0 & 0 \end{pmatrix}, \quad (2.3)$$

where l is a linear function representing the affine motion of points $b_t^{(i)}$, R_t is an invertible 2×2 rotation matrix, and $R_t \in R^2$, $M_t^{(i)}$ is a $2 \times i$ translation vector $M_t^{(i)} \in R^2$. The second form of (2.1) using the exponential map is for the convenience of denoting the affine transformation of the imaging process.

2.3.2. Tracking by Particle Filter with Affine Motion

State estimate and sample for tracked points in each frame: the efficiency of the particle filter tracking algorithm mainly relies on the importance of random sampling and calculate the weights $w_t^{(i,j)}$. The measurement likelihood $p(y_t^{(i)} | X_t^{(i)})$ is independent of the current state particles $X_{0:t}^{(i,j)}$. The measurement state equation can be expressed in the discrete setting as

$$y(X_t^{(i)}) = f \left(X_{t-1}^{(i)} \exp \left(a_i \log \left((X_{t-2}^{(i)})^{-1} X_{t-1}^{(i)} \right) + \sum_{m=1}^3 \xi_m \varepsilon \tau_t \right) \right) + v_t, \quad (2.4)$$

where v_t is a measurement zero-mean Gaussian noise. a_i is the AR process parameter. τ_t is zero-mean Gaussian noise. $\varepsilon = (1/12)^{1/2}$ represents obtaining 12 frames per second. The affine transformation basis elements $\xi_m (m = 1, \dots, 3)$ denoting the templates have produced translation, shearing or scaling transformation.

It is necessary to approximate the particles $\{R_t^{(i,1)}, \dots, R_t^{(i,G)}\}$ and resample according to their weights at every timestep. To optimize the computational procedure and avoid directly calculating, all the resample particles are expected to be quite similar to each other. Denote $\bar{M}_t^{(i)}$ as the arithmetic mean of $M_t^{(i,j)}$, the sample mean of $X_t^{(i,j)}$ can be approximated as

$$\begin{bmatrix} \bar{R}_t^{(i)} & \bar{M}_t^{(i)} \\ 0 & 1 \end{bmatrix} = \exp \begin{pmatrix} \bar{\Gamma} & \bar{\gamma} \\ 0 & 0 \end{pmatrix} = \begin{bmatrix} R_{t,\max}^{(i)} \exp \left(\frac{1}{G} \sum_{j=1}^G \log \left(R_{t,\max}^{(i)-1} R_t^{(i,j)} \right) \right) & \bar{M}_t^{(i)} \\ 0 & 1 \end{bmatrix}. \quad (2.5)$$

Calculate covariance with descriptor for the tracked point templates: the spatial structure constraint vector $h = (x_t^{(i)}, y_t^{(i)}, b_t^{(i,d)}, b_t^{(i,\theta)}, I(X_t^{(i)}), I_x, I_y, \tan^{-1}(I_x/I_y), I_{xx}, I_{yy})^T$ is added in our

method for minimizing the influence of illumination, distortion, motion-blur, and noise interference. Here $b_t^{(i,d)}$, $b_t^{(i,\theta)}$ represent the distance and the angle between origin (or polar axis) and each tracked pixel in polar coordinate system. $I(X_t^{(i)})$ denotes the image pixel intensity. I_x , I_y , I_{xx} , I_{yy} represent the first- and second-order image derivatives in the Cartesian coordinates system [14, 16]. s is the size of template window \hat{h} is the mean value of $h(X_t^{(i)})$. The covariance descriptor S of the point templates patches can be given as

$$S_{X_t^{(i)}} = \frac{1}{s-1} \sum_{p=1}^s \left(h_{(X_t^{(i)})_p} - \hat{h} \right) \left(h_{(X_t^{(i)})_p} - \hat{h} \right)^T. \quad (2.6)$$

Measure relative distance: for ensuring the covariance descriptors are changing successively, it is necessary to collect image covariance descriptors and calculate the principal eigenvector and the geodesic distance between two group elements $\{S(X_t^{(i)}), \bar{S}\}$ and $\{I(X_t^{(i)}), \bar{I}\}$. The measurement function is defined using the distance-from-feature space, distance-in-feature space, and similarity comparison purposes [18] then the measurement equation in [19] can be more explicitly expressed as

$$\mathbf{y}_t^{(i)} = \begin{bmatrix} \sqrt{\left\| \log S_{X_t^{(i)}} - \log(\bar{S}) \right\|^2 - \sum_{n=1}^M c_n^2} \\ \sqrt{\sum_{n=1}^M \frac{c_n^2}{\rho_n}} \\ \left\| I_{X_t^{(i)}} - \bar{I} \right\| \end{bmatrix} + \mathbf{v}_t^{(i)}, \quad (2.7)$$

where c_n is the projection coefficient and ρ_n is the eigenvalues corresponding to principal eigenvector, and the two parameters are used to calculate the distance-in-feature space, and \bar{I} represents the point mean intensity. The measurement likelihood is described as

$$p\left(\mathbf{y}_t^{(i)} \mid X_t^{(i)}\right) \propto \exp\left(\mathbf{y}_t^{(i)T} R^{-1} \mathbf{y}_t^{(i)}\right), \quad (2.8)$$

where R is the covariance of zero-mean Gaussian noise \mathbf{v}_t . When the measurement likelihood $p(\mathbf{y}_t^{(i)} \mid X_t^{(i)})$ is gained, we can calculate and normalize the importance weights for the tracked points and then realize multiple SFPs' tracking in long microscopic sequences.

3. Experimental Results and Discussion

The SFPs' detecting/matching experiments are implemented on a standard microchip with 3D microstructures. Some of the results by both PS algorithm and proposed method are shown in Figure 3, while the contrast accumulated error measured by the Euclidean distance in pixels of the tracking points to the ground truth positions is shown in Figure 4. It proves that our method enhances the correct detecting rate of SFPs' for images under large changed illumination in Mode 1. The reason for this improvement is that the changed gray value is

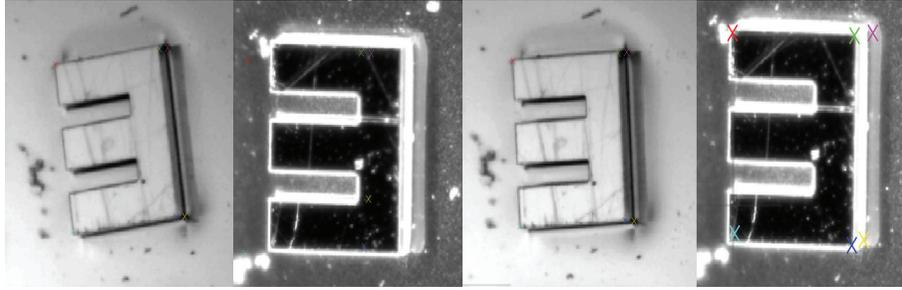


Figure 3: Tracking results in Mode 1 by PS algorithm (left two images) and our proposed method (right two images). Each image pairs are corresponding to different illumination conditions.

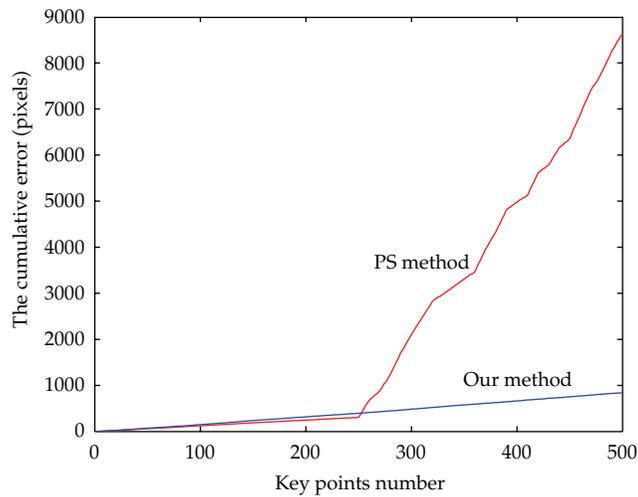


Figure 4: Comparing the accumulated error distribution curves of PS method and Similarity-PS method in the Mode 1. Points number from 0 to 250 are tracking points on the images which are captured under different illumination condition with the rest ones.

largely caused by the varying illumination while the local structure keeps static. Therefore, it is a good choice to introduce the local self-similarity descriptors into PS for our application.

We also provide some SFPs' tracking results in Mode 2 by our proposed method and KLT for comparison, as shown in Figure 5. The tracking results by the proposed method obviously show a higher localization accuracy of the SFPs' in long micro image sequence.

For further demonstrating the effectiveness of our method, we present some of the 3D projective reconstruction results based on the tracking results in Figure 6. Obviously the results of the proposed method perfectly reconstruct the 3D structure from the microscopic sequences, while the KLT method failed.

4. Conclusions

This paper developed a framework of SFPs' extracting and matching in microscopic images for 3D micro stereoscopic measurement in two stereo imaging modes. The proposed SFPs' tracking framework ensures the illumination invariance and the robustness in the

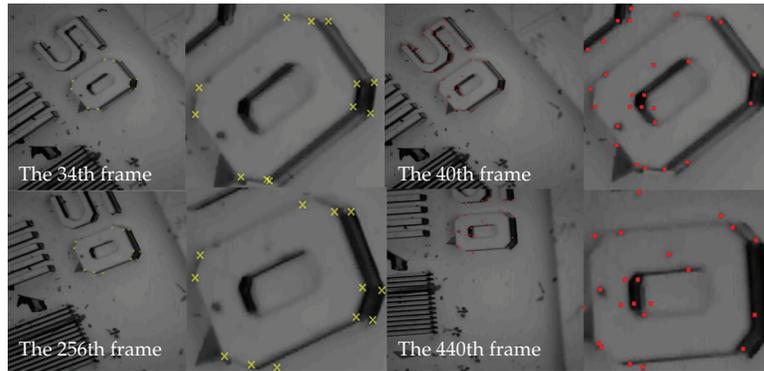


Figure 5: The first and second columns show the accurate tracking results of proposed PF with affine transformation, while the third and fourth columns demonstrate the tracking results of KLT algorithm with lost or dislocated tracks.

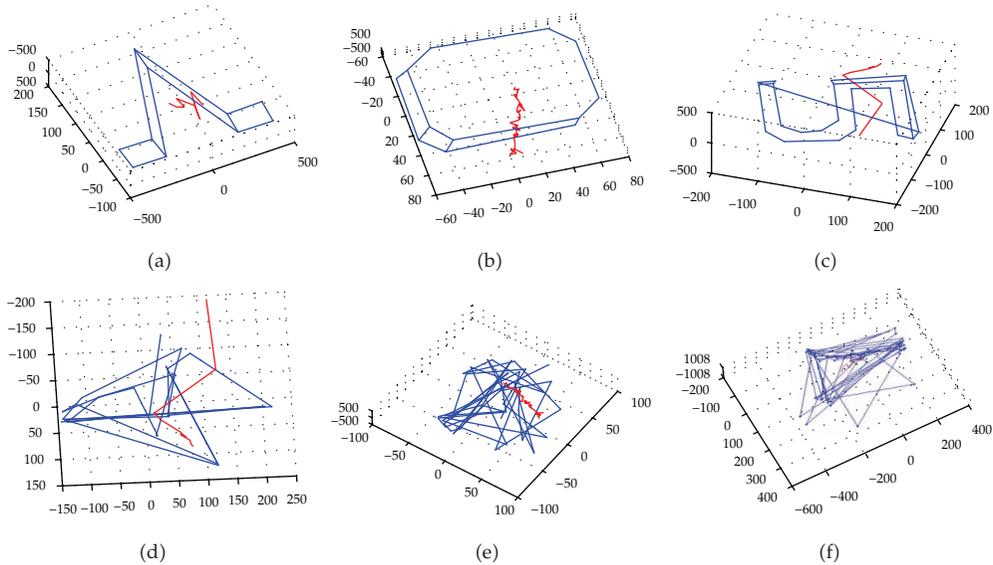


Figure 6: 3D projective reconstruction results by using multiple-point tracking results by PF with affine transformation (the first row) and multiple-point tracking results by KLT algorithm (the second row) respectively. The irregular red curves represent the camera relative moving trajectory. Obviously our method outperforms KLT greatly in the application of 3D reconstruction from microscopic sequences.

fixed-positioning stereo mode and continuous-rotational stereo mode, respectively. The effectiveness of our tracking framework has been empirically verified in visual 3D projective reconstruction with microscopic images. In Mode 2 of our system, there is an inevitable tracking error caused by motion blur. Therefore we plan to use the method described in [20] to deal with this problem in our future work. Our future research will also focus on the micro visual measurement planning (similar to the method described in [21]) and optimal 3D micro model representation [22].

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC-60870002, 60802087, 60573123, 60605013), NCET, and the Science and Technology Department of Zhejiang Province (2009C21008, 2010R10006, Y1090592, Y1110688).

References

- [1] S. Y. Chen, Y. F. Li, and J. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
- [2] 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.
- [3] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, pp. 15–18, August 2006.
- [4] H. W. Schreier, D. Garcia, and M. A. Sutton, "Advances in light microscope stereo vision," *Experimental Mechanics*, vol. 44, no. 3, pp. 278–288, 2004.
- [5] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the Alvey Vision Conference*, pp. 189–192, 1988.
- [6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [7] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [8] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.
- [9] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, pp. II506–II513, July 2004.
- [10] Z. Luo, Y. Zhuang, Y. Pan, and F. Liu, "Feature tracking algorithms based on two cameras," *Journal of Computer-Aided Design and Computer Graphics*, vol. 14, no. 7, pp. 646–650, 2002.
- [11] L. Ye and Y. Wang, "Real-time tracking of the shoot point from light pen based on camshift," in *Proceedings of the 1st International Conference on Intelligent Networks and Intelligent Systems (ICINIS '08)*, pp. 560–564, November 2008.
- [12] Y.-S. Yao and R. Chellappa, "Dynamic feature point tracking in an image sequence (EKF)," in *Proceedings of the International Conference on Pattern Recognition (ICPR '94)*, vol. 12, pp. 654–657, 1994.
- [13] A. Buchanan and A. Fitzgibbon, "Combining local and global motion models for feature point tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.
- [14] J. Kwon, K. M. Lee, and F. C. Park, "Visual tracking via geometric particle filtering on the affine group with optimal importance functions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 991–998, June 2009.
- [15] J. Kwon and F. C. Park, "Visual tracking via particle filtering on the affine group," *International Journal of Robotics Research*, vol. 29, no. 2-3, pp. 198–217, 2010.
- [16] 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.
- [17] X. Tan, F. Song, Z. H. Zhou, and S. Chen, "Enhanced pictorial structures for precise eye localization under uncontrolled conditions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 1621–1628, Miami, Fla, USA, June 2009.
- [18] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: a unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.
- [19] J. Kwon and F. C. Park, "Visual tracking via particle filtering on the affine group," *International Journal of Robotics Research*, vol. 29, no. 2-3, pp. 198–217, 2010.
- [20] S. Y. Chen and Y. F. Li, "Determination of stripe edge blurring for depth sensing," *IEEE Sensors Journal*, vol. 11, no. 2, Article ID 5585653, pp. 389–390, 2011.

- [21] 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.
- [22] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 480–487, 2011.

Review Article

Markov Models for Image Labeling

S. Y. Chen,¹ Hanyang Tong,¹ and Carlo Cattani²

¹ College of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China

² Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (Sa), Italy

Correspondence should be addressed to S. Y. Chen, sy@ieee.org

Received 2 May 2011; Accepted 18 May 2011

Academic Editor: Gani Aldashev

Copyright © 2012 S. Y. Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Markov random field (MRF) is a widely used probabilistic model for expressing interaction of different events. One of the most successful applications is to solve image labeling problems in computer vision. This paper provides a survey of recent advances in this field. We give the background, basic concepts, and fundamental formulation of MRF. Two distinct kinds of discrete optimization methods, that is, belief propagation and graph cut, are discussed. We further focus on the solutions of two classical vision problems, that is, stereo and binary image segmentation using MRF model.

1. Introduction

Many tasks in computer vision and image analysis can be formulated as a labeling problem where the correct label has to be assigned to each pixel or clique. The label of a pixel represents some property in the real scene, such as the same object or the disparity. Such problems can be naturally represented in Markov random field (MRF) model. MRF is firstly introduced into vision by S. Geman and D. Geman [1] in 1984 and has been widely used in both low-level and high-level vision perception in recent years.

Basically, humans understand a scene mainly by using the spatial and visual information which is assimilated through their eyes. Inversely, given an image or images, this information such as boundary or object, mainly based on the contextual constraints, is extremely necessary for scene interpretation. We hope to model the vision problem to capture the full interaction between pixels. On the other hand, due to the sensor noise and complexity of the real world, exact interpretation is rather difficult for computers. As a result, researchers have realized that the solution of vision problems should be solved by using optimization methods. As the most popular models for gridded image-like data, the MRF

provides a series of mathematical theories to find such optimal solutions under the contextual visual information in the images. The context-dependent object in digital images can be modeled in a convenient and consistent way through MRF theory. It is achieved through characterizing mutual influences among such entities using conditional MRF distributions [2]. Besides, the images we captured are always piecewise smooth, which can be encoded as a prior distribution. Thus, we can use the MRF model whose negative log-likelihood is proportional to a robustified measure of image smoothness [3]. Moreover, we may know that some premise or external knowledge of the scene such as the specify object might exist in the environment. With these priors, we can get more reliable understanding of the images.

In the latest two decades, the renaissance of the MRF model in computer vision has begun due to powerful energy minimization algorithms. A lot of inference algorithms have been developed to solve the MRF optimization problems, such as graph cut [4], belief propagation [5], tree-reweighted message passing [6], dual decomposition [7], fusion move [8], iterated conditional modes, and their extensions. In the literature [9], Szeliski et al. gave a set of energy minimization benchmarks and used them to evaluate the solution quality and runtime of several energy minimization algorithms. Felzenszwalb and Zabih [10] reviewed the dynamic programming and graph algorithms and then discussed their applications on computer vision. A review for the linear programming to solve max-sum problem was given in [11]. On the other hand, a framework of learning image priors for MRF was introduced by Roth and Black [12]. Schmidt et al. [13] revisited the generative aspects of MRF and analyzed the quality of common image priors in a fully application-neutral setting. New models based on MRF such as MPF were proposed. It was proved that the convex energy MPF can be used to encourage arbitrary marginal statistics [14]. Some excellent books about MRF models in image analysis such as [2] are also available.

The MRF has been successfully applied to image analysis such as restoration, matting [15], and segmentation, as well as two-dimensional (2D) fields such as stereo matching, super resolution [16], optical flow, image inpainting [17], motion estimation, and 2D-3D registration [18]. The MRF was also used to solve the high-level vision problems such as object classification [19, 20], face analysis [21], face recognition [22], and text recognition [23]. Many optimization problems can be formulated in the MRF, for example, color to gray transformation [24], feature detection scale-selection [25], and so forth. Additionally, Boykov and Funka-Lea [26] presented a survey of various energy-based techniques for binary object segmentation. S. Geman and D. Geman [1] firstly applied the MRF to image restoration. Sun et al. [27] used belief propagation algorithm and combinet it with occlusion to solve the stereo problem. Detry et al. [28] proposed an object representation framework that encodes probabilistic spatial relations between 3D features. Then the authors of [28] organize the proposed framework in the MRF.

In the remainder of the paper, Section 2 gives a sketchy of MRF and related concepts. Section 3 provides two most frequently used inference algorithms for MRF. Section 4 briefly introduces two labeling applications of MRF in low-level vision. Section 5 summarizes the contribution and offers the future works in the topic.

2. Problem Formulation with MRF

As a branch of probability theory [2], MRF is an undirected graphical model in which a set of random variables have a Markov property. To solve a special computer vision problem involving pixel interaction and partially observed information into an optimization problem using MRF model, we will go over the graphical models that visualize the structure of the

5	4	3	4	5
4	2	1	2	4
3	1	0	1	3
4	2	1	2	4
5	4	3	4	5

Figure 1: An example of the 5th-order neighbor system.

probabilistic models using diagrammatic representations. A graph consists of nodes and edges. Each node means an event, and each edge represents the relationship between the events. MRF is used to find the most optimal label configuration.

For a labeling problem, we need to specify a set of nodes, labels, and edges. Without loss of generality, let \mathbf{M} be a set of indexes $\mathbf{M} = \{1, \dots, m\}$, and let $\mathbf{P} = \{p_1, \dots, p_m\}$ be a set of observed nodes. In vision problem, a node often represents the pixel intensity or some other image features. Let \mathbf{L} be a set of labels. \mathbf{L} can be continuous or discrete, but in most cases, all the labels we set are discrete: $\mathbf{L} = \{l_1, \dots, l_m\}$.

As stated above, a label means some quantity of the real scene. The simplest case is binary form where $\mathbf{L} = \{0, 1\}$. Such black and white model is often used to classify the foreground and background regions in the image. In general cases, the label value is more meaningful. For example, in stereo and image restoration problem, larger label value means depth information or lighter pixel intensity. Additionally, \mathbf{L} also can be unordered labels of which the value has no semantic meaning, such as for object classification.

$\mathbf{N} = \{N_i \mid \forall i \in \mathbf{M}\}$ represents the neighbor system to indicate the interrelationship between nodes or the order of MRF. The edges are added between one node P_i and its neighbors N_i . Usually, the neighbor system should satisfy [2] the following:

- (1) a site does not neighbor with itself: $i \notin N_i$,
- (2) the neighboring relationship is mutual: $i \in N_j \Leftrightarrow j \in N_i$.

The definition of the neighbor system is important because it reflects how far the contextual constraint is. For a regular array data, as in Figure 1, the neighbors of i are defined as the set of sites within a radius of \sqrt{r} from i where r is the order of the neighbor system. $N_i = \{i, j \in \mathbf{M} \mid [f(\text{pixel}_i, \text{pixel}_j)]^2 \leq r, i \neq j\}$ where $f(a, b)$ measures the Euclidean distance between a and b .

Another concept here is "clique." A clique is a subset of \mathbf{P} which plays the near role of the neighbor system. However, the nodes in a clique are ordered which means that (P_i, P_j) is a different form (P_j, P_i) . Figure 2 shows some examples of clique types.

Though we could get more static information of the problem domain with larger neighboring system, the computational complexity of the problem will also increase exponentially with the size of neighborhood. In most cases, a 4-neighborhood system is used for simplification and efficiency. MRF is an undirected graph where a set of random variables have a Markov property. In the random field, each random variable p_i in the set \mathbf{P} can take a label from \mathbf{L} . Usually, a mapping function $\mathbf{F} = \{f_1, \dots, f_m\}$ in which $\mathbf{F} : \mathbf{P} \rightarrow \mathbf{L}$

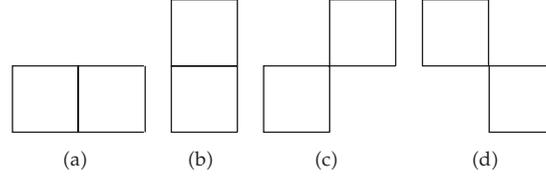


Figure 2: (a) Horizontal pair-site cliques, (b) vertical pair-site cliques, (c) and (d) diagonal pair-site cliques.

can represent for this processing. \mathbf{F} is also called configuration. Denote $\Pr(f_i(p_i) = l_i)$ as the probability of a pixel p_i taking the label l_i . Then the configuration is a joint probability: $\Pr(\mathbf{F}(\mathbf{P}) = \mathbf{L}) = \Pr(f_1(p_1) = l_1, \dots, f_m(p_m) = l_m)$. Note that $\Pr(f_i(p_i) = l_i) > 0$ for all $p_i \in \mathbf{P}$. The Markov property is a basic property if the conditional probability

$$\Pr(p_i | p_{M \setminus i}) = \Pr(p_i | p_{N_i}), \quad (2.1)$$

where $M \setminus i$ means the entire element in \mathbf{M} other than i and N_i is the neighbor system of p_i . The Markov property means that the state probability of one node only depends on its neighbors rather than other remaining nodes. Gibbs random field is a random field in which the probability obeys the Gibbs distribution in the form of

$$\Pr(f) = Z^{-1} \times e^{-(1/T)E(f)}, \quad (2.2)$$

where $Z = \sum_{f \in F} e^{-(1/T)E(f)}$ is a normalizing constant called the partition function, and T is a constant which shall be assumed to be 1 unless otherwise stated. $E(F) = \sum_{c \in C_i} V_c(f)$ $f \in c$, $i \in M$ is the energy function. C is the clique defined on the graph, and $V_c(f)$ is the clique potential function.

Hammersley-Clifford theorem states that if a probability distribution is a positive and satisfies the Markov properties in an undirected graph \mathbf{G} , the distribution is a Gibbs random field. That is, its probability can be factorized over the cliques of the graph. This theorem provides a simple way to calculate the joint probability using the clique potential. According to the Bayes' rule, the posterior distribution for a given set \mathbf{y} and their evidence $p(\mathbf{y} | \mathbf{x})$, combined with a prior $p(\mathbf{x})$ over the unknowns \mathbf{x} , is given by

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}. \quad (2.3)$$

If we do not know the prior information, the maximum likelihood (ML) criterion may be used where $\mathbf{x}^* = \operatorname{argmax} p(\mathbf{y} | \mathbf{x})$. However, sometimes we can still obtain the knowledge about the prior distribution of \mathbf{x} . Thus, the maximum of *a posteriori* (MAP) estimation is the best way to get the optimization where $\mathbf{x}^* = \operatorname{argmax} p(\mathbf{y} | \mathbf{x})p(\mathbf{x})$. Figure 3 illustrates the difference between ML criterion and MAP criterion. MAP probability is one of the most popular statistical criteria for optimization, and in fact, it is the first choice in MRF vision modeling.

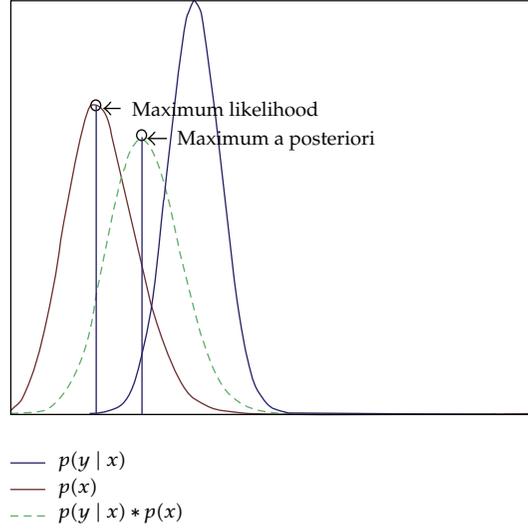


Figure 3: Schematic comparison of ML and MAP methods.

Logarithmize both sides, and then we can obtain the negative posterior log-likelihood

$$-\log p(x | y) = -\log p(y | x) - \log p(x) + \log p(y), \quad (2.4)$$

where $\log p(y)$ is a constant used to make integration of $p(x | y)$ equal to 1. To find the MAP solution, we simply minimize (2.4). Rewrite the clique potential $E(F) = \sum_{c \in C_i} V_c(f)$ $f \in c, i \in M$, then (2.4) becomes an energy function

$$E(x, y) = E_d(x, y) + E_s(x), \quad (2.5)$$

where the $E_d(x, y)$ can be treated as the clique potential whose clique size is 1, and $E_s(x)$ is the remaining clique potential or the observed image prior distribution. In most vision problems, the single-site clique potential is also called unary potential or data energy. Similarly, $E_s(x)$ is called smooth potential or smooth energy. With (2.1), (2.5) can be rewritten as

$$E(f_p, N_p) = E_d(f_p, N_p) + E_s(f_p), \quad (2.6)$$

where $E_s(f_p) = \sum_{q \in N_p \setminus q} E_s(f_p, f_q)$. Therefore,

$$E(F) = \sum_{p \in P} \sum_{q \in N_p \setminus q} E_s(f_p, f_q) + \sum_{p \in P} E_d(f_p, N_p). \quad (2.7)$$

Most vision problems map the minimization of an energy function over an MRF. In some degree, the energy function can be seen as a mathematical representation of the scene and should precisely measure the global quantity of the solution as well as can be easy to find

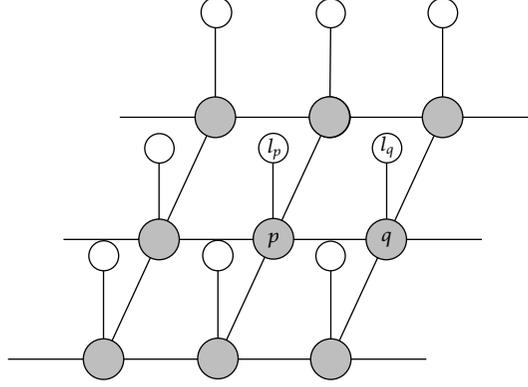


Figure 4: The model for a 4-neighbor MRF (i.e., the pairwise MRF). The dash circles are the observed nodes, and the white circles are the unobserved labels.

the global minimization. When the energy function (2.7) is minimized, the corresponding posteriori $\Pr(x | y)$ gets the maximum.

To solve a specific problem, we need to determine the energy form and the parameters involved. Though there are many types of clique potential functions, there exists a unique normalized potential, called the canonical potential. In literature, the energy function can be expressed as either a parametric form or a nonparametric form [2]. Here, we take the second-order clique potential, for example, which is also called the pairwise model. The pairwise MRF is the most commonly used model in which each node interacts with its adjacent nodes. It is the lowest-order constraint to convey contextual information and is widely used due to its simple form and low computational cost [2]. The pairwise MRF models the statistics of the first derivative in the image structure (Figure 4). The corresponding energy function is

$$E(F) = \sum_{p \in P} \sum_{q \in N_p \setminus q} E_s(f_p, f_q) + \sum_{p \in P} E_d(f_p, N_p) \cdots N = 4. \quad (2.8)$$

Usually, E_d is the local evidence of p taking the label f_p such as the intensity or the color value. Equation (2.8) can be rewritten as

$$E(F) = \sum_{p \in P} \sum_{q \in N_p \setminus q} E_s(f_p, f_q) + \sum_{p \in P} E_d(f_p) \cdots N = 4. \quad (2.9)$$

In the binary MRF case, $E_s(f_p, f_q) = f_p \times f_q$, where $f_p \in \{0, 1\}$. In the multilabel case, Potts model is the most widely used one which can prevent the edges of objects from oversmoothing. Usually, Potts model takes the form

$$V(f_p, f_q) = \begin{cases} 0 \cdots f_p = f_q, \\ \alpha \cdots f_p \neq f_q, \end{cases} \quad (2.10)$$

where α may be a constant or $\alpha = |f_p - f_q|$.

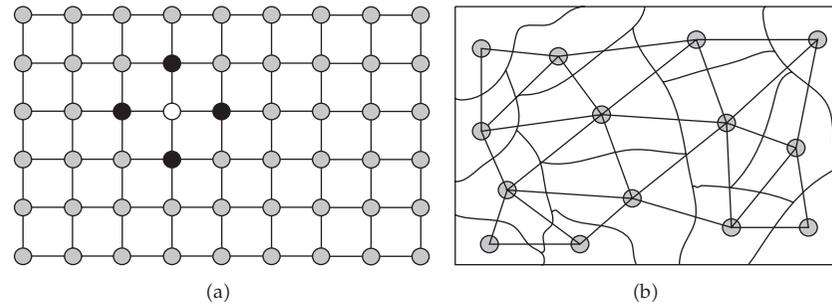


Figure 5: (a) Standard pairwise MRF model with image-grid data, where the circle represents a node or a pixel. The black circles are neighbors of the white one; (b) an example of MRF used in segmentation, where the nodes of neighboring segments are connected by applying Delaunay triangulation method.

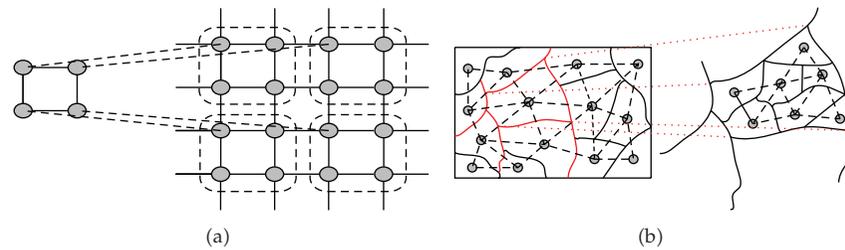


Figure 6: Two examples of hierarchical MRFs. (a) is the most common quadtree where four-grid nodes on the low level are decomposed into one node on the high level. (b) is an irregular pyramid with a large-scale MRF and a corresponding small-scale MRF.

As is illustrated in Figure 5, in the pairwise MRF model, a node is attached to a pixel in the image, while edges are constructed between the node and its four neighborhoods. With such model, the corresponding energy function can be efficiently minimized using many inference algorithms. Other graph structures are also used. For example, in image segmentation, an image is partitioned into several regions. Each region can be regarded as a node, and edges may be constructed between adjacent segmented regions. To make the optimization more efficient, a hierarchical MRF model is used. It mainly uses the pyramid structure and performs in a coarse-to-fine scheme which uses a coarser solution to initialize a finer solution. It is well known that hierarchical methods can significantly improve the convergence rate and reduce the execution time. In [29–31], a regular pyramid downsampling method was applied, while Zitnick and Kang [32] used an irregular pyramid. Figure 6 illustrates an example.

Although most MRFs use the pairwise model due to its simplicity, a scheme of more complex interaction, for example, 8-neighborhood or more numbers of pairwise terms, is also used sometimes. People usually use 26-neighborhood in 3D volumetric images or video analysis. Higher-order clique potentials can capture more complex interactions of random variables. For example, calculating the curvature of an object requires interaction of at least three nodes. Computational time for the clique potential increases exponentially with the size of the clique and poses a difficult energy minimization scenario, which poses a tough question. Recently, there have been many attempts to go beyond pairwise MRF. One approach is to transform the higher-order problem into pairwise problem by adding auxiliary variables.

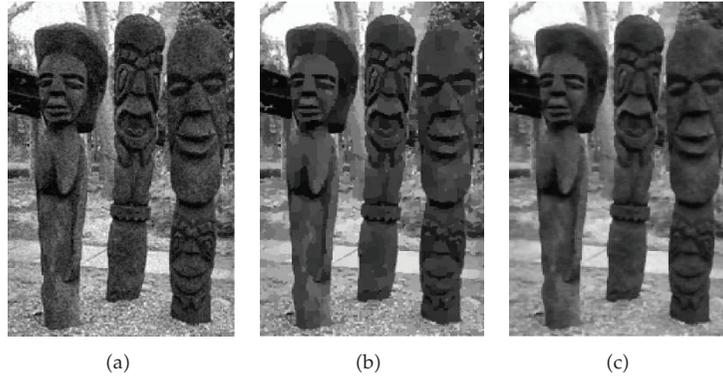


Figure 7: An example of image denoising. (a) is the original noisy image. (c) is the denoising result by BP using pairwise interaction [5]. (b) is the result by BP with the learned 2-by-2 model. The results no longer exhibit any piecewise constancy. Not only edges are preserved, but also smoothly varying regions are preserved better than those with higher-order clique [39].

For instance, Kohli et al. [33] proposed an efficient graph cut method based on special class of higher-order potential, that is, robust P^n Potts model. Rother et al. [34] transformed the minimizing sparse higher-order energy function into an equivalent quadratic minimization problem. Potetz and Lee [35] introduced an efficient belief propagation algorithm where the computational complexity increases linearly with the clique size. Kwon et al. [36] decomposed high-order cliques as hierarchical auxiliary nodes and used hierarchical gradient nodes to reduce the computational complexity. Another way is to perform direct computing using factor graph representation [37]. Kwon et al. [38] proposed a nonrigid registration method using the MRF with a higher-order spatial prior. Experiments show that using high-order potential the performances of image denoising are significantly improved, as is shown in Figure 7.

3. Inference Methods

Over the years, a large number of inference algorithms have been developed, which can be mainly classified into two categories, that is, message passing algorithms such as loopy belief propagation and move making algorithms such as graph cuts. In this section, we briefly introduce two classic inference methods for approximating energy minimums, that is, belief propagation and graph cut.

3.1. Graph Cut

Graph cut (GC) was first applied in computer vision by Greig et al. [40], which describes a large family of MRF inference algorithms based on solving min-cut/max-flow problems. Given a type of computer vision problems which can be formulated in terms of an energy function, GC can get the minimum energy configuration corresponds to the MAP theory.

Suppose that $G = (V, E)$ is a directed graph in which the edge weight is nonnegative, V represents vertices, and E denotes edges. The graph has two special terminals (vertices), that is, the source s and the sink t . A cut $C = (S, T)$ is a partition of V . An s - t cut is a cut that splits

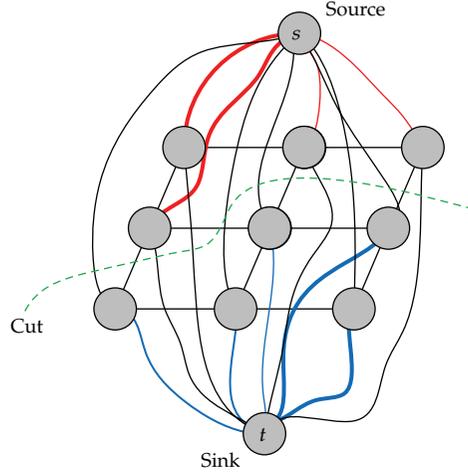


Figure 8: An example of min-cut/max-flow graph cut. The gray circles represent the nodes, and the solid lines are the edges between the nodes. The curve of each “flow” is connected to the source terminal or sink terminal. The potential of flow is measured by the width of line. The dotted line indicates a cut of graph partition.

the source and the sink to be in different subsets where $s \in S$ and $t \in T$. Besides, according to graph theory, the potential of a cut can be measured by the sum of the weights of the edge crossing the cut. To find a cut which can minimize s - t cut problem is equivalent to compute the maximum flow from the source s to the sink t . Maximum flow is the maximum “amount of water” that can be sent from the source to the sink by interpreting graph edges as directed “pipes” with capacities equal to edge weights. As illustrated in Figure 8, the GC algorithm is ideally designed to solve the max-flow problem.

It was reported that GC can obtain the exact solution in the binary label case. In multilabel case, GC requires solving a series of related binary inferences and then obtains the approximated global optimal solutions. Two of the most popular GC algorithms are α - β swap and α -expansion. In the α - β algorithm, a swap move takes some subset of nodes that currently label with α and assign their label with β , and vice versa. The α -expansion algorithm increases the set of nodes taking α by moving it to other nodes. When there is no more swap or expansion move, a local minimum is found. Comparing the two algorithms, α -expansion is more accurate and efficient. Also α -expansion can produce a result with lower energy. However, the condition of α -expansion is more strict. When using the α -expansion, the interaction potential must be metric, that is,

$$V_{ij}(\alpha, \alpha) + V_{ij}(\beta, \beta) \leq V_{ij}(\alpha, \beta) + V_{ij}(\beta, \alpha). \quad (3.1)$$

For α - β swap, it must be semimetric, that is,

$$V_{ij}(\alpha, \alpha) + V_{ij}(\beta, \beta) \leq V_{ij}(\alpha, \beta) + V_{ij}(\beta, \alpha). \quad (3.2)$$

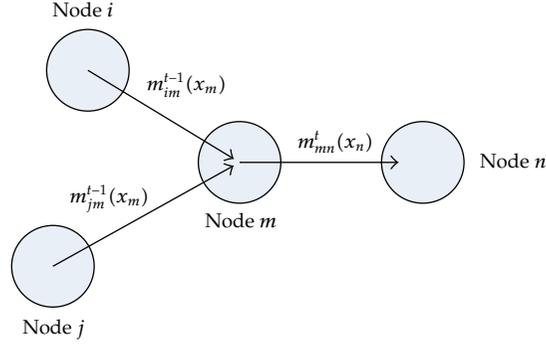


Figure 9: Message passing in BP. $m_{ij}(x_i)^t$ is a message from node i to its neighbor j and indicates what the state should be in node j .

More details about α - β swap and α -expansion can be found in [4]. In addition, Kolmogorov and Roth [41] wrote a survey about graph cut and pointed out that GC can be applied to both submodular and nonsubmodular functions. Other more recent developments in GC include order-preserving GC [42] and combination GC [43].

3.2. Belief Propagation

Belief propagation is a power inference tool originally developed for tree-Bayesian networks [45]. It is recently extended to those “cycle” graphs such as MRF. Although BP can only guarantee convergence with the Bethe free energy in MRF [46], it can obtain reasonable results in practice. In standard BP with pairwise MRF, a variable $m_{ij}(x_j)$ can be treated as a “message” from a node i to its neighbor j which contains the information about what the state of node j should be in. The message is a vector of the same dimension as the number of possible labels. The value of each dimension manifests how this label might be corresponding to the node.

Let $\phi_{ij}(x_i, x_j)$ be the pairwise interaction potential of p_i with p_j , and $\phi_i(x_i, x_j)$ is the “local evidence” of p_i . Usually, the message must be nonnegative. A large value of the message means that the node “believes” the posterior probability of X_j is high. The message updating rule is

$$m_{ij}(x_i)^t = \sum_{x_i} \varphi_i(x_i) \phi_{ij}(x_i, x_j) \prod_{k \in N_i \setminus j} m_{ki}(x_i)^{t-1}, \quad (3.3)$$

where t represents the number of interaction T as showed in Figure 9.

The belief is the product of “local evidence” of the node and all messages send to it

$$b_i(x_i) = k \phi_i(x_i) \sum_{j \in N_i} m_{ij}(x_i). \quad (3.4)$$

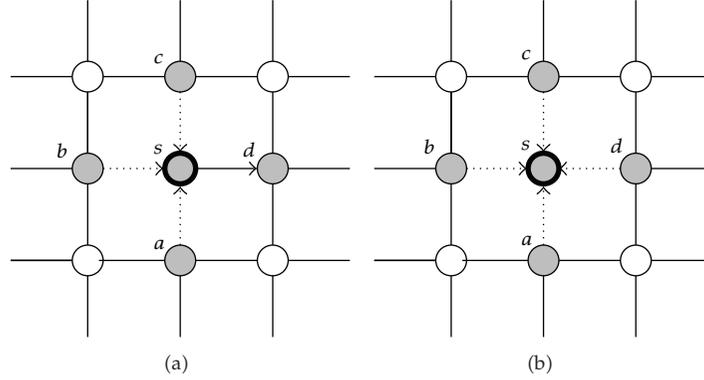


Figure 10: (a) Messages passing from node s to its neighbor u . (b) The belief of node s is calculated according to its neighbors' messages.

The standard BP described above is also called *sum-product* BP. There is another variant BP which is more simple to use, that is, *max-product* (or *max-sum* in log domain). In *max-product* BP, (3.3) and (3.4) are represented as

$$m_{ij}(x_i)^t = \max_{x_i} \left(\varphi_i(x_i) + \phi_{ij}(x_i, x_j) + \sum_{k \in N_i \setminus j} m_{ki}(x_i)^{t-1} \right), \quad (3.5)$$

$$b_i(x_i) = k \left(\sum_{j \in N_i} m_{ij}(x_i) + \varphi_i(x_i) \right). \quad (3.6)$$

The sketch map of this process is illustrated in Figure 10. Several speed-up techniques are attempted, for example, distance transformation, checkerboard updating, and multiscale BP [5], so that the belief propagation can converge efficiently. In another way, Yu et al. [47] used the predictive coding, linear transform coding, and envelope point transform to improve the BP efficiency.

Although BP is an implicitly efficient inference algorithm for MRF with loops, it can only converge to the stationary point of the Bethe approximation of the free energy. Recently, a generalized belief propagation (GBP) algorithm proposed by [48] has received more attention due to its better convergence property against BP. It can converge to a more accurate stationary point of Kikuchi free energy [46]. More details about the GBP algorithm can be found in [48].

BP and graph cut are both good optimal techniques which can find "global" minima over cliques and produce plausible results in practice. A comparison between the two different approaches for stereo vision was described in [49]. GC can get lower energy, but the performance of BP is comparative to GC relative to the ground truth.

In addition to the two typical methods, many other inference algorithms have been proposed in latest few years. Fusion move [8] is proposed for multilabel MRF. By employing QPBO graph cut, the fusion move can efficiently combine two proposal labels in a theoretically sound way, which is in practice often globally optimal. Alahari et al. [50] improved the computational and memory efficiency of algorithms for solving multilabel

energy functions arising from discrete MRF by recycling, reducing, and reusing. Kumar et al. [51] provided an analysis of linear programming relaxation, the quadratic programming relaxation, and the second-order cone programming relaxation to obtain the maximum a posteriori estimate of a general discrete MRF. Komodakis and Tziritis [52] proposed an exemplar-based framework and used priority BP to find MRF solutions. Ishikawa [53] introduced a method to exactly solve a first-order MRF optimization problem in more generality than previous ones. Cho et al. [54] used patch transform representation to manipulate images in the patch domain. The patch transform is posed as a patch assignment problem on an MRF, where each patch should be used only once, and neighboring patches should fit to form a plausible image.

4. Applications

Here, we provide MRF solutions for two typical problems in computer vision, that is, stereo matching and image segmentation. These problems require labeling each pixel with a value to represent the disparity and foreground or background. They can be easily modeled using MRF and solved by energy minimization.

4.1. Stereo Matching

Stereo matching has always been one of the most challenging and fundamental problems in computer vision. Comprehensive research has been done in the last decade [32, 55–58]. A latest evaluation of these various methods can be found in [59]. In the last few years, as is shown in [44], the global methods based on MRF have reached the top performance.

For MAP estimation, let \mathbf{P} be the set of the image pixels in image pair, and let \mathbf{L} be the set of disparity. The initial data cost, which is calculated by the truncated linear transform which is robust to noise or outlier, is defined as

$$D(f_p) = \lambda \cdot \min \left(\sqrt{\sum_{c \in \{L, a, b\}} (\mathbf{I}_c^L(p) - \mathbf{I}_c^R(p - f_p))^2}, T \right), \quad (4.1)$$

where λ is the cost weight which determines the portion of energy that data cost possesses in the whole energy, and T represents the truncating value. The parameters can be set with empirical values from experiments. $\mathbf{I}_c^L(p)$ represents p 's intensity in the left image of channel c . $\mathbf{I}_c^R(p)$ is similarly defined. Birchfield and Tomasi's pixel dissimilarity is used to improve the robustness against the image sampling noise. The smooth cost which expresses the compatibility between neighboring variables embedded in the truncated linear model, is defined as:

$$V(f_p, f_q) = \min(|f_p - f_q|, K), \quad (4.2)$$

where K is the truncating value. The smooth cost based on the truncated linear model is also referred to as discontinuity preserving cost since it can prevent the edges of objects from



Figure 11: Comparison of different optimization algorithms. (a) image is the original left image. (b) is the result BP. (c) is the result by α -expansion. (d) is the result TRW [44].

oversmoothing. The corresponding energy function used here is the most conspicuous one and is defined as

$$E(f) = \sum_{p \in \mathbf{P}} D(f_p) + \sum_{(p,q) \in \mathbf{N}} V(f_p, f_q), \quad (4.3)$$

where \mathbf{N} contains the edges in the four-connected neighborhood set.

The objective is to find a solution which minimizes (4.3). The solution means the correct depth information in the scene. Figure 11 shows the results of “Tsukuba” data set using different energy minimization methods available in [44]. In the past decades, segment-based stereos [32] have been boomed as they perform well in reducing the ambiguity associated with textureless regions and enhancing noise tolerance by aggregating over pixels with homogenous properties. Usually, those algorithms firstly segment the source image. Then the matching cost is computed over the entire segment. A plane fit method is applied to refine the result.

4.2. Binary Image Segmentation

Binary image segmentation is widely used in medical image analysis and object recognition. Here, each pixel is assigned with a label l with $0 \leq l \leq 1$. In the simplest case, we have $l \in \{0, 1\}$, where 0 represents the pixel belonging to the background and 1 to the foreground. The segmentation result should be accurate and fine enough for successful applications such as object category, photo editing, and image retrieval. Although segmentation is regarded as one of the most difficult problems due to the complexity of real scene and noise corruption, MRF model can often successfully deal with this challenging problem.

The corresponding energy function is represented the same as (2.9). The data cost E_d represents whether the pixel property is consistent with the statistic distribute of possible region. It may be simple to take such an absolute difference of pixel intensity and the mean of region gray level. Alternatively, the complex data term often leads to better results. For example, in [60], the data cost uses the color data model which is the log-likelihood of a pixel and is modeled as two separate Gaussian mixture models. The smoothness term is a simple Potts model

$$E_s(f_q, f_p) = K \times \text{Dis}(m, n)^{-1} [f_q = f_p]. \quad (4.4)$$

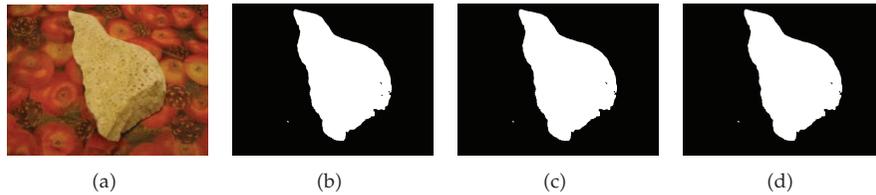


Figure 12: Comparison of different optimization algorithms for binary segmentation of the image “sponge.” (a) image is original. (b) is the result of BP. (c) is the result of α -expansion. (d) is the result of TRW. There is a slight difference among the three results. Besides, since this is a binary labeling problem, the α -expansion finds the global optimum in a single iteration [44].

In (4.4), $\text{Dis}(m, n)$ is the Euclidean distance of pixel m and pixel n , and $[f_p = f_q]$ denotes the indicator function taking 0 and 1. K is a constant. If $K = 1$, the smoothness term recovers the Ising model which encourages smoothness everywhere. K determines how coherent the similar grey level in a region is. Recently, user interaction was proposed to refine the results in [60–62]. Usually, the user first marks some pixels to indicate the background and foreground. With those labeled pixels, we can get the corresponding region statistics.

GC is the most common optimal tool for binary MRF combined with both color (texture) information and edge information. Further, the marked pixels can be used as the seeds in the cut-based algorithm. A graph cut extension, that is, grabcut [60], was proposed for iterative minimization of the energy. Figure 12 shows the results of binary segmentation with different methods using identity parameters [44].

Considering multilabel segmentation, Micusik and Pajdla [63] formulated single-image multi-label segmentation into coherent regions in texture and color as a max-sum problem. As a region merging method, Mignotte [64] used MRF fusion model combining several segmentation results to achieve a more reliable and accurate result.

More recently, Panda and Nanda [65] proposed an unsupervised color image segmentation scheme using the homotopy continuation method and compound MRF model. Chen et al. [66] proposed image segmentation method based on MAP or ML estimation. Li [67] introduced a multiresolution MRF approach to texture segmentation problems. Rivera et al. [68] presented a new MRF model for parametric image segmentation. Some other works [69–71] carried out for learning of the prior distribution. MRF is also widely used in medical image segmentation. Zhang et al. [72] proposed segmentation of brain MR images through a hidden MRF. Scherrer et al. [73] used expectation maximization to segment the images in an MRF model. Anguelov et al. [74] segment 3D scanned data into objects using GC. Hower et al. [75] investigated in the context of neuroimaging segmentation. As a low level vision problem, the segment is often applied for object classification. Honghui et al. [76] proposed a robust supervised label transfer method for semantic segmentation of street scenes. Feng et al. [77] recently proposed a method to optimize the MRF, which can automatically determine the number of labels in balance of accuracy and efficiency.

5. Conclusion

It is now acknowledged that MRF is one of the most successful approaches for solving labeling problems in computer vision and image analysis. The most challenge of MRF models is to develop its efficient inference algorithm in order to find the low-energy configuration.

As in computer vision, there are too many nodes. For example, consider two frame images with the size of $a \times b$. If each node takes N possible labels, the computation space is $N^{a \times b}$. Clearly, the inference algorithm should be efficient enough to overcome this dilemma. Secondly, constructing reasonable MRF also plays key roles, especially for some new vision applications. For instance, there are many different grid topologies and nonlocal topologies. Thirdly, the parameters of MRF model should be efficiently learned from image instead of manually or empirically chosen. Furthermore, further studies can focus on the energy functions which can not be efficiently solved by using state-of-the-art methods.

Acknowledgments

This work was supported by the National Natural Science Foundation of China and Microsoft Research Asia (NSFC-60870002, 60802087), NCET, Zhejiang Provincial S&T Department (2010R10006, 2010C33095), and Zhejiang Provincial Natural Science Foundation (R1110679).

References

- [1] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [2] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, Springer, Berlin, Germany, 2009.
- [3] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, Berlin, Germany, 1st edition, 2011.
- [4] 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.
- [5] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 41–54, 2006.
- [6] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1568–1583, 2006.
- [7] N. Komodakis, N. Paragios, and G. Tziritas, "MRF energy minimization and beyond via dual decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 531–552, 2011.
- [8] V. Lempitsky, C. Rother, S. Roth, and A. Blake, "Fusion moves for markov random field optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1392–1405, 2010.
- [9] R. Szeliski, R. Zabih, D. Scharstein et al., "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, 2008.
- [10] P. F. Felzenszwalb and R. Zabih, "Dynamic programming and graph algorithms in computer vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 721–740, 2011.
- [11] T. Werner, "A linear programming approach to max-sum problem: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1165–1179, 2007.
- [12] S. Roth and M. J. Black, "Fields of experts: a framework for learning image priors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR '05)*, vol. 2, pp. 860–867, San Diego, Calif, USA, June 2005.
- [13] U. Schmidt, Q. Gao, and S. Roth, "A generative perspective on MRFs in low-level vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR '10)*, pp. 1751–1758, San Francisco, Calif, USA, June 2010.
- [14] O. Woodford, C. Rother, and V. Kolmogorov, "A global perspective on MAP inference for low-level vision," in *Proceedings of the 12th International Conference on Computer Vision, (ICCV '09)*, pp. 2319–2326, October 2009.
- [15] S. Y. Lin and J. Y. Shi, "A Markov random field model-based approach to natural image matting," *Journal of Computer Science and Technology*, vol. 22, no. 1, pp. 161–167, 2007.
- [16] T. Kasetkasem, M. K. Arora, and P. K. Varshney, "Super-resolution land cover mapping using a Markov random field based approach," *Remote Sensing of Environment*, vol. 96, no. 3-4, pp. 302–314, 2005.

- [17] H. Heng-Feng, L. Jin-Jang, L. Cheng-Shian, and C. Hsuan-Ying, "Image inpainting using structure-guided priority belief propagation and label transformations," in *Proceeding of the 20th International Conference on Pattern Recognition, (ICPR '10)*, pp. 4492–4495, August 2010.
- [18] D. Zikic, B. Glocker, O. Kutter et al., "Markov random field optimization for intensity-based 2D-3D registration," in *Medical Imaging 2010: Image Processing*, vol. 7623 of *Proceedings of SPIE*, 2010.
- [19] J. Verbeek and B. Triggs, "Region classification with Markov field aspect models," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR '07)*, pp. 1059–1066, June 2007.
- [20] D. Larlus and F. Jurie, "Combining appearance models and Markov random fields for category level object segmentation," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, (CVPR '08)*, vol. 1–20, pp. 864–870, June 2008.
- [21] M. Lievin and F. Luthon, "Nonlinear color space and spatiotemporal MRF for hierarchical segmentation of face features in video," *IEEE Transactions on Image Processing*, vol. 13, no. 1, pp. 63–71, 2004.
- [22] X. G. Wang and X. O. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009.
- [23] S. Espana-Boquera, M. J. Castro-Bleda, J. Gorbe-Moya, and F. Zamora-Martinez, "Improving offline handwritten text recognition with hybrid HMM/ANN models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 767–779, 2011.
- [24] M. L. Song, D. C. Tao, C. Chen, X. L. Li, and C. W. Chen, "Color to gray: visual cue preservation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1537–1552, 2010.
- [25] H. Mirzaalian and G. Hamarneh, "Vessel scale-selection using MRF optimization," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR '10)*, pp. 3273–3279, June 2010.
- [26] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.
- [27] J. Sun, N. N. Zheng, and H. Y. Shum, "Stereo matching using belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 787–800, 2003.
- [28] R. Detry, N. Pugeault, and J. H. Piater, "A probabilistic framework for 3D visual object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1790–1803, 2009.
- [29] A. S. Willsky, "Multiresolution Markov models for signal and image processing," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1396–1458, 2002.
- [30] Z. Kato, M. Berthod, and J. Zerubia, "Multiscale Markov random field models for parallel image classification," in *Proceedings of the IEEE 4th International Conference on Computer Vision*, pp. 253–257, May 1993.
- [31] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy, "Sonar image segmentation using an unsupervised hierarchical MRF model," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1216–1231, 2000.
- [32] C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 49–65, 2007.
- [33] P. Kohli, L. Ladicky, and P. H. S. Torr, "Robust higher order potentials for enforcing label consistency," *International Journal of Computer Vision*, vol. 82, no. 3, pp. 302–324, 2009.
- [34] C. Rother, P. Kohli, W. Feng, and J. Y. Jia, "Minimizing sparse higher order energy functions of discrete variables," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (CVPR Workshops '09)*, vol. 1–4, pp. 1382–1389, June 2009.
- [35] B. Potetz and T. S. Lee, "Efficient belief propagation for higher-order cliques using linear constraint nodes," *Computer Vision and Image Understanding*, vol. 112, no. 1, pp. 39–54, 2008.
- [36] D. Kwon, K. J. Lee, I. D. Yun, and S. U. Lee, "Solving MRFs with higher-order smoothness priors using hierarchical gradient nodes," in *Proceedings of the 10th Asian Conference on Computer Vision*, pp. 121–134, November 2010.
- [37] D. Tarlow, I. Givoni, and R. S. Zemel, "HOP-MAP: efficient message passing with high order potentials," *Journal of Machine Learning Research*, vol. 9, pp. 812–819, 2010.
- [38] D. Kwon, K. J. Lee, I. D. Yun, and S. U. Lee, "Nonrigid image registration using dynamic higher-order MRF model," in *Proceedings of the 10th European Conference on Computer Vision, (ECCV '08)*, vol. 5302, part 1 of *Lecture Notes in Computer Science*, pp. 373–386, Marseille, France, October 2008.
- [39] X. Y. Lan, S. Roth, D. Huttenlocher, and M. J. Black, "Efficient belief propagation with learned higher-order Markov random fields," in *Proceedings of the 9th European Conference on Computer Vision (ECCV '06)*, vol. 3952, part 2, pp. 269–282, Graz, Austria, May 2006.

- [40] D. M. Greig, B. T. Porteous, and A. H. Seheult, "Exact maximum a posteriori estimation for binary images," *Journal of the Royal Statistical Society Series B Methodological*, vol. 51, no. 2, pp. 271–279, 1989.
- [41] V. Kolmogorov and C. Rother, "Minimizing nonsubmodular functions with graph cuts—a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1274–1279, 2007.
- [42] X. Liu, O. Veksler, and J. Samarabandu, "Order-preserving moves for graph-cut-based optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1182–1196, 2010.
- [43] N. Komodakis and G. Tziritas, "Approximate labeling via graph cuts based on linear programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1436–1453, 2007.
- [44] D. Scharstein and R. Szeliski, "Middlebury Stereo Vision Research Page," 2011, <http://vision.middlebury.edu/stereo/eval>.
- [45] J. Pearl, "Reverend bayes on inference engines: a distributed hierarchical approach," in *Proceedings of the National Conference on Artificial Intelligence*, pp. 133–136, Pittsburgh, Pa, USA, 1982.
- [46] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2282–2312, 2005.
- [47] T. L. Yu, R. S. Lin, B. Super, and B. Tang, "Efficient message representations for belief propagation," in *Proceedings of the 11th IEEE International Conference on Computer Vision, (ICCV '07)*, vol. 1–6, pp. 604–611, October 2007.
- [48] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Generalized belief propagation," *Advances in Neural Information Processing Systems*, vol. 13, pp. 689–695, 2001.
- [49] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," in *Proceedings of the 9th IEEE International Conference on Computer Vision*, vol. 2, pp. 900–907, October 2003.
- [50] K. Alahari, P. Kohli, and P. H. S. Torr, "Dynamic hybrid algorithms for MAP inference in discrete MRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1846–1857, 2010.
- [51] M. P. Kumar, V. Kolmogorov, and P. H. S. Torr, "An analysis of convex relaxations for MAP estimation of discrete MRFs," *Journal of Machine Learning Research*, vol. 10, pp. 71–106, 2009.
- [52] N. Komodakis and G. Tziritas, "Image completion using efficient belief propagation via priority scheduling and dynamic pruning," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2649–2661, 2007.
- [53] H. Ishikawa, "Exact optimization for Markov random fields with convex priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1333–1336, 2003.
- [54] T. S. Cho, S. Avidan, and W. T. Freeman, "The patch transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1489–1501, 2010.
- [55] Q. X. Yang, L. Wang, R. G. Yang, H. Stewenius, and D. Nister, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 492–504, 2009.
- [56] A. Barbu and S. C. Zhu, "Incorporating visual knowledge representation in stereo reconstruction," in *Proceedings of the 10th IEEE International Conference on Computer Vision, (ICCV '05)*, vol. 1–2, pp. 572–579, Beijing, China, October 2005.
- [57] N. D. F. Campbell, G. Vogiatzis, C. Hernandez, and R. Cipolla, "Using multiple hypotheses to improve depth-maps for multi-view stereo," in *Proceedings of the 10th European Conference on Computer Vision, (ECCV '08)*, vol. 5302, part 1, pp. 766–779, 2008.
- [58] J. Starck and A. Hilton, "Correspondence labelling for wide-timeframe free-form surface matching," in *Proceedings of the 11th IEEE International Conference on Computer Vision, (ICCV '07)*, vol. 1–6, pp. 2189–2196, October 2007.
- [59] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1–3, pp. 7–42, 2002.
- [60] C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut' - interactive foreground extraction using iterated graph cuts," *Acm Transactions on Graphics*, vol. 23, no. 3, pp. 309–314, 2004.
- [61] Y. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proceedings of the 8th International Conference on Computer Vision*, vol. 1, pp. 105–112, July 2001.
- [62] I. Grinias, N. Komodakis, and G. Tziritas, "Flooding and MRF-based algorithms for interactive segmentation," in *Proceedings of the 20th International Conference on Pattern Recognition, (ICPR '10)*, pp. 3943–3946, August 2010.

- [63] B. Micusik and T. Pajdla, "Multi-label image segmentation via max-sum solver," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1–8, pp. 1994–1999, 2007.
- [64] M. Mignotte, "A label field fusion bayesian model and its penalized maximum rand estimator for image segmentation," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1610–1624, 2010.
- [65] S. Panda and P. K. Nanda, "Unsupervised color image segmentation using compound Markov random field model," in *Proceedings of the 3rd International Conference on Pattern Recognition and Machine Intelligence*, vol. 5909, pp. 291–296, December 2009.
- [66] S. F. Chen, L. L. Cao, J. Z. Liu, and X. O. Tang, "Iterative MAP and ML estimations for image segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR '07)*, vol. 1–8, pp. 289–294, June 2007.
- [67] C. T. Li, "A conditional random field approach to unsupervised texture image segmentation," *Eurasip Journal on Advances in Signal Processing*, vol. 2010, 2010.
- [68] M. Rivera, O. Ocegueda, and J. L. Marroquin, "Entropy-controlled quadratic Markov measure field models for efficient image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 3047–3057, 2007.
- [69] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4085–4098, 2010.
- [70] X. Feng, C. K. I. Williams, and S. N. Felderhof, "Combining belief networks and neural networks for scene segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 467–483, 2002.
- [71] X. Jianxiong and Q. Long, "Multiple view semantic segmentation for street view images," in *Proceedings of the 12th International Conference on Computer Vision, (ICCV '09)*, pp. 686–693, October 2009.
- [72] Y. Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.
- [73] B. Scherrer, M. Dojat, F. Forbes, and C. Garbay, "Agentification of Markov model-based segmentation: application to magnetic resonance brain scans," *Artificial Intelligence in Medicine*, vol. 46, no. 1, pp. 81–95, 2009.
- [74] 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)*, vol. 2, pp. 169–176, June 2005.
- [75] D. Hoyer, V. Singh, and S. Johnson, "Label set perturbation for MRF based neuroimaging segmentation," in *Proceedings of the 12th International Conference on Computer Vision, (ICCV '09)*, pp. 849–856, October 2009.
- [76] Z. Honghui, X. Jianxiong, and Q. Long, "Supervised label transfer for semantic segmentation of street scenes," in *Proceedings of the 11th European Conference on Computer Vision, (ECCV '10)*, vol. 6315, pp. 561–574, September 2010.
- [77] W. Feng, J. Y. Jia, and Z. Q. Liu, "Self-validated labeling of Markov random fields for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1871–1887, 2010.

Research Article

Numerov's Method for a Class of Nonlinear Multipoint Boundary Value Problems

Yuan-Ming Wang,^{1,2} Wen-Jia Wu,¹ and Massimo Scalia³

¹ Department of Mathematics, East China Normal University, Shanghai 200241, China

² Scientific Computing Key Laboratory of Shanghai Universities, Division of Computational Science, E-Institute of Shanghai Universities, Shanghai Normal University, Shanghai 200234, China

³ Department of Mathematics, University of Rome "La Sapienza", Piazzale le Aldo Moro 2, 00185 Rome, Italy

Correspondence should be addressed to Yuan-Ming Wang, ymwang@math.ecnu.edu.cn

Received 28 March 2011; Revised 16 May 2011; Accepted 13 June 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Yuan-Ming Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of this paper is to give a numerical treatment for a class of nonlinear multipoint boundary value problems. The multipoint boundary condition under consideration includes various commonly discussed boundary conditions, such as the three- or four-point boundary condition. The problems are discretized by the fourth-order Numerov's method. The existence and uniqueness of the numerical solution are investigated by the method of upper and lower solutions. The convergence and the fourth-order accuracy of the method are proved. An accelerated monotone iterative algorithm with the quadratic rate of convergence is developed for solving the resulting nonlinear discrete problems. Some applications and numerical results are given to demonstrate the high efficiency of the approach.

1. Introduction

Multipoint boundary value problems arise in various fields of applied science. An often discussed problem is the following nonlinear second-order multipoint boundary value problem:

$$\begin{aligned} -u''(x) &= f(x, u(x)), \quad 0 < x < 1, \\ u(0) &= \sum_{i=1}^p \alpha_i u(\xi_i), \quad u(1) = \sum_{i=1}^p \beta_i u(\eta_i), \end{aligned} \tag{1.1}$$

where $f(x, u)$ is a continuous function of its arguments and for each i , $\alpha_i, \beta_i \in [0, \infty)$ and $\xi_i, \eta_i \in (0, 1)$. An application of this problem appears in the design of a large-size bridge with multipoint supports, where $u(x)$ denotes the displacement of the bridge from the unloaded position (e.g., see [1]). For other applications of problem (1.1), we see [2–4] and the references therein. It is allowed in (1.1) that $\alpha_i = 0$ or $\beta_i = 0$ for some or all i . This implies that the boundary condition in (1.1) includes various commonly discussed multipoint boundary conditions. In particular, the boundary condition in (1.1) is reduced to

$$u(0) = 0, \quad u(1) = \sum_{i=1}^p \beta_i u(\eta_i), \quad (1.1_a)$$

if $\alpha_i = 0$ for all i (see [5–14]), to the form

$$u(0) = \sum_{i=1}^p \alpha_i u(\xi_i), \quad u(1) = 0, \quad (1.1_b)$$

if $\beta_i = 0$ for all i (see [15]), to the four-point boundary condition

$$u(0) = \alpha u(\xi), \quad u(1) = \beta u(\eta), \quad (1.1_c)$$

if $p = 1$ and $\xi_1 = \xi, \eta_1 = \eta$ (see [11, 15–17]), and to the two-point boundary condition

$$u(0) = u(1) = 0, \quad (1.1_d)$$

if $\alpha_i = 0$ and $\beta_i = 0$ for all i . Condition (1.1_c) includes the three-point boundary condition when $\xi = \eta$ (see [16, 17]).

The study of multipoint boundary value problems for linear second-order ordinary differential equations was initiated in [18, 19] by Il'in and Moiseev. In [20], Gupta studied a three-point boundary value problem for nonlinear second-order ordinary differential equations. Since then, more general nonlinear second-order multipoint boundary value problems in the form (1.1) have been studied. Most of the discussions were concerned with the existence and multiplicity of solutions by using different methods. Applying the fixed point index theorem in cones, the works in [5–14] showed the existence of one or more solutions to the problem (1.1)-(1.1_a), while the works in [15–17] were devoted to the existence of solutions for the three- or four-point boundary value problem (1.1)-(1.1_c). For the problem (1.1) with the more general multipoint boundary conditions, some existence results were obtained in [21, 22] by using the fixed point index theory or the topological degree theory. Based on the method of upper and lower solutions, the authors of [17, 23] obtained some sufficient conditions so that (1.1) or its some special form has at least one solution. Additional works that deal with the existence problem of nonlinear second-order multipoint boundary value problems can be found in [24–29].

On the other hand, there are also some works that are devoted to numerical methods for the solutions of multipoint boundary value problems. The work in [30] made use of the Chebyshev series for approximating solutions of nonlinear first-order multipoint

boundary value problems, and the work in [31] showed how an adaptive finite difference technique can be developed to produce efficient approximations to the solutions of nonlinear multipoint boundary value problems for first-order systems of equations. Another method for computing the solutions of nonlinear first-order multipoint boundary value problems was described in [32], where a multiple shooting technique was developed. Some other works for the computational methods of first-order multipoint boundary value problems can be seen in [33–35]. In [36–38] the authors gave several constructive methods for the solutions of multipoint discrete boundary value problems, including the method of adjoints, the invariant embedding method, and the shooting-type method. In the case of second-order multipoint boundary value problems, there are only a few computational algorithms in the literature. The paper [39] set up a reproducing kernel Hilbert space method for the solution of a second-order three-point boundary value problem. Based upon the shooting technique, a numerical method was developed in [1] for approximating solutions and fold bifurcation solutions of a class of second-order multipoint boundary value problems.

As we know, Numerov's method is one of the well-known difference methods to solve the second-order ordinary differential equation $-u'' = f(x, u)$. Because Numerov's method possesses the fourth-order accuracy and a compact property, it has attracted considerable attention and has been extensively applied in practical computations (cf. [40–51]). Although many theoretical investigations have focused on Numerov's method for two-point boundary conditions such as (1.1_d) (cf. [40, 41, 43, 44, 47–51]), there is relatively little discussion on the analysis of Numerov's method applied to fully multipoint boundary conditions in (1.1). The study presented in this paper is aimed at filling in such a gap by considering Numerov's method for the numerical solution of the multipoint boundary value problem (1.1) with the more general boundary conditions, including the boundary conditions (1.1_a), (1.1_b), and (1.1_c). It is not difficult to give a Numerov's difference approximation to (1.1) in the same manner as that for two-point boundary value problems. However, a lack of explicit information about the boundary value of the solution in the multipoint boundary conditions prevents us from using the standard analysis process of treating two-point boundary value problems, and so we here develop a different approach for the analysis of Numerov's difference approximation to (1.1). Our specific goals are (1) to establish the existence and uniqueness of the numerical solution, (2) to show the convergence of the numerical solution to the analytic solution with the fourth-order accuracy, and (3) to develop an efficient computational algorithm for solving the resulting nonlinear discrete problems. To achieve the above goals, we use the method of upper and lower solutions and its associated monotone iterations. It should be mentioned that the proposed fourth-order Numerov's discretization methodology may be straightforwardly extended to the following nonhomogeneous multipoint boundary condition:

$$u(0) = \sum_{i=1}^p \alpha_i u(\xi_i) + \lambda_1, \quad u(1) = \sum_{i=1}^p \beta_i u(\eta_i) + \lambda_2, \quad (1.2)$$

where λ_1 and λ_2 are two prescribed constants.

The outline of the paper is as follows. In Section 2, we discretize (1.1) into a finite difference system by Numerov's technique. In Section 3, we deal with the existence and

uniqueness of the numerical solution by using the method of upper and lower solutions. The convergence of the numerical solution and the fourth-order accuracy of the method are proved in Section 4. Section 5 is devoted to an accelerated monotone iterative algorithm for solving the resulting nonlinear discrete problem. Using an upper solution and a lower solution as initial iterations, the iterative algorithm yields two sequences that converge monotonically from above and below, respectively, to a unique solution of the resulting nonlinear discrete problem. It is shown that the rate of convergence for the sum of the two produced sequences is quadratic (the error metric is the sum of the infinity norm of the error between the m th-iteration of the upper solution and the true solution with the infinity norm of the error between the m th-iteration of the lower solution and the true solution) and under an additional requirement, the quadratic rate of convergence is attained for one of these two sequences. In Section 6, we give some applications to three model problems and present some numerical results demonstrating the monotone and rapid convergence of the iterative sequences and the fourth-order accuracy of the method. We also compare our method with the standard finite difference method and show its advantages. The final section contains some concluding remarks.

2. Numerov's Method

Let $h = 1/L$ be the mesh size, and let $x_i = ih$ ($0 \leq i \leq L$) be the mesh points in $[0, 1]$. Assume that for all $1 \leq i \leq p$, the points ξ_i and η_i in the boundary condition of (1.1) serve as mesh points. This assumption is always satisfied by a proper choice of mesh size h . For convenience, we use the following notations:

$$S_\alpha[u(\xi)] = \sum_{i=1}^p \alpha_i u(\xi_i), \quad S_\beta[u(\eta)] = \sum_{i=1}^p \beta_i u(\eta_i) \quad (2.1)$$

and introduce the finite difference operators δ_h^2 and \mathcal{D}_h as follows:

$$\begin{aligned} \delta_h^2 u(x_i) &= u(x_{i-1}) - 2u(x_i) + u(x_{i+1}), \quad 1 \leq i \leq L-1, \\ \mathcal{D}_h u(x_i) &= \frac{h^2}{12} (u(x_{i-1}) + 10u(x_i) + u(x_{i+1})), \quad 1 \leq i \leq L-1. \end{aligned} \quad (2.2)$$

Using the following Numerov's formula (cf. [52, 53]):

$$\delta_h^2 u(x_i) = \mathcal{D}_h u''(x_i) + \mathcal{O}(h^6), \quad 1 \leq i \leq L-1, \quad (2.3)$$

we have from (1.1) and (2.1) that

$$\begin{aligned} -\delta_h^2 u(x_i) &= \mathcal{D}_h f(x_i, u(x_i)) + \mathcal{O}(h^6), \quad 1 \leq i \leq L-1, \\ u(0) &= S_\alpha[u(\xi)], \quad u(1) = S_\beta[u(\eta)]. \end{aligned} \quad (2.4)$$

After dropping the $\mathcal{O}(h^6)$ term, we derive a Numerov's difference approximation to (1.1) as follows:

$$\begin{aligned} -\delta_h^2 u_h(x_i) &= \mathcal{D}_h f(x_i, u_h(x_i)), \quad 1 \leq i \leq L-1, \\ u_h(0) &= S_\alpha[u_h(\xi)], \quad u_h(1) = S_\beta[u_h(\eta)], \end{aligned} \quad (2.5)$$

where $u_h(x_i)$ represents the approximation of $u(x_i)$.

For two constants \underline{M} and \overline{M} satisfying $\overline{M} \geq \underline{M} > -\pi^2$, we define

$$h(\underline{M}, \overline{M}) = \begin{cases} \sqrt{\frac{12}{\underline{M}}}, & \underline{M} > -8, \overline{M} > 0, \\ 1, & \underline{M} > -8, \overline{M} \leq 0, \\ \min \left\{ \sqrt{\frac{12}{\underline{M}}}, \sqrt{\frac{12}{\pi^2} \left(1 + \frac{\underline{M}}{\pi^2}\right)} \right\}, & \underline{M} \leq -8, \overline{M} > 0, \\ \sqrt{\frac{12}{\pi^2} \left(1 + \frac{\underline{M}}{\pi^2}\right)}, & \underline{M} \leq -8, \overline{M} \leq 0. \end{cases} \quad (2.6)$$

A fundamental and useful property of the operators δ_h^2 and \mathcal{D}_h is stated below.

Lemma 2.1 (See Lemma 3.1 of [50]). *Let \underline{M} , \overline{M} , and M_i be some constants satisfying*

$$-\pi^2 < \underline{M} \leq M_i \leq \overline{M}, \quad 0 \leq i \leq L. \quad (2.7)$$

If

$$\begin{aligned} -\delta_h^2 u_h(x_i) + \mathcal{D}_h(M_i u_h(x_i)) &\geq 0, \quad 1 \leq i \leq L-1, \\ u_h(0) &\geq 0, \quad u_h(1) \geq 0, \end{aligned} \quad (2.8)$$

and $h < h(\underline{M}, \overline{M})$, then $u_h(x_i) \geq 0$ for all $0 \leq i \leq L$.

The following results are also useful for our forthcoming discussions. Their proofs will be given in the appendix.

Lemma 2.2. *Assume*

$$\sigma \equiv \max \left\{ \sum_{i=1}^p \alpha_i, \sum_{i=1}^p \beta_i \right\} < 1. \quad (2.9)$$

Let \underline{M} , \overline{M} , and M_i be the given constants such that

$$-8(1 - \sigma) < \underline{M} \leq M_i \leq \overline{M}, \quad 0 \leq i \leq L. \quad (2.10)$$

If

$$\begin{aligned} -\delta_h^2 u_h(x_i) + \mathcal{D}_h(M_i u_h(x_i)) &\geq 0, \quad 1 \leq i \leq L-1, \\ u_h(0) &\geq S_\alpha[u_h(\xi)], \quad u_h(1) \geq S_\beta[u_h(\eta)], \end{aligned} \quad (2.11)$$

and $h < h(\underline{M}, \overline{M})$, then $u_h(x_i) \geq 0$ for all $0 \leq i \leq L$.

Lemma 2.3. Let the condition (2.9) be satisfied, and let \underline{M} , \overline{M} , and M_i be the given constants satisfying (2.10). Assume that the functions $u_h(x_i)$ and $g(x_i)$ satisfy

$$\begin{aligned} -\delta_h^2 u_h(x_i) + \mathcal{D}_h(M_i u_h(x_i)) &= g(x_i), \quad 1 \leq i \leq L-1, \\ u_h(0) &= S_\alpha[u_h(\xi)], \quad u_h(1) = S_\beta[u_h(\eta)]. \end{aligned} \quad (2.12)$$

Then when $h < h(\underline{M}, \overline{M})$,

$$\|u_h\|_\infty \leq \frac{\|g\|_\infty}{(8(1-\sigma) + \min(\underline{M}, 0))h^2}, \quad (2.13)$$

where $\|u_h\|_\infty = \max_{1 \leq i \leq L-1} |u_h(x_i)|$ denotes discrete infinity norm for any mesh function $u_h(x_i)$.

Remark 2.4. It is clear from Lemma 2.1 that if $\sigma = 0$ then the condition (2.10) in Lemma 2.2 can be replaced by the weaker condition (2.7). Lemmas 2.1 and 2.2 guarantee that the linear problems based on (2.8) and (2.11) with the inequality relation “ \geq ” replaced by the equality relation “ $=$ ” are well posed.

3. The Existence and Uniqueness of the Solution

To investigate the existence and uniqueness of the solution of (2.5), we use the method of upper and lower solutions. The definition of the upper and lower solutions is given as follows.

Definition 3.1. A function $\tilde{u}_h(x_i)$ is called an upper solution of (2.5) if

$$\begin{aligned} -\delta_h^2 \tilde{u}_h(x_i) &\geq \mathcal{D}_h f(x_i, \tilde{u}_h(x_i)), \quad 1 \leq i \leq L-1, \\ \tilde{u}_h(0) &\geq S_\alpha[\tilde{u}_h(\xi)], \quad \tilde{u}_h(1) \geq S_\beta[\tilde{u}_h(\eta)]. \end{aligned} \quad (3.1)$$

Similarly, a function $\hat{u}_h(x_i)$ is called a lower solution of (2.5) if it satisfies the above inequalities in the reversed order. A pair of upper and lower solutions $\tilde{u}_h(x_i)$ and $\hat{u}_h(x_i)$ are said to be ordered if $\tilde{u}_h(x_i) \geq \hat{u}_h(x_i)$ for all $0 \leq i \leq L$.

It is clear that every solution of (2.5) is an upper solution as well as a lower solution. For a given pair of ordered upper and lower solutions $\tilde{u}_h(x_i)$ and $\hat{u}_h(x_i)$, we set

$$\begin{aligned} \langle \hat{u}_h, \tilde{u}_h \rangle &= \{u_h; \hat{u}_h(x_i) \leq u_h(x_i) \leq \tilde{u}_h(x_i) (0 \leq i \leq L)\}, \\ [\hat{u}_h(x_i), \tilde{u}_h(x_i)] &= \{u_i \in \mathbf{R}; \hat{u}_h(x_i) \leq u_i \leq \tilde{u}_h(x_i)\} \end{aligned} \quad (3.2)$$

and make the following basic hypotheses:

(H₁) For each $0 \leq i \leq L$, there exists a constant M_i such that $M_i > -\pi^2$ and

$$f(x_i, v_i) - f(x_i, v'_i) \geq -M_i(v_i - v'_i) \quad (3.3)$$

whenever $\hat{u}_h(x_i) \leq v'_i \leq v_i \leq \tilde{u}_h(x_i)$;

(H₂) $h < h(\underline{M}, \overline{M})$, where $\overline{M} = \max_{0 \leq i \leq L} M_i$ and $\underline{M} = \min_{0 \leq i \leq L} M_i$.

The existence of the constant M_i in (H₁) is trivial if $f(x_i, u)$ is a C^1 -function of $u \in [\hat{u}_h(x_i), \tilde{u}_h(x_i)]$. In fact, M_i may be taken as any nonnegative constant satisfying

$$M_i \geq \max\{-f_u(x_i, u); u \in [\hat{u}_h(x_i), \tilde{u}_h(x_i)]\}. \quad (3.4)$$

Theorem 3.2. *Let $\tilde{u}_h(x_i)$ and $\hat{u}_h(x_i)$ be a pair of ordered upper and lower solutions of (2.5), and let hypotheses (H₁) and (H₂) be satisfied. Then system (2.5) has a maximal solution $\bar{u}_h(x_i)$ and a minimal solution $\underline{u}_h(x_i)$ in $\langle \hat{u}_h, \tilde{u}_h \rangle$. Here, the maximal property of $\bar{u}_h(x_i)$ means that for any solution $u_h(x_i)$ of (2.5) in $\langle \hat{u}_h, \tilde{u}_h \rangle$, one has $u_h(x_i) \leq \bar{u}_h(x_i)$ for all $0 \leq i \leq L$. The minimal property of $\underline{u}_h(x_i)$ is similarly understood.*

Proof. The proof is constructive. Using the initial iterations $\bar{u}_h^{(0)}(x_i) = \tilde{u}_h(x_i)$ and $\underline{u}_h^{(0)}(x_i) = \hat{u}_h(x_i)$ we construct two sequences $\{\bar{u}_h^{(m)}(x_i)\}$ and $\{\underline{u}_h^{(m)}(x_i)\}$, respectively, from the following iterative scheme:

$$\begin{aligned} -\delta_h^2 \underline{u}_h^{(m)}(x_i) + \mathcal{D}_h(M_i \underline{u}_h^{(m)}(x_i)) &= \mathcal{D}_h(M_i \underline{u}_h^{(m-1)}(x_i) + f(x_i, \underline{u}_h^{(m-1)}(x_i))), \quad 1 \leq i \leq L-1, \\ \underline{u}_h^{(m)}(0) &= S_\alpha[u_h^{(m-1)}(\xi)], \quad \underline{u}_h^{(m)}(1) = S_\beta[u_h^{(m-1)}(\eta)], \end{aligned} \quad (3.5)$$

where M_i is the constant in (H₁). By Lemma 2.1, these two sequences are well defined. We shall first prove that for all $m = 0, 1, \dots$,

$$\underline{u}_h^{(m)}(x_i) \leq \underline{u}_h^{(m+1)}(x_i) \leq \bar{u}_h^{(m+1)}(x_i) \leq \bar{u}_h^{(m)}(x_i), \quad 0 \leq i \leq L. \quad (3.6)$$

Let $\bar{w}_h^{(0)}(x_i) = \bar{u}_h^{(0)}(x_i) - \bar{u}_h^{(1)}(x_i)$. Then by (3.1) and (3.5),

$$\begin{aligned} -\delta_h^2 \bar{w}_h^{(0)}(x_i) + \mathcal{D}_h(M_i \bar{w}_h^{(0)}(x_i)) &\geq 0, \quad 1 \leq i \leq L-1, \\ \bar{w}_h^{(0)}(0) &\geq 0, \quad \bar{w}_h^{(0)}(1) \geq 0. \end{aligned} \quad (3.7)$$

It follows from Lemma 2.1 that $\bar{w}_h^{(0)}(x_i) \geq 0$, that is, $\bar{u}_h^{(0)}(x_i) \geq \bar{u}_h^{(1)}(x_i)$ for all $0 \leq i \leq L$. A similar argument using the property of a lower solution gives $\underline{u}_h^{(1)}(x_i) \geq \underline{u}_h^{(0)}(x_i)$ for all $0 \leq i \leq L$. Let $w_h^{(1)}(x_i) = \bar{u}_h^{(1)}(x_i) - \underline{u}_h^{(1)}(x_i)$. We have from (3.3) and (3.5) that

$$\begin{aligned} -\delta_h^2 w_h^{(1)}(x_i) + \mathcal{D}_h \left(M_i w_h^{(1)}(x_i) \right) &\geq 0, \quad 1 \leq i \leq L-1, \\ w_h^{(1)}(0) &\geq 0, \quad w_h^{(1)}(1) \geq 0. \end{aligned} \quad (3.8)$$

Again by Lemma 2.1, $w_h^{(1)}(x_i) \geq 0$, that is, $\bar{u}_h^{(1)}(x_i) \geq \underline{u}_h^{(1)}(x_i)$ for all $0 \leq i \leq L$. This proves (3.6) for $m = 0$. Finally, an induction argument leads to the desired result (3.6) for all $m = 0, 1, \dots$

In view of (3.6), the limits

$$\lim_{m \rightarrow \infty} \bar{u}_h^{(m)}(x_i) = \bar{u}_h(x_i), \quad \lim_{m \rightarrow \infty} \underline{u}_h^{(m)}(x_i) = \underline{u}_h(x_i), \quad 0 \leq i \leq L \quad (3.9)$$

exist and satisfy

$$\underline{u}_h^{(m)}(x_i) \leq \underline{u}_h^{(m+1)}(x_i) \leq \underline{u}_h(x_i) \leq \bar{u}_h(x_i) \leq \bar{u}_h^{(m+1)}(x_i) \leq \bar{u}_h^{(m)}(x_i), \quad 0 \leq i \leq L, \quad m = 0, 1, \dots \quad (3.10)$$

Letting $m \rightarrow \infty$ in (3.5) shows that both $\bar{u}_h(x_i)$ and $\underline{u}_h(x_i)$ are solutions of (2.5).

Now, if $u_h(x_i)$ is a solution of (2.5) in $\langle \hat{u}_h, \tilde{u}_h \rangle$, then the pair $u_h(x_i)$ and $\hat{u}_h(x_i)$ are also a pair of ordered upper and lower solutions of (2.5). The above arguments imply that $\underline{u}_h(x_i) \leq u_h(x_i)$ for all $0 \leq i \leq L$. Similarly, we have $u_h(x_i) \leq \bar{u}_h(x_i)$ for all $0 \leq i \leq L$. This shows that $\bar{u}_h(x_i)$ and $\underline{u}_h(x_i)$ are the maximal and the minimal solutions of (2.5) in $\langle \hat{u}_h, \tilde{u}_h \rangle$, respectively. The proof is completed. \square

Theorem 3.2 shows that the system (2.5) has a maximal solution $\bar{u}_h(x_i)$ and a minimal solution $\underline{u}_h(x_i)$ in $\langle \hat{u}_h, \tilde{u}_h \rangle$. If $\bar{u}_h(x_i) = \underline{u}_h(x_i)$ for all $0 \leq i \leq L$, then $\bar{u}_h(x_i)$ or $\underline{u}_h(x_i)$ is a unique solution of (2.5) in $\langle \hat{u}_h, \tilde{u}_h \rangle$. In general, these two solutions do not coincide. Consider, for example, the case

$$\sum_{i=1}^p \alpha_i = \sum_{i=1}^p \beta_i = 1. \quad (3.11)$$

If there exist two different constants \bar{c} and \underline{c} such that $f(x, \bar{c}) = f(x, \underline{c}) = 0$ for all $x \in (0, 1)$ then both \bar{c} and \underline{c} are solutions of (2.5). Hence to show the uniqueness of a solution it is necessary to impose some additional conditions on α_i , β_i and f . Assume that there exists a constant \underline{M}_u such that

$$f(x_i, v_i) - f(x_i, v'_i) \leq -\underline{M}_u (v_i - v'_i), \quad 0 \leq i \leq L \quad (3.12)$$

whenever $\hat{u}_h(x_i) \leq v'_i \leq v_i \leq \tilde{u}_h(x_i)$. This condition is trivially satisfied if $f(x_i, u)$ is a C^1 -function of $u \in [\hat{u}_h(x_i), \tilde{u}_h(x_i)]$ for all $0 \leq i \leq L$. In fact, \underline{M}_u may be taken as

$$\underline{M}_u = \min_{0 \leq i \leq L} \min \{ -f_u(x_i, u); u \in [\hat{u}_h(x_i), \tilde{u}_h(x_i)] \}. \quad (3.13)$$

The following theorem gives a sufficient condition for the uniqueness of a solution.

Theorem 3.3. *Let the conditions in Theorem 3.2 hold. If, in addition, the conditions (2.9) and (3.12) hold and either*

$$-8(1 - \sigma) < \underline{M}_u \leq 0 \quad \text{or} \quad \underline{M}_u > 0, h < \sqrt{\frac{12}{\underline{M}_u}}, \quad (3.14)$$

then the system (2.5) has a unique solution $u_h^*(x_i)$ in $\langle \hat{u}_h, \tilde{u}_h \rangle$. Moreover, the relation (3.10) holds with $\bar{u}_h(x_i) = \underline{u}_h(x_i) = u_h^*(x_i)$ for all $0 \leq i \leq L$.

Proof. It suffices to show $\bar{u}_h(x_i) = \underline{u}_h(x_i)$ for all $0 \leq i \leq L$, where $\bar{u}_h(x_i)$ and $\underline{u}_h(x_i)$ are the limits in (3.9). Let $w_h(x_i) = \bar{u}_h(x_i) - \underline{u}_h(x_i)$. Then $w_h(x_i) \geq 0$, and by (2.5),

$$\begin{aligned} -\delta_h^2 w_h(x_i) &= \mathcal{D}_h(f(x_i, \bar{u}_h(x_i)) - f(x_i, \underline{u}_h(x_i))), \quad 1 \leq i \leq L - 1, \\ w_h(0) &= S_\alpha[w_h(\xi)], \quad w_h(1) = S_\beta[w_h(\eta)]. \end{aligned} \quad (3.15)$$

Therefore, we have from (3.12) that

$$\begin{aligned} -\delta_h^2 w_h(x_i) + \mathcal{D}_h(\underline{M}_u w_h(x_i)) &\leq 0, \quad 1 \leq i \leq L - 1, \\ w_h(0) &= S_\alpha[w_h(\xi)], \quad w_h(1) = S_\beta[w_h(\eta)]. \end{aligned} \quad (3.16)$$

By Lemma 2.2, $w_h(x_i) \leq 0$ for all $0 \leq i \leq L$. This proves $\bar{u}_h(x_i) = \underline{u}_h(x_i)$ for all $0 \leq i \leq L$.

To give another sufficient condition, we assume that there exists a nonnegative constant \overline{M}_u^* such that

$$|f(x_i, v_i) - f(x_i, v'_i)| \leq \overline{M}_u^* |v_i - v'_i|, \quad 0 \leq i \leq L \quad (3.17)$$

whenever $\hat{u}_h(x_i) \leq v'_i \leq v_i \leq \tilde{u}_h(x_i)$. If $f(x_i, u)$ is a C^1 -function of $u \in [\hat{u}_h(x_i), \tilde{u}_h(x_i)]$ for all $0 \leq i \leq L$, the above condition is clearly satisfied by

$$\overline{M}_u^* = \max_{0 \leq i \leq L} \max \{ |f_u(x_i, u)|; u \in [\hat{u}_h(x_i), \tilde{u}_h(x_i)] \}. \quad (3.18) \quad \square$$

Theorem 3.4. *Let the conditions in Theorem 3.2 hold. If, in addition, the conditions (2.9) and (3.17) hold and*

$$\overline{M}_u^* < 8(1 - \sigma), \quad (3.19)$$

then the conclusions of Theorem 3.3 are also valid.

Proof. Applying Lemma 2.3 with $M_i = 0$ to (3.15) leads to

$$\|w_h\|_\infty \leq \frac{\|g\|_\infty}{8(1-\sigma)h^2}, \quad (3.20)$$

where $g(x_i) = \mathcal{D}_h(f(x_i, \bar{u}_h(x_i)) - f(x_i, \underline{u}_h(x_i)))$. By (3.17), we obtain $\|g\|_\infty \leq h^2 \bar{M}_u^* \|w_h\|_\infty$. Consequently,

$$\|w_h\|_\infty \leq \frac{\bar{M}_u^* \|w_h\|_\infty}{8(1-\sigma)}. \quad (3.21)$$

This together with (3.19) implies $w_h(x_i) = 0$, that is, $\bar{u}_h(x_i) = \underline{u}_h(x_i)$ for all $0 \leq i \leq L$. \square

It is seen from the proofs of Theorems 3.2–3.4 that the iterative scheme (3.5) not only leads to the existence and uniqueness of the solution of (2.5) but also provides a monotone iterative algorithm for computing the solution. However, the rate of convergence of the iterative scheme (3.5) is only of linear order because it is of Picard type. A more efficient monotone iterative algorithm with the quadratic rate of convergence will be developed in Section 5.

4. Convergence of Numerov's Method

In this section, we deal with the convergence of the numerical solution and show the fourth-order accuracy of Numerov's scheme (2.5). Throughout this section, we assume that the function $f(x, u)$ and the solution $u(x)$ of (1.1) are sufficiently smooth.

Let $u(x_i)$ be the value of the solution of (1.1) at the mesh point x_i , and let $u_h(x_i)$ be the solution of (2.5). We consider the error $e_h(x_i) = u(x_i) - u_h(x_i)$. In fact, we have from (2.4) and (2.5) that

$$\begin{aligned} -\delta_h^2 e_h(x_i) &= \mathcal{D}_h(f(x_i, u(x_i)) - f(x_i, u_h(x_i))) + \mathcal{O}(h^6), \quad 1 \leq i \leq L-1, \\ e_h(0) &= S_\alpha[e_h(\xi)], \quad e_h(1) = S_\beta[e_h(\eta)]. \end{aligned} \quad (4.1)$$

Theorem 4.1. *Let the condition (2.9) hold, and let $[u_{*,i}, u_i^*]$ be an interval in \mathbf{R} such that $u(x_i), u_h(x_i) \in [u_{*,i}, u_i^*]$. Assume that*

$$\max_{0 \leq i \leq L} \max\{f_u(x_i, u); u \in [u_{*,i}, u_i^*]\} < 8(1-\sigma). \quad (4.2)$$

Then for sufficiently small h ,

$$\|u - u_h\|_\infty \leq C^* h^4, \quad (4.3)$$

where C^ is a positive constant independent of h .*

Proof. Applying the mean value theorem to the first equality of (4.1), we have

$$\begin{aligned} -\delta_h^2 e_h(x_i) + \mathcal{D}_h(M_i e_h(x_i)) &= \mathcal{O}(h^6), \quad 1 \leq i \leq L-1, \\ e_h(0) &= S_\alpha[e_h(\xi)], \quad e_h(1) = S_\beta[e_h(\eta)], \end{aligned} \quad (4.4)$$

where $M_i = -f_u(x_i, \theta_i)$ and $\theta_i \in [u_{*,i}, u_i^*]$. Let $\overline{M} = \min_i M_i$ and $\underline{M} = \max_i M_i$. Then by (4.2), $-8(1-\sigma) < \underline{M} \leq M_i \leq \overline{M}$. We, therefore, obtain from Lemma 2.3 that when $h < h(\underline{M}, \overline{M})$,

$$\|e_h\|_\infty \leq \frac{C_1 \|\mathcal{O}(h^6)\|_\infty}{h^2}, \quad (4.5)$$

where C_1 is a positive constant independent of h . Finally, the error estimate (4.3) follows from $\|\mathcal{O}(h^6)\|_\infty \leq C_2 h^6$ for some positive constant independent of h . \square

Theorem 4.1 shows that Numerov's scheme (2.5) possesses the fourth-order accuracy under the conditions of the theorem.

5. An Accelerated Monotone Iterative Algorithm

The iterative scheme (3.5) gives an algorithm for solving the system (2.5). However, as already mentioned in Section 3, its rate of convergence is only of linear order because it is of Picard type. To raise the rate of convergence while maintaining the monotone convergence of the sequence, we propose an accelerated monotone iterative algorithm. An advantage of this algorithm is that its rate of convergence for the sum of the two produced sequences is quadratic (in the sense mentioned in Section 1) with only the usual differentiability requirement on the function $f(\cdot, u)$. If the function $f_u(\cdot, u)$ possesses a monotone property in u , this algorithm is reduced to Newton's method, and one of the two produced sequences converges quadratically.

5.1. Monotone Iterative Algorithm

Let $\tilde{u}_h(x_i)$ and $\hat{u}_h(x_i)$ be a pair of ordered upper and lower solutions of (2.5) and assume that $f(\cdot, u)$ is a C^1 -function of $u \in \langle \hat{u}_h, \tilde{u}_h \rangle$. It follows from Theorems 3.2–3.4 that (2.5) has a unique solution $u_h^*(x_i)$ in $\langle \hat{u}_h, \tilde{u}_h \rangle$ under the conditions of the theorems. To compute this solution, we use the following iterative scheme:

$$\begin{aligned} -\delta_h^2 u_h^{(m)}(x_i) + \mathcal{D}_h(M_i^{(m-1)} u_h^{(m)}(x_i)) \\ = \mathcal{D}_h(M_i^{(m-1)} u_h^{(m-1)}(x_i) + f(x_i, u_h^{(m-1)}(x_i))), \quad 1 \leq i \leq L-1, \\ u_h^{(m)}(0) = S_\alpha[u_h^{(m)}(\xi)], \quad u_h^{(m)}(1) = S_\beta[u_h^{(m)}(\eta)], \end{aligned} \quad (5.1)$$

where $u_h^{(0)}(x_i)$ is either $\tilde{u}_h(x_i)$ or $\hat{u}_h(x_i)$, and for each i ,

$$M_i^{(m)} = \max\left\{-f_u(x_i, u); u \in \left[\underline{u}_h^{(m)}(x_i), \overline{u}_h^{(m)}(x_i)\right]\right\}. \quad (5.2)$$

The functions $\overline{u}_h^{(m)}(x_i)$ and $\underline{u}_h^{(m)}(x_i)$ in the definition of $M_i^{(m)}$ are obtained from (5.1) with $u_h^{(0)}(x_i) = \tilde{u}_h(x_i)$ and $u_h^{(0)}(x_i) = \hat{u}_h(x_i)$, respectively. It is clear from (5.2) that

$$f(x_i, v_i) - f(x_i, v'_i) \geq -M_i^{(m)}(v_i - v'_i), \quad 0 \leq i \leq L, \quad (5.3)$$

whenever $\underline{u}_h^{(m)}(x_i) \leq v'_i \leq v_i \leq \overline{u}_h^{(m)}(x_i)$. Moreover,

$$M_i^{(m)} = \begin{cases} -f_u(x_i, \overline{u}_h^{(m)}(x_i)), & \text{if } f_u(x_i, u) \text{ is monotone nonincreasing in } u \in \left[\underline{u}_h^{(m)}(x_i), \overline{u}_h^{(m)}(x_i)\right], \\ -f_u(x_i, \underline{u}_h^{(m)}(x_i)), & \text{if } f_u(x_i, u) \text{ is monotone nondecreasing in } u \in \left[\underline{u}_h^{(m)}(x_i), \overline{u}_h^{(m)}(x_i)\right]. \end{cases} \quad (5.4)$$

Hence, if $f_u(x_i, u)$ is monotone nonincreasing/nondecreasing in $u \in \left[\underline{u}_h^{(m)}(x_i), \overline{u}_h^{(m)}(x_i)\right]$ for all $0 \leq i \leq L$, then the iterative scheme (5.1) for $\{\overline{u}_h^{(m)}(x_i)\} / \{\underline{u}_h^{(m)}(x_i)\}$ is reduced to Newton's form:

$$\begin{aligned} & -\delta_h^2 \underline{u}_h^{(m)}(x_i) - \mathcal{D}_h \left(f_u(x_i, \underline{u}_h^{(m-1)}(x_i)) \underline{u}_h^{(m)}(x_i) \right) \\ & = -\mathcal{D}_h \left(f_u(x_i, \underline{u}_h^{(m-1)}(x_i)) \underline{u}_h^{(m-1)}(x_i) - f(x_i, \underline{u}_h^{(m-1)}(x_i)) \right), \quad 1 \leq i \leq L-1, \quad (5.5) \\ & \underline{u}_h^{(m)}(0) = S_\alpha \left[\underline{u}_h^{(m)}(\xi) \right], \quad \underline{u}_h^{(m)}(1) = S_\beta \left[\underline{u}_h^{(m)}(\eta) \right]. \end{aligned}$$

To show that the sequences given by (5.1) are well-defined and monotone for an arbitrary C^1 -function $f(\cdot, u)$, we let \underline{M}_u be given by (3.13) and let

$$\overline{M}_u = \max_{0 \leq i \leq L} \max\left\{-f_u(x_i, u); u \in [\hat{u}_h(x_i), \tilde{u}_h(x_i)]\right\}. \quad (5.6)$$

Lemma 5.1. *Let the condition (2.9) hold, and let $\tilde{u}_h(x_i)$ and $\hat{u}_h(x_i)$ be a pair of ordered upper and lower solutions of (2.5). Assume that $\underline{M}_u > -8(1 - \sigma)$ and $h < h(\underline{M}_u, \overline{M}_u)$. Then the sequences $\{\overline{u}_h^{(m)}(x_i)\}$, $\{\underline{u}_h^{(m)}(x_i)\}$, and $\{M_i^{(m)}\}$ given by (5.1) and (5.2) with $\overline{u}_h^{(0)}(x_i) = \tilde{u}_h(x_i)$ and $\underline{u}_h^{(0)}(x_i) = \hat{u}_h(x_i)$ are all well defined and possess the monotone property*

$$\underline{u}_h^{(m)}(x_i) \leq \underline{u}_h^{(m+1)}(x_i) \leq \overline{u}_h^{(m+1)}(x_i) \leq \overline{u}_h^{(m)}(x_i), \quad 0 \leq i \leq L, \quad m = 0, 1, \dots \quad (5.7)$$

Proof. Since $M_i^{(0)} = \max\{-f_u(x_i, u); u_i \in [\hat{u}_h(x_i), \tilde{u}_h(x_i)]\}$, $-8(1-\sigma) < \underline{M}_u \leq M_i^{(0)} \leq \overline{M}_u$ and $h < h(\underline{M}_u, \overline{M}_u)$, we have from Lemma 2.2 that the first iterations $\overline{u}_h^{(1)}(x_i)$ and $\underline{u}_h^{(1)}(x_i)$ are well defined. Let $\overline{w}_h^{(0)}(x_i) = \overline{u}_h^{(0)}(x_i) - \underline{u}_h^{(1)}(x_i)$. Then, by (3.1) and (5.1),

$$\begin{aligned} -\delta_h^2 \overline{w}_h^{(0)}(x_i) + \rho_h \left(M_i^{(0)} \overline{w}_h^{(0)}(x_i) \right) &\geq 0, \quad 1 \leq i \leq L-1, \\ \overline{w}_h^{(0)}(0) &\geq S_\alpha \left[\overline{w}_h^{(0)}(\xi) \right], \quad \overline{w}_h^{(0)}(1) \geq S_\beta \left[\overline{w}_h^{(0)}(\eta) \right]. \end{aligned} \quad (5.8)$$

We have from Lemma 2.2 that $\overline{w}_h^{(0)}(x_i) \geq 0$, that is, $\overline{u}_h^{(0)}(x_i) \geq \underline{u}_h^{(1)}(x_i)$ for every $0 \leq i \leq L$. Similarly by the property of a lower solution, $\underline{u}_h^{(1)}(x_i) \geq \underline{u}_h^{(0)}(x_i)$ for every $0 \leq i \leq L$. Let $w_h^{(1)}(x_i) = \overline{u}_h^{(1)}(x_i) - \underline{u}_h^{(1)}(x_i)$. Then by (5.1) and (5.3),

$$\begin{aligned} -\delta_h^2 w_h^{(1)}(x_i) + \rho_h \left(M_i^{(0)} w_h^{(1)}(x_i) \right) &\geq 0, \quad 1 \leq i \leq L-1, \\ w_h^{(1)}(0) &= S_\alpha \left[w_h^{(1)}(\xi) \right], \quad w_h^{(1)}(1) = S_\beta \left[w_h^{(1)}(\eta) \right]. \end{aligned} \quad (5.9)$$

It follows from Lemma 2.2 that $w_h^{(1)}(x_i) \geq 0$, that is, $\overline{u}_h^{(1)}(x_i) \geq \underline{u}_h^{(1)}(x_i)$ for every $0 \leq i \leq L$. This proves the monotone property (5.7) for $m = 0$.

Assume, by induction, that there exists some integer $m_0 \geq 0$ such that for all $0 \leq m \leq m_0$, the iterations $\overline{u}_h^{(m)}(x_i)$, $\underline{u}_h^{(m)}(x_i)$, $\underline{u}_h^{(m)}(x_i)$, and $\underline{u}_h^{(m+1)}(x_i)$ are well-defined and satisfy (5.7). Then $M_i^{(m_0+1)}$ is well defined and $-8(1-\sigma) < \underline{M}_u \leq M_i^{(m_0+1)} \leq \overline{M}_u$. Since $h < h(\underline{M}_u, \overline{M}_u)$, we have from Lemma 2.2 that the iterations $\overline{u}_h^{(m_0+2)}(x_i)$ and $\underline{u}_h^{(m_0+2)}(x_i)$ exist uniquely. Let $\overline{w}_h^{(m_0+1)}(x_i) = \overline{u}_h^{(m_0+1)}(x_i) - \underline{u}_h^{(m_0+2)}(x_i)$. Since

$$\begin{aligned} M_i^{(m_0+1)} \overline{w}_h^{(m_0+1)}(x_i) \\ = \left(M_i^{(m_0+1)} - M_i^{(m_0)} \right) \overline{u}_h^{(m_0+1)}(x_i) + M_i^{(m_0)} \overline{u}_h^{(m_0+1)}(x_i) - M_i^{(m_0+1)} \underline{u}_h^{(m_0+2)}(x_i), \end{aligned} \quad (5.10)$$

the iterative scheme (5.1) implies that

$$\begin{aligned} &-\delta_h^2 \overline{w}_h^{(m_0+1)}(x_i) + \rho_h \left(M_i^{(m_0+1)} \overline{w}_h^{(m_0+1)}(x_i) \right) \\ &= \rho_h \left(M_i^{(m_0)} \left(\overline{u}_h^{(m_0)}(x_i) - \underline{u}_h^{(m_0+1)}(x_i) \right) + f \left(x_i, \overline{u}_h^{(m_0)}(x_i) \right) - f \left(x_i, \underline{u}_h^{(m_0+1)}(x_i) \right) \right), \quad 1 \leq i \leq L-1, \\ &\overline{w}_h^{(m_0+1)}(0) = S_\alpha \left[\overline{w}_h^{(m_0+1)}(\xi) \right], \quad \overline{w}_h^{(m_0+1)}(1) = S_\beta \left[\overline{w}_h^{(m_0+1)}(\eta) \right]. \end{aligned} \quad (5.11)$$

Using the relation (5.3) yields

$$\begin{aligned} -\delta_h^2 \bar{w}_h^{(m_0+1)}(x_i) + \mathcal{D}_h \left(M_i^{(m_0+1)} \bar{w}_h^{(m_0+1)}(x_i) \right) &\geq 0, \quad 1 \leq i \leq L-1, \\ \bar{w}_h^{(m_0+1)}(0) = S_\alpha \left[\bar{w}_h^{(m_0+1)}(\xi) \right], \quad \bar{w}_h^{(m_0+1)}(1) &= S_\beta \left[\bar{w}_h^{(m_0+1)}(\eta) \right]. \end{aligned} \quad (5.12)$$

By Lemma 2.2, $\bar{w}_h^{(m_0+1)}(x_i) \geq 0$, that is, $\bar{u}_h^{(m_0+1)}(x_i) \geq \bar{u}_h^{(m_0+2)}(x_i)$ for every $0 \leq i \leq L$. Similarly we have $\underline{u}_h^{(m_0+2)}(x_i) \geq \underline{u}_h^{(m_0+1)}(x_i)$ for every $0 \leq i \leq L$. Let $w_h^{(m_0+2)}(x_i) = \bar{u}_h^{(m_0+2)}(x_i) - \underline{u}_h^{(m_0+2)}(x_i)$. Then by (5.1) and (5.3), $w_h^{(m_0+2)}(x_i)$ satisfies (5.12) with $\bar{w}_h^{(m_0+1)}(x_i)$ replaced by $w_h^{(m_0+2)}(x_i)$. Therefore, by Lemma 2.2, $w_h^{(m_0+2)}(x_i) \geq 0$, that is, $\bar{u}_h^{(m_0+2)}(x_i) \geq \underline{u}_h^{(m_0+2)}(x_i)$ for every $0 \leq i \leq L$. This shows that the monotone property (5.7) is also true for $m = m_0 + 1$. Finally, the conclusion of the lemma follows from the principle of induction. \square

We next show monotone convergence of the sequences $\{\bar{u}_h^{(m)}(x_i)\}$ and $\{\underline{u}_h^{(m)}(x_i)\}$.

Theorem 5.2. *Let the hypothesis in Lemma 5.1 hold. Then the sequences $\{\bar{u}_h^{(m)}(x_i)\}$ and $\{\underline{u}_h^{(m)}(x_i)\}$ given by (5.1) converge monotonically to the unique solution $u_h^*(x_i)$ of (2.5) in $\langle \hat{u}_h, \tilde{u}_h \rangle$, respectively. Moreover,*

$$\underline{u}_h^{(m)}(x_i) \leq \underline{u}_h^{(m+1)}(x_i) \leq u_h^*(x_i) \leq \bar{u}_h^{(m+1)}(x_i) \leq \bar{u}_h^{(m)}(x_i), \quad 0 \leq i \leq L, \quad m = 0, 1, \dots \quad (5.13)$$

Proof. It follows from the monotone property (5.7) that the limits

$$\lim_{m \rightarrow \infty} \bar{u}_h^{(m)}(x_i) = \bar{u}_h(x_i), \quad \lim_{m \rightarrow \infty} \underline{u}_h^{(m)}(x_i) = \underline{u}_h(x_i), \quad 0 \leq i \leq L \quad (5.14)$$

exist and they satisfy (3.10). Since the sequence $\{M_i^{(m)}\}$ is monotone nonincreasing and is bounded from below by \underline{M}_u given in (3.13), it converges as $m \rightarrow \infty$. Letting $m \rightarrow \infty$ in (5.1) shows that both $\bar{u}_h(x_i)$ and $\underline{u}_h(x_i)$ are solutions of (2.5) in $\langle \hat{u}_h, \tilde{u}_h \rangle$. Since $\underline{M}_u > -8(1 - \sigma)$ and $h < h(\underline{M}_u, \bar{M}_u)$, the condition (3.14) of Theorem 3.3 is satisfied. Thus by Theorem 3.3, $\bar{u}_h(x_i) = \underline{u}_h(x_i) (\equiv u_h^*(x_i))$ and $u_h^*(x_i)$ is the unique solution of (2.5) in $\langle \hat{u}_h, \tilde{u}_h \rangle$. The monotone property (5.13) follows from (3.10). \square

When $f_u(x_i, u)$ is monotone nonincreasing/nondecreasing in $u \in [\underline{u}_h^{(m)}(x_i), \bar{u}_h^{(m)}(x_i)]$ for all $0 \leq i \leq L$, the iterative scheme (5.1) for $\{\bar{u}_h^{(m)}(x_i)\} / \{\underline{u}_h^{(m)}(x_i)\}$ is reduced to Newton iteration (5.5). As a consequence of Theorem 5.2, we have the following conclusion.

Corollary 5.3. *Let the hypothesis in Lemma 5.1 be satisfied. If $f_u(x_i, u)$ is monotone nonincreasing in $u \in [\underline{u}_h^{(m)}(x_i), \bar{u}_h^{(m)}(x_i)]$ for all $0 \leq i \leq L$, the sequence $\{\bar{u}_h^{(m)}(x_i)\}$ given by (5.5) with $\bar{u}_h^{(0)}(x_i) = \tilde{u}_h(x_i)$ converges monotonically from above to the unique solution $u_h^*(x_i)$ of (2.5) in $\langle \hat{u}_h, \tilde{u}_h \rangle$. Otherwise, if $f_u(x_i, u)$ is monotone nondecreasing in $u \in [\underline{u}_h^{(m)}(x_i), \bar{u}_h^{(m)}(x_i)]$ for all $0 \leq i \leq L$, the sequence $\{\underline{u}_h^{(m)}(x_i)\}$ given by (5.5) with $\underline{u}_h^{(0)}(x_i) = \hat{u}_h(x_i)$ converges monotonically from below to $u_h^*(x_i)$.*

5.2. Rate of Convergence

We now show the quadratic rate of convergence of the sequences given by (5.1). Assume that there exists a nonnegative constant \overline{M}^* such that

$$|f_u(x_i, v_i) - f_u(x_i, v'_i)| \leq \overline{M}^* |v_i - v'_i| \quad \forall v_i, v'_i \in [\hat{u}_h(x_i), \tilde{u}_h(x_i)], \quad 0 \leq i \leq L. \quad (5.15)$$

Clearly, this assumption is satisfied if $f(\cdot, u)$ is a C^2 -function of u .

Theorem 5.4. *Let the hypotheses in Lemma 5.1 and (5.15) hold. Also let $\{\overline{u}_h^{(m)}(x_i)\}$ and $\{\underline{u}_h^{(m)}(x_i)\}$ be the sequences given by (5.1) and let $u_h^*(x_i)$ be the unique solution of (2.5) in $\langle \hat{u}_h, \tilde{u}_h \rangle$. Then there exists a constant ρ , independent of m , such that*

$$\left\| \overline{u}_h^{(m)} - u_h^* \right\|_\infty + \left\| \underline{u}_h^{(m)} - u_h^* \right\|_\infty \leq \rho \left(\left\| \overline{u}_h^{(m-1)} - u_h^* \right\|_\infty + \left\| \underline{u}_h^{(m-1)} - u_h^* \right\|_\infty \right)^2, \quad m = 1, 2, \dots \quad (5.16)$$

Proof. Let $\overline{w}_h^{(m)}(x_i) = \overline{u}_h^{(m)}(x_i) - u_h^*(x_i)$. Subtracting (2.5) from (5.1) gives

$$\begin{aligned} & -\delta_h^2 \overline{w}_h^{(m)}(x_i) + \rho_h \left(M_i^{(m-1)} \overline{w}_h^{(m)}(x_i) \right) \\ &= \rho_h \left(M_i^{(m-1)} \overline{w}_h^{(m-1)}(x_i) + f \left(x_i, \overline{u}_h^{(m-1)}(x_i) \right) - f \left(x_i, u_h^*(x_i) \right) \right), \quad 1 \leq i \leq L-1, \quad (5.17) \\ & \overline{w}_h^{(m)}(0) = S_\alpha \left[\overline{w}_h^{(m)}(\xi) \right], \quad \overline{w}_h^{(m)}(1) = S_\beta \left[\overline{w}_h^{(m)}(\eta) \right]. \end{aligned}$$

By the intermediate value theorem,

$$M_i^{(m-1)} = -f_u \left(x_i, \theta_i^{(m-1)} \right), \quad (5.18)$$

where $\theta_i^{(m-1)} \in [\underline{u}_h^{(m-1)}(x_i), \overline{u}_h^{(m-1)}(x_i)]$, and by the mean value theorem,

$$f \left(x_i, \overline{u}_h^{(m-1)}(x_i) \right) - f \left(x_i, u_h^*(x_i) \right) = f_u \left(x_i, \gamma_i^{(m-1)} \right) \overline{w}_h^{(m-1)}(x_i), \quad (5.19)$$

where $\gamma_i^{(m-1)} \in [u_h^*(x_i), \overline{u}_h^{(m-1)}(x_i)]$. Let

$$g_i^{(m-1)} = \left(f_u \left(x_i, \gamma_i^{(m-1)} \right) - f_u \left(x_i, \theta_i^{(m-1)} \right) \right) \overline{w}_h^{(m-1)}(x_i). \quad (5.20)$$

Then we have from (5.17) that

$$\begin{aligned} & -\delta_h^2 \overline{w}_h^{(m)}(x_i) + \rho_h \left(M_i^{(m-1)} \overline{w}_h^{(m)}(x_i) \right) = \rho_h g_i^{(m-1)}, \quad 1 \leq i \leq L-1, \\ & \overline{w}_h^{(m)}(0) = S_\alpha \left[\overline{w}_h^{(m)}(\xi) \right], \quad \overline{w}_h^{(m)}(1) = S_\beta \left[\overline{w}_h^{(m)}(\eta) \right]. \end{aligned} \quad (5.21)$$

Since $-8(1 - \sigma) < \underline{M}_u \leq M_i^{(m-1)} \leq \overline{M}_u$ and $h < h(\underline{M}_u, \overline{M}_u)$, it follows from Lemma 2.3 that there exists a constant ρ_1 , independent of m , such that

$$\left\| \overline{w}_h^{(m)} \right\|_{\infty} \leq \frac{\rho_1 \left\| \rho_h g_i^{(m-1)} \right\|_{\infty}}{h^2}. \quad (5.22)$$

To estimate $g_i^{(m-1)}$, we observe from (5.15) that

$$\left| g_i^{(m-1)} \right| \leq \overline{M}^* \left| \gamma_i^{(m-1)} - \theta_i^{(m-1)} \right| \cdot \left| \overline{w}_h^{(m-1)}(x_i) \right|. \quad (5.23)$$

Since both $\gamma_i^{(m-1)}$ and $\theta_i^{(m-1)}$ are in $[\underline{u}_h^{(m-1)}(x_i), \overline{u}_h^{(m-1)}(x_i)]$, the above estimate implies that

$$\left| g_i^{(m-1)} \right| \leq \overline{M}^* \left| \overline{u}_h^{(m-1)}(x_i) - \underline{u}_h^{(m-1)}(x_i) \right| \cdot \left| \overline{w}_h^{(m-1)}(x_i) \right|. \quad (5.24)$$

Using this estimate in (5.22), we obtain

$$\left\| \overline{w}_h^{(m)} \right\|_{\infty} \leq \rho_1 \overline{M}^* \left\| \overline{u}_h^{(m-1)} - \underline{u}_h^{(m-1)} \right\|_{\infty} \left\| \overline{w}_h^{(m-1)} \right\|_{\infty}, \quad (5.25)$$

or

$$\left\| \overline{u}_h^{(m)} - u_h^* \right\|_{\infty} \leq \rho_1 \overline{M}^* \left\| \overline{u}_h^{(m-1)} - \underline{u}_h^{(m-1)} \right\|_{\infty} \left\| \overline{u}_h^{(m-1)} - u_h^* \right\|_{\infty}. \quad (5.26)$$

Similarly, we have

$$\left\| \underline{u}_h^{(m)} - u_h^* \right\|_{\infty} \leq \rho_1 \overline{M}^* \left\| \overline{u}_h^{(m-1)} - \underline{u}_h^{(m-1)} \right\|_{\infty} \left\| \underline{u}_h^{(m-1)} - u_h^* \right\|_{\infty}. \quad (5.27)$$

Addition of (5.26) and (5.27) gives

$$\begin{aligned} & \left\| \overline{u}_h^{(m)} - u_h^* \right\|_{\infty} + \left\| \underline{u}_h^{(m)} - u_h^* \right\|_{\infty} \\ & \leq \rho_1 \overline{M}^* \left\| \overline{u}_h^{(m-1)} - \underline{u}_h^{(m-1)} \right\|_{\infty} \left(\left\| \overline{u}_h^{(m-1)} - u_h^* \right\|_{\infty} + \left\| \underline{u}_h^{(m-1)} - u_h^* \right\|_{\infty} \right). \end{aligned} \quad (5.28)$$

Then the estimate (5.16) follows immediately. \square

Theorem 5.4 gives a quadratic convergence for the sum of the sequences $\{\overline{u}_h^{(m)}(x_i)\}$ and $\{\underline{u}_h^{(m)}(x_i)\}$ in the sense of (5.16). If $f_u(x_i, u)$ is monotone nonincreasing/nondecreasing in $u \in [\underline{u}_h^{(m)}(x_i), \overline{u}_h^{(m)}(x_i)]$ for all $0 \leq i \leq L$, the sequence $\{\overline{u}_h^{(m)}(x_i)\} / \{\underline{u}_h^{(m)}(x_i)\}$ has the quadratic convergence. This result is stated as follows.

Theorem 5.5. *Let the conditions in Theorem 5.4 hold. Then there exists a constant ρ , independent of m , such that*

$$\left\| \bar{u}_h^{(m)} - u_h^* \right\|_\infty \leq \rho \left\| \bar{u}_h^{(m-1)} - u_h^* \right\|_\infty^2, \quad m = 1, 2, \dots \quad (5.29)$$

if $f_u(x_i, u)$ is monotone nonincreasing in $u \in [\underline{u}_h^{(m)}(x_i), \bar{u}_h^{(m)}(x_i)]$ for all $0 \leq i \leq L$ and

$$\left\| \underline{u}_h^{(m)} - u_h^* \right\|_\infty \leq \rho \left\| \underline{u}_h^{(m-1)} - u_h^* \right\|_\infty^2, \quad m = 1, 2, \dots \quad (5.30)$$

if $f_u(x_i, u)$ is monotone nondecreasing in $u \in [\underline{u}_h^{(m)}(x_i), \bar{u}_h^{(m)}(x_i)]$ for all $0 \leq i \leq L$.

Proof. Consider the monotone nonincreasing case. In this case, the sequence $\{\bar{u}_h^{(m)}(x_i)\}$ is given by (5.5) with $\bar{u}_h^{(0)}(x_i) = \tilde{u}_h(x_i)$. This implies that $\theta_i^{(m-1)} = \bar{u}_h^{(m-1)}(x_i)$, where $\theta_i^{(m-1)}$ is the intermediate value in (5.18). Since $\gamma_i^{(m-1)}$ in (5.19) is in $[u_h^*(x_i), \bar{u}_h^{(m-1)}(x_i)]$, we see that

$$\left| \gamma_i^{(m-1)} - \theta_i^{(m-1)} \right| \leq \left| \bar{u}_h^{(m-1)}(x_i) - u_h^*(x_i) \right|. \quad (5.31)$$

Thus, (5.24) is now reduced to

$$\left| g_i^{(m-1)} \right| \leq \bar{M}^* \left| \bar{w}_h^{(m-1)}(x_i) \right|^2. \quad (5.32)$$

The argument in the proof of Theorem 5.4 shows that (5.29) holds with $\rho = \rho_1 \bar{M}^*$, where ρ_1 is the constant in (5.22). The proof of (5.30) is similar. \square

6. Applications and Numerical Results

In this section, we give some applications of the results in the previous sections to three model problems. We present some numerical results to demonstrate the monotone and rapid convergence of the sequence from (5.1) and to show the fourth-order accuracy of Numerov's scheme (2.5), as predicted in the analysis.

In order to implement the monotone iterative algorithm (5.1), it is necessary to find a pair of ordered upper and lower solutions of (2.5). The construction of this pair depends mainly on the function $f(\cdot, u)$, and much discussion on the subject can be found in [54] for continuous problems. To demonstrate some techniques for the construction of ordered upper and lower solutions of (2.5), we assume that $f(x, 0) \geq 0$ for all $x \in [0, 1]$ and there exists a nonnegative constant C such that

$$f(x, C) \leq 0, \quad x \in [0, 1]. \quad (6.1)$$

Then $-\delta_h^2 C = 0 \geq \mathcal{D}_h f(x_i, C)$ for all $1 \leq i \leq L-1$. This implies that $\tilde{u}_h(x_i) \equiv C$ and $\hat{u}_h(x_i) \equiv 0$ are a pair of ordered upper and lower solutions of (2.5) if, in addition, the condition (2.9) holds. On the other hand, assume that there exist nonnegative constants a, b with $a < 8(1-\sigma)$ such that

$$f(x, u) \leq au + b \quad \text{for } x \in [0, 1], u \geq 0, \quad (6.2)$$

where $\sigma < 1$ is defined by (2.9). We have from Lemma 2.2 that the solution $\tilde{u}_h(x_i)$ of the linear problem

$$\begin{aligned} -\delta_h^2 \tilde{u}_h(x_i) - a \mathcal{D}_h \tilde{u}_h(x_i) &= h^2 b, \quad 1 \leq i \leq L-1, \\ \tilde{u}_h(0) &= S_\alpha[\tilde{u}_h(\xi)], \quad \tilde{u}_h(1) = S_\beta[\tilde{u}_h(\eta)] \end{aligned} \quad (6.3)$$

exists uniquely and is nonnegative. Clearly by (6.2), this solution is a nonnegative upper solution of (2.5).

As applications of the above construction of upper and lower solutions, we next consider three specific examples. In each of these examples, the analytic solution $u(x)$ of (1.1) is explicitly known, against which we can compare the numerical solution $u_h^*(x_i)$ of the scheme (2.5) to demonstrate the fourth-order accuracy of the scheme. The order of accuracy is calculated by

$$\text{error}_\infty(h) = \|u - u_h^*\|_\infty, \quad \text{order}_\infty(h) = \log_2 \left(\frac{\text{error}_\infty(h)}{\text{error}_\infty(h/2)} \right). \quad (6.4)$$

All computations are carried out by using a MATLAB subroutine on a Pentium 4 computer with 2G memory, and the termination criterion of iterations for (5.1) is given by

$$\left\| \bar{u}_h^{(m)} - \underline{u}_h^{(m)} \right\|_\infty < 10^{-14}. \quad (6.5)$$

Example 6.1. Consider the four-point boundary value problem:

$$\begin{aligned} -u''(x) &= \theta u(x)(1-u(x)) + q(x), \quad 0 < x < 1, \\ u(0) &= \frac{1}{9} u\left(\frac{1}{2}\right), \quad u(1) = \frac{1}{8} u\left(\frac{1}{4}\right), \end{aligned} \quad (6.6)$$

where θ is a positive constant and $q(x)$ is a nonnegative continuous function. Clearly, problem (6.6) is a special case of (1.1) with

$$f(x, u) = \theta u(1-u) + q(x). \quad (6.7)$$

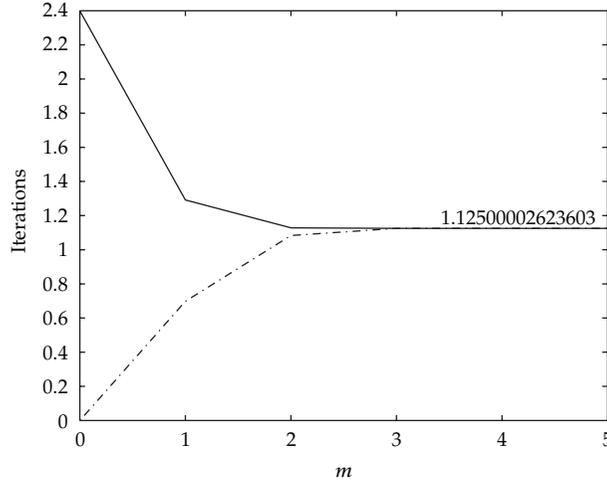


Figure 1: The monotone convergence of $(\{\bar{u}_h^{(m)}(x_i)\}, \{\underline{u}_h^{(m)}(x_i)\})$ at $x_i = 0.5$ for Example 6.1.

To obtain an explicit analytic solution of (6.6), we choose

$$q(x) = 8 + \left(\frac{\pi^2}{2}\right) \sin(2\pi x) - \theta z(x)(1 - z(x)), \quad z(x) = 4x(1 - x) + \frac{1 + \sin(2\pi x)}{8}. \quad (6.8)$$

Then the function $u(x) = z(x)$ is a solution of (6.6). Moreover, $q(x) \geq 0$ in $[0, 1]$ if $\theta \leq 32 - 2\pi^2$.

For problem (6.6), the corresponding Numerov scheme (2.5) is now reduced to

$$\begin{aligned} -\delta_h^2 u_h(x_i) &= \mathcal{P}_h f(x_i, u_h(x_i)), \quad 1 \leq i \leq L - 1, \\ u_h(0) &= \frac{1}{9} u_h\left(\frac{1}{2}\right), \quad u_h(1) = \frac{1}{8} u_h\left(\frac{1}{4}\right). \end{aligned} \quad (6.9)$$

To find a pair of ordered upper and lower solutions of (6.9), we observe from (6.7) that $f(x, 0) = q(x) \geq 0$ for all $x \in [0, 1]$, and, therefore, $\hat{u}_h(x_i) \equiv 0$ is a lower solution. Since $q(x) \leq 14$, we have from (6.7) that the condition (6.2) is satisfied for the present function f with $a = \theta$ and $b = 14$. Therefore, the solution $\tilde{u}_h(x_i)$ of (6.3) (corresponding to (6.9)) with $a = \theta$ and $b = 14$ is a nonnegative upper solution if $\theta < 7$. This implies that $\tilde{u}_h(x_i)$ and $\hat{u}_h(x_i) \equiv 0$ are a pair of ordered upper and lower solutions of (6.9).

Let $\theta = \pi/2$. Using $\bar{u}_h^{(0)}(x_i) = \tilde{u}_h(x_i)$ and $\underline{u}_h^{(0)}(x_i) = 0$, we compute the sequences $\{\bar{u}_h^{(m)}(x_i)\}$ and $\{\underline{u}_h^{(m)}(x_i)\}$ from the iterative scheme (5.1) for (6.9) and various values of h . In all the numerical computations, the basic feature of monotone convergence of the sequences was observed. Let $h = 1/32$. In Figure 1, we present some numerical results of these sequences at $x_i = 0.5$, where the solid line denotes the sequence $\{\bar{u}_h^{(m)}(x_i)\}$ and the dashed-dotted line stands for the sequence $\{\underline{u}_h^{(m)}(x_i)\}$. As described in Theorem 5.2, the sequences converge to the same limit as $m \rightarrow \infty$, and their common limit $u_h^*(x_i)$ is the unique solution of (6.9) in $\langle 0, \tilde{u}_h \rangle$. Besides, these sequences converge rapidly (in five iterations). More numerical results

Table 1: Solutions $u_h^*(x_i)$ and $u(x_i)$ of Example 6.1.

x_i	$u_h^*(x_i)$	$u(x_i)$
1/16	0.40721072351325	0.40721042904564
1/8	0.65088888830290	0.65088834764832
3/16	0.84986064600007	0.84985994156391
1/4	1.00000076215892	1
5/16	1.09986064798991	1.09985994156391
3/8	1.15088889437534	1.15088834764832
7/16	1.15721073688765	1.15721042904564
1/2	1.12500002623603	1.12500000000000

Table 2: The accuracy of the numerical solution $u_h^*(x_i)$ of Example 6.1.

h	Scheme (6.9)		SFD scheme	
	$\text{error}_\infty(h)$	$\text{order}_\infty(h)$	$\text{error}_\infty(h)$	$\text{order}_\infty(h)$
1/4	$3.43422969288754e-03$	4.10451040194224	$2.82813777940529e-02$	2.12179230623150
1/8	$1.99640473104834e-04$	4.02647493758726	$6.49796599949681e-03$	2.03066791458431
1/16	$1.22506422453039e-05$	4.00662171813206	$1.59032351665434e-03$	2.00765985597618
1/32	$7.62158923750533e-07$	4.00165559258019	$3.95475554192615e-04$	2.00191427472146
1/64	$4.75802997002006e-08$	4.00040916222332	$9.87377889736241e-05$	2.00047851945960
1/128	$2.97292546136418e-09$	4.00024505190767	$2.46762611546547e-05$	2.00011961094792
1/256	$1.85776283245787e-10$	4.03319325068531	$6.16855384505399e-06$	2.00002977942424
1/512	$1.13469234008790e-11$		$1.54210662950405e-06$	

of $u_h^*(x_i)$ at various x_i are explicitly given in Table 1. We also list the values of the analytic solution $u(x_i)$. Clearly, the numerical solution $u_h^*(x_i)$ meets the analytic solution $u(x_i)$ closely.

To further demonstrate the accuracy of the numerical solution $u_h^*(x_i)$, we list the maximum error $\text{error}_\infty(h)$ and the order $\text{order}_\infty(h)$ in the first three columns of Table 2 for various values of h . The data demonstrate that the numerical solution $u_h^*(x_i)$ has the fourth-order accuracy. This coincides with the analysis very well.

For comparison, we also solve (6.6) by the standard finite difference (SFD) method. This method leads to a difference scheme in the form (6.9) with $\mathcal{P}_h = \mathcal{O}$ (an identical operator). Thus, a similar iterative scheme as (5.1) can be used in actual computations. The corresponding maximum error $\text{error}_\infty(h)$ and the order $\text{order}_\infty(h)$ are listed in the last two columns of Table 2. We see that the standard finite difference method possesses only the second-order accuracy.

Example 6.2. Our second example is for the following five-point boundary value problem:

$$\begin{aligned}
 -u''(x) &= \frac{\theta u(x)}{1+u(x)} + q(x), \quad 0 < x < 1, \\
 u(0) &= \frac{\sqrt{3}}{6}u\left(\frac{1}{4}\right) + \frac{1}{4}u\left(\frac{1}{2}\right), \quad u(1) = \frac{1}{4}u\left(\frac{1}{2}\right) + \frac{\sqrt{3}}{6}u\left(\frac{3}{4}\right),
 \end{aligned} \tag{6.10}$$

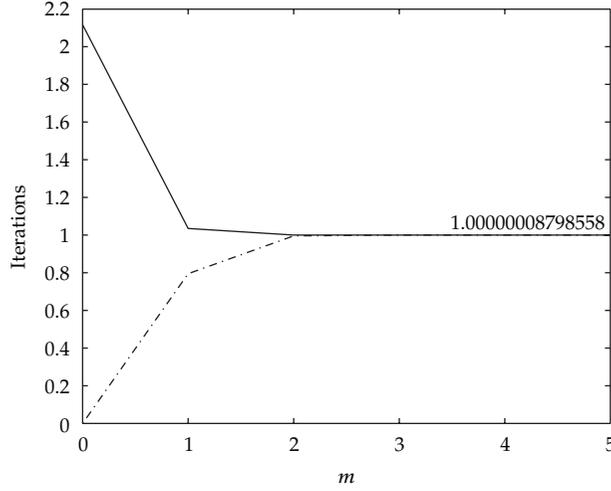


Figure 2: The monotone convergence of $(\{\bar{u}_h^{(m)}(x_i)\}, \{\underline{u}_h^{(m)}(x_i)\})$ at $x_i = 0.5$ for Example 6.2.

where θ is a positive constant and $q(x)$ is a nonnegative continuous function. The corresponding Numerov's scheme (2.5) for this example is given by

$$\begin{aligned}
 -\delta_h^2 u_h(x_i) &= \rho_h f(x_i, u_h(x_i)), \quad 1 \leq i \leq L-1, \\
 u_h(0) &= \frac{\sqrt{3}}{6} u_h\left(\frac{1}{4}\right) + \frac{1}{4} u_h\left(\frac{1}{2}\right), \quad u_h(1) = \frac{1}{4} u_h\left(\frac{1}{2}\right) + \frac{\sqrt{3}}{6} u_h\left(\frac{3}{4}\right),
 \end{aligned} \tag{6.11}$$

where

$$f(x_i, u_h(x_i)) = \frac{\theta u_h(x_i)}{1 + u_h(x_i)} + q(x_i), \quad 0 \leq i \leq L. \tag{6.12}$$

Let

$$q(x) = \left(\kappa^2 - \frac{\theta}{1 + \sin(\kappa x + \pi/6)} \right) \sin\left(\kappa x + \frac{\pi}{6}\right), \quad \kappa = \frac{2\pi}{3}. \tag{6.13}$$

Then $q(x) \geq 0$ in $[0, 1]$ if $\theta \leq 3\kappa^2/2$, and $u(x) = \sin(\kappa x + \pi/6)$ is a solution of (6.10). Clearly, $\hat{u}_h(x_i) \equiv 0$ is a lower solution of (6.11). On the other hand, the condition (6.2) is satisfied for the present problem with $a = \theta$ and $b = \kappa^2$. Therefore, the solution $\tilde{u}_h(x_i)$ of (6.3) (corresponding to (6.11)) with $a = \theta$ and $b = \kappa^2$ is a nonnegative upper solution of (6.11) if $\theta < 2(9 - 2\sqrt{3})/3$.

Let $\theta = \kappa$. Using $\bar{u}_h^{(0)}(x_i) = \tilde{u}_h(x_i)$ and $\underline{u}_h^{(0)}(x_i) = 0$, we compute the sequences $\{\bar{u}_h^{(m)}(x_i)\}$ and $\{\underline{u}_h^{(m)}(x_i)\}$ from the iterative scheme (5.1) for (6.11). Let $h = 1/32$. Some numerical results of these sequences at $x_i = 0.5$ are plotted in Figure 2, where the solid line denotes the sequence $\{\bar{u}_h^{(m)}(x_i)\}$ and the dashed-dotted line stands for the sequence $\{\underline{u}_h^{(m)}(x_i)\}$. We see that the sequences possess the monotone convergence given in

Table 3: The accuracy of the numerical solution $u_h^*(x_i)$ of Example 6.2.

h	Scheme (6.11)		SFD scheme	
	$\text{error}_\infty(h)$	$\text{order}_\infty(h)$	$\text{error}_\infty(h)$	$\text{order}_\infty(h)$
1/4	3.64212838788403e-04	4.01156400486606	2.66151346177448e-02	2.01341045503011
1/8	2.25815711794031e-05	4.00292832787037	6.59222051766362e-03	2.00338885776990
1/16	1.40848640284297e-06	4.00073477910484	1.64418842865244e-03	2.00084988143279
1/32	8.79855768243232e-08	4.00018946428474	4.10805033531858e-04	2.00021264456991
1/64	5.49837642083162e-09	3.99948214984524	1.02686121950857e-04	2.00005317293630
1/128	3.43771899835588e-10	3.99684542351672	2.56705843380001e-05	2.00001326834657
1/256	2.15327755626049e-11	4.03735960917218	6.41758706221296e-06	2.00000333671353
1/512	1.31139543668724e-12		1.60439305485482e-06	

Theorem 5.2 and converge rapidly (in five iterations) to the unique solution $u_h^*(x_i)$ of (6.11) in $\langle 0, \tilde{u}_h \rangle$. The maximum error $\text{error}_\infty(h)$ and the order $\text{order}_\infty(h)$ of the numerical solution $u_h^*(x_i)$ by the scheme (6.11) and the SFD scheme are presented in Table 3. The numerical results clearly indicate that the proposed scheme (6.11) is more efficient than the SFD scheme.

Example 6.3. Our last example is given by

$$\begin{aligned}
 -u''(x) &= \theta \left(q^4(x) - u^4(x) \right), \quad 0 < x < 1, \\
 u(0) &= \frac{\sqrt{2}}{8} u\left(\frac{1}{8}\right) + \frac{\sqrt{3}}{12} u\left(\frac{1}{4}\right) + \frac{1}{4} u\left(\frac{1}{2}\right), \\
 u(1) &= \frac{\sqrt{3}}{12} u\left(\frac{1}{4}\right) + \frac{1}{4} u\left(\frac{1}{2}\right) + \frac{\sqrt{3}}{12} u\left(\frac{3}{4}\right),
 \end{aligned} \tag{6.14}$$

where θ is a positive constant and $q(x)$ is a continuous function. For this example, the corresponding Numerov scheme (2.5) is reduced to

$$\begin{aligned}
 -\delta_h^2 u_h(x_i) &= \mathcal{D}_h f(x_i, u_h(x_i)), \quad 1 \leq i \leq L-1, \\
 u_h(0) &= \frac{\sqrt{2}}{8} u_h\left(\frac{1}{8}\right) + \frac{\sqrt{3}}{12} u_h\left(\frac{1}{4}\right) + \frac{1}{4} u_h\left(\frac{1}{2}\right), \\
 u_h(1) &= \frac{\sqrt{3}}{12} u_h\left(\frac{1}{4}\right) + \frac{1}{4} u_h\left(\frac{1}{2}\right) + \frac{\sqrt{3}}{12} u_h\left(\frac{3}{4}\right),
 \end{aligned} \tag{6.15}$$

where

$$f(x_i, u_h(x_i)) = \theta \left(q^4(x_i) - u_h^4(x_i) \right), \quad 0 \leq i \leq L. \tag{6.16}$$

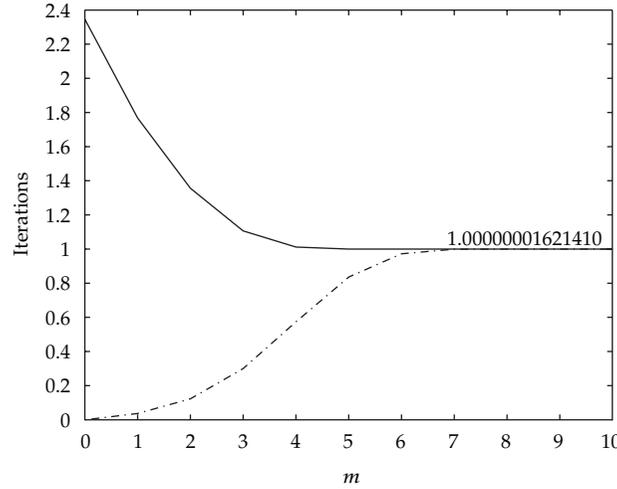


Figure 3: The monotone convergence of $(\{\bar{u}_h^{(m)}(x_i)\}, \{\underline{u}_h^{(m)}(x_i)\})$ at $x_i = 0.5$ for Example 6.3

Table 4: Table 1 The accuracy of the numerical solution $u_h^*(x_i)$ of Example 6.3.

h	Scheme (6.15)		SFD scheme	
	$\text{error}_\infty(h)$	$\text{order}_\infty(h)$	$\text{error}_\infty(h)$	$\text{order}_\infty(h)$
1/8	$4.16044850615194e-06$	4.00262945151917	$1.20005715017424e-03$	1.99011955907584
1/16	$2.59554536974349e-07$	4.00071651350112	$3.02076017246300e-04$	1.99756596363740
1/32	$1.62141038373420e-08$	4.00019299964688	$7.56465233968662e-05$	1.99939372559153
1/64	$1.01324593160257e-09$	4.00017895476307	$1.89195798938613e-05$	1.99984571226086
1/128	$6.33200158972613e-11$	4.00592116975355	$4.73040083492915e-06$	1.99995314107427
1/256	$3.94129173741931e-12$	4.06977174278409	$1.18263862036727e-06$	1.99999043828107
1/512	$2.34701147405758e-13$		$2.95661614635456e-07$	

To accommodate the analytical solution of $u(x) = \sin(\kappa x + \pi/6)$ where $\kappa = 2\pi/3$, we let

$$q(x) = \left(\frac{\kappa^2}{\theta} \sin\left(\kappa x + \frac{\pi}{6}\right) + \sin^4\left(\kappa x + \frac{\pi}{6}\right) \right)^{1/4}. \tag{6.17}$$

As in the previous examples, $\hat{u}_h(x_i) \equiv 0$ is a lower solution of (6.15) and the solution $\tilde{u}_h(x_i)$ of (6.3) (corresponding to (6.15)) with $a = 0$ and $b = \kappa^2 + \theta$ is a nonnegative upper solution.

Let $\theta = \pi^2/2$. We compute the corresponding sequences $\{\bar{u}_h^{(m)}(x_i)\}$ and $\{\underline{u}_h^{(m)}(x_i)\}$ from the iterative scheme (5.1) with the initial iterations $\bar{u}_h^{(0)}(x_i) = \tilde{u}_h(x_i)$ and $\underline{u}_h^{(0)}(x_i) = 0$. Let $h = 1/32$. Figure 3 shows the monotone and rapid convergence of these sequences at $x_i = 0.5$, where the solid line denotes the sequence $\{\bar{u}_h^{(m)}(x_i)\}$ and the dashed-dotted line stands for the sequence $\{\underline{u}_h^{(m)}(x_i)\}$ as before. The data in Table 4 show the maximum error $\text{error}_\infty(h)$ and the order $\text{order}_\infty(h)$ of the numerical solution $u_h^*(x_i)$ by the scheme (6.15) and the SFD scheme for various values of h . The fourth-order accuracy of the numerical solution $u_h^*(x_i)$ by the present Numerov scheme is demonstrated in this table.

7. Conclusions

In this paper, we have given a numerical treatment for a class of nonlinear multipoint boundary value problems by the fourth-order Numerov method. The existence and uniqueness of the numerical solution and the convergence of the method have been discussed. An accelerated monotone iterative algorithm with the quadratic rate of convergence has been developed for solving the resulting nonlinear discrete problem. The proposed Numerov method is more attractive due to its fourth-order accuracy, compared to the standard finite difference method.

In this work, we have generalized the method of upper and lower solutions to nonlinear multipoint boundary value problems. We have also developed a technique for designing and analyzing compact and monotone finite difference schemes with high accuracy. They are very useful for accurate numerical simulations of many other nonlinear problems, such as those related to integrodifferential equations (e.g., [55, 56]) and those in information modeling (e.g., [57–60]).

Appendix

A. Proofs of Lemmas 2.2 and 2.3

Lemmas 2.2 and 2.3 are the special cases of Lemmas 2.2 and 2.3 in [61]. We include their proofs here in order to make the paper self-contained. Define

$$\alpha_i^* = \begin{cases} \alpha_{i'}, & x_i = \xi_{i'} \text{ for some } i', \\ 0, & \text{otherwise,} \end{cases} \quad \beta_i^* = \begin{cases} \beta_{i'}, & x_i = \eta_{i'} \text{ for some } i', \\ 0, & \text{otherwise,} \end{cases} \quad 1 \leq i \leq L-1. \quad (\text{A.1})$$

Let $A = (a_{i,j})$, $B = (b_{i,j})$, and $D = (d_{i,j})$ be the $(L-1)$ th-order matrices with

$$a_{i,j} = 2\delta_{i,j} - \delta_{i,j-1} - \delta_{i,j+1}, \quad b_{i,j} = \frac{5}{6}\delta_{i,j} + \frac{1}{12}\delta_{i,j-1} + \frac{1}{12}\delta_{i,j+1}, \quad d_{i,j} = \delta_{i,1}\alpha_j^* + \delta_{i,L-1}\beta_j^*, \quad (\text{A.2})$$

where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$.

Lemma A.1. *Let the condition (2.9) be satisfied. Then the inverse $(A - D)^{-1} > 0$, and*

$$\|(A - D)^{-1}\|_\infty \leq \frac{1}{8(1 - \sigma)h^2}. \quad (\text{A.3})$$

Proof. It can be checked by Corollary 3.20 on Page 91 of [62] that the inverse $(A - D)^{-1} > 0$. Let $E = (1, 1, \dots, 1)^T \in \mathbf{R}^{L-1}$ and let $S = (A - D)^{-1}E$. Then $\|(A - D)^{-1}\|_\infty = \|S\|_\infty$. It is known

that the inverse $A^{-1} = (J_{i,j})$ exists, and its elements $J_{i,j}$ are given by

$$J_{i,j} = \begin{cases} \frac{(L-j)i}{L}, & i \leq j, \\ \frac{(L-i)j}{L}, & i > j. \end{cases} \quad (\text{A.4})$$

A simple calculation shows that $\|A^{-1}\|_{\infty} \leq L^2/8 = 1/(8h^2)$ and $J_{i,1} + J_{i,L-1} = 1$ for each $1 \leq i \leq L-1$. This implies

$$S = A^{-1}E + A^{-1}DS \leq \|A^{-1}\|_{\infty} E + \sigma \|S\|_{\infty} E \leq \left(\frac{1}{8h^2}\right)E + \sigma \|S\|_{\infty} E. \quad (\text{A.5})$$

Thus the estimate (A.3) follows immediately. \square

Proof of Lemma 2.2. Define the following $(L-1)$ th-order matrices or vectors:

$$\begin{aligned} U_h &= (u_h(x_1), u_h(x_2), \dots, u_h(x_{L-1}))^T, \\ M &= \text{diag}(M_1, M_2, \dots, M_{L-1}), \quad M_b = \text{diag}(M_0, 0, \dots, 0, M_L), \\ G_b &= \left(\left(1 - \frac{h^2}{12}M_0\right)u_h(0), 0, \dots, 0, \left(1 - \frac{h^2}{12}M_L\right)u_h(1) \right)^T. \end{aligned} \quad (\text{A.6})$$

Using the matrices A and B defined by (A.2), we have from (2.11) that

$$(A + h^2BM)U_h \geq G_b. \quad (\text{A.7})$$

Since $\underline{M} > -8(1 - \sigma)$ and $h < h(\underline{M}, \overline{M})$, it is easy to check that $1 - (h^2/12)M_i \geq 0 (i = 0, L)$. Thus by the boundary condition in (2.11),

$$G_b \geq DU_h - \frac{h^2}{12}M_bDU_h, \quad (\text{A.8})$$

where D is the $(L-1)$ th-order matrix defined by (A.2). This leads to

$$\left(A - D + h^2BM + \frac{h^2}{12}M_bD \right) U_h \geq 0. \quad (\text{A.9})$$

Let $Q \equiv A - D + h^2BM + (h^2/12)M_bD$. To prove $u_h(x_i) \geq 0$ for all $0 \leq i \leq L$, it suffices to show that the inverse of Q exists and is nonnegative.

Case 1 ($\underline{M} \geq 0$). In this case, the matrix Q satisfies the condition of Corollary 3.20 on Page 91 of [62], and, therefore, its inverse Q^{-1} exists and is positive.

Case 2 ($0 > \underline{M} > -8(1 - \sigma)$). For this case, we define

$$M^+ = \text{diag}(M_1^+, M_2^+, \dots, M_{L-1}^+), \quad M_i^+ = \max\{M_i, 0\}, \quad M^- = M - M^+. \quad (\text{A.10})$$

The matrices M_b^+ and M_b^- can be similarly defined. Let $\bar{Q} = A - D + h^2 BM^+ + (h^2/12)M_b^+ D$. We know from Case 1 that \bar{Q}^{-1} exists and is positive. Since

$$Q = \bar{Q} + h^2 BM^- + \frac{h^2}{12} M_b^- D = \bar{Q} \left(I + h^2 \bar{Q}^{-1} \left(BM^- + \frac{1}{12} M_b^- D \right) \right), \quad (\text{A.11})$$

we need only to prove that the inverse $(I + h^2 \bar{Q}^{-1} (BM^- + (1/12)M_b^- D))^{-1}$ exists and is nonnegative. By Theorem 3 on Page 298 of [63], this is true if

$$\left\| h^2 \bar{Q}^{-1} \left(BM^- + \frac{1}{12} M_b^- D \right) \right\|_\infty < 1. \quad (\text{A.12})$$

Since $\bar{Q} \geq A - D$ which implies $0 \leq \bar{Q}^{-1} \leq (A - D)^{-1}$, we have from Lemma A.1 that

$$\left\| \bar{Q}^{-1} \right\|_\infty \leq \left\| (A - D)^{-1} \right\|_\infty \leq \frac{1}{8(1 - \sigma)h^2}. \quad (\text{A.13})$$

It is clear that $\|B + (1/12)D\|_\infty = 1$, $\|M^-\|_\infty \leq -\underline{M}$ and $\|M_b^-\|_\infty \leq -\underline{M}$. Thus, we have

$$\left\| h^2 \bar{Q}^{-1} \left(BM^- + \frac{1}{12} M_b^- D \right) \right\|_\infty \leq \frac{-\underline{M}}{8(1 - \sigma)}. \quad (\text{A.14})$$

The estimate (A.12) follows from $\underline{M} > -8(1 - \sigma)$. □

Proof of Lemma 2.3. Using the same notation as before, the system (2.12) can be written as

$$QU_h = G, \quad (\text{A.15})$$

where $G = (g(x_1), g(x_2), \dots, g(x_{L-1}))^T$.

Case 1 ($\underline{M} \geq 0$). Since the inverse Q^{-1} exists and is positive, we have $0 < Q^{-1} \leq (A - D)^{-1}$. This shows

$$\left\| Q^{-1} \right\|_\infty \leq \left\| (A - D)^{-1} \right\|_\infty \leq \frac{1}{8(1 - \sigma)h^2}. \quad (\text{A.16})$$

Thus, by (A.15), $\|U_h\|_\infty \leq \|G\|_\infty / (8(1 - \sigma)h^2)$ which implies the desired estimate (2.13).

Case 2 ($0 > \underline{M} > -8(1 - \sigma)$). It follows from (A.11) that

$$\|Q^{-1}\|_{\infty} \leq \|\overline{Q}^{-1}\|_{\infty} \left\| \left(I + h^2 \overline{Q}^{-1} \left(BM^{-} + \frac{1}{12} M_b^{-} D \right) \right)^{-1} \right\|_{\infty}. \quad (\text{A.17})$$

By (A.13) and (A.14),

$$\|Q^{-1}\|_{\infty} \leq \frac{1}{8(1 - \sigma)h^2} \cdot \frac{8(1 - \sigma)}{8(1 - \sigma) + \underline{M}} = \frac{1}{(8(1 - \sigma) + \underline{M})h^2}. \quad (\text{A.18})$$

This together with (A.15) leads to the estimate (2.13). \square

Acknowledgments

The authors would like to thank the editor and the referees for their valuable comments and suggestions which improved the presentation of the paper. This work was supported in part by the National Natural Science Foundation of China No. 10571059, E-Institutes of Shanghai Municipal Education Commission No. E03004, the Natural Science Foundation of Shanghai No. 10ZR1409300, and Shanghai Leading Academic Discipline Project No. B407.

References

- [1] Y. Zou, Q. Hu, and R. Zhang, "On numerical studies of multi-point boundary value problem and its fold bifurcation," *Applied Mathematics and Computation*, vol. 185, no. 1, pp. 527–537, 2007.
- [2] J. Adem and M. Moshinsky, "Self-adjointness of a certain type of vectorial boundary value problems," *Boletín de la Sociedad Matemática Mexicana*, vol. 7, pp. 1–17, 1950.
- [3] M. S. Berger and L. E. Fraenkel, "Nonlinear desingularization in certain free-boundary problems," *Communications in Mathematical Physics*, vol. 77, no. 2, pp. 149–172, 1980.
- [4] M. Greguš, F. Neuman, and F. M. Arscott, "Three-point boundary value problems in differential equations," *Journal of the London Mathematical Society*, vol. 2, pp. 429–436, 1971.
- [5] C.-Z. Bai and J.-X. Fang, "Existence of multiple positive solutions for nonlinear m -point boundary-value problems," *Applied Mathematics and Computation*, vol. 140, no. 2-3, pp. 297–305, 2003.
- [6] P. W. Eloe and Y. Gao, "The method of quasilinearization and a three-point boundary value problem," *Journal of the Korean Mathematical Society*, vol. 39, no. 2, pp. 319–330, 2002.
- [7] W. Feng and J. R. L. Webb, "Solvability of m -point boundary value problems with nonlinear growth," *Journal of Mathematical Analysis and Applications*, vol. 212, no. 2, pp. 467–480, 1997.
- [8] Y. Guo, W. Shan, and W. Ge, "Positive solutions for second-order m -point boundary value problems," *Journal of Computational and Applied Mathematics*, vol. 151, no. 2, pp. 415–424, 2003.
- [9] C. P. Gupta, "A Dirichlet type multi-point boundary value problem for second order ordinary differential equations," *Nonlinear Analysis*, vol. 26, no. 5, pp. 925–931, 1996.
- [10] R. Ma, "Positive solutions of a nonlinear m -point boundary value problem," *Computers & Mathematics with Applications*, vol. 42, no. 6-7, pp. 755–765, 2001.
- [11] B. Liu, "Solvability of multi-point boundary value problem at resonance (II)," *Applied Mathematics and Computation*, vol. 136, no. 2-3, pp. 353–377, 2003.
- [12] X. Liu, J. Qiu, and Y. Guo, "Three positive solutions for second-order m -point boundary value problems," *Applied Mathematics and Computation*, vol. 156, no. 3, pp. 733–742, 2004.
- [13] R. Ma, "Multiplicity results for a three-point boundary value problem at resonance," *Nonlinear Analysis*, vol. 53, no. 6, pp. 777–789, 2003.
- [14] R. Ma, "Positive solutions for nonhomogeneous m -point boundary value problems," *Computers & Mathematics with Applications*, vol. 47, no. 4-5, pp. 689–698, 2004.

- [15] B. Liu and J. Yu, "Solvability of multi-point boundary value problem at resonance (III)," *Applied Mathematics and Computation*, vol. 129, no. 1, pp. 119–143, 2002.
- [16] W. Feng and J. R. L. Webb, "Solvability of three point boundary value problems at resonance," *Nonlinear Analysis*, vol. 30, no. 6, pp. 3227–3238, 1997.
- [17] Z. Zhang and J. Wang, "Positive solutions to a second order three-point boundary value problem," *Journal of Mathematical Analysis and Applications*, vol. 285, no. 1, pp. 237–249, 2003.
- [18] V. A. Il'in and E. I. Moiseev, "Nonlocal boundary value problem of the first kind for a Sturm-Liouville operator in its differential and finite difference aspects," *Differential Equations*, vol. 23, no. 7, pp. 803–810, 1987.
- [19] V. A. Il'in and E. I. Moiseev, "Nonlocal boundary value problem of the second kind for a Sturm-Liouville operator," *Differential Equations*, vol. 23, no. 8, pp. 979–987, 1987.
- [20] C. P. Gupta, "Solvability of a three-point nonlinear boundary value problem for a second order ordinary differential equation," *Journal of Mathematical Analysis and Applications*, vol. 168, no. 2, pp. 540–551, 1992.
- [21] C. P. Gupta, "A generalized multi-point boundary value problem for second order ordinary differential equations," *Applied Mathematics and Computation*, vol. 89, no. 1–3, pp. 133–146, 1998.
- [22] B. Liu, "Solvability of multi-point boundary value problem at resonance. Part IV," *Applied Mathematics and Computation*, vol. 143, no. 2–3, pp. 275–299, 2003.
- [23] Z. Zhang and J. Wang, "The upper and lower solution method for a class of singular nonlinear second order three-point boundary value problems," *Journal of Computational and Applied Mathematics*, vol. 147, no. 1, pp. 41–52, 2002.
- [24] C. P. Gupta, S. K. Ntouyas, and P. Ch. Tsamatos, "On an m -point boundary-value problem for second-order ordinary differential equations," *Nonlinear Analysis*, vol. 23, no. 11, pp. 1427–1436, 1994.
- [25] W. Feng, "On a m -point nonlinear boundary value problem," *Nonlinear Analysis*, vol. 30, pp. 5369–5374, 1997.
- [26] R. Ma, "Existence theorems for a second order m -point boundary value problem," *Journal of Mathematical Analysis and Applications*, vol. 211, no. 2, pp. 545–555, 1997.
- [27] S. A. Marano, "A remark on a second-order three-point boundary value problem," *Journal of Mathematical Analysis and Applications*, vol. 183, no. 3, pp. 518–522, 1994.
- [28] C. P. Gupta, "A second order m -point boundary value problem at resonance," *Nonlinear Analysis*, vol. 24, no. 10, pp. 1483–1489, 1995.
- [29] C. P. Gupta, "Existence theorems for a second order m -point boundary value problem at resonance," *International Journal of Mathematics and Mathematical Sciences*, vol. 18, no. 4, pp. 705–710, 1995.
- [30] M. Urabe, "Numerical solution of multi-point boundary value problems in Chebyshev series theory of the method," *Numerische Mathematik*, vol. 9, pp. 341–366, 1967.
- [31] M. Lentini and V. Pereyra, "A variable order finite difference method for nonlinear multipoint boundary value problems," *Mathematics of Computation*, vol. 28, pp. 981–1003, 1974.
- [32] F. R. de Hoog and R. M. M. Mattheij, "An algorithm for solving multipoint boundary value problems," *Computing*, vol. 38, no. 3, pp. 219–234, 1987.
- [33] R. P. Agarwal, "The numerical solution of multipoint boundary value problems," *Journal of Computational and Applied Mathematics*, vol. 5, no. 1, pp. 17–24, 1979.
- [34] T. Ojika and W. Welsh, "A numerical method for multipoint boundary value problems with application to a restricted three-body problem," *International Journal of Computer Mathematics*, vol. 8, no. 4, pp. 329–344, 1980.
- [35] W. Welsh and T. Ojika, "Multipoint boundary value problems with discontinuities. I. Algorithms and applications," *Journal of Computational and Applied Mathematics*, vol. 6, no. 2, pp. 133–143, 1980.
- [36] R. P. Agarwal, "On multi-point boundary value problems for discrete equations," *Journal of Mathematical Analysis and Applications*, vol. 96, pp. 520–534, 1983.
- [37] R. P. Agarwal and R. A. Usmani, "The formulation of invariant imbedding method to solve multipoint discrete boundary value problems," *Applied Mathematics Letters*, vol. 4, no. 4, pp. 17–22, 1991.
- [38] R. C. Gupta and R. P. Agarwal, "A new shooting method for multipoint discrete boundary value problems," *Journal of Mathematical Analysis and Applications*, vol. 112, no. 1, pp. 210–220, 1985.
- [39] F. Geng, "Solving singular second order three-point boundary value problems using reproducing kernel Hilbert space method," *Applied Mathematics and Computation*, vol. 215, no. 6, pp. 2095–2102, 2009.
- [40] R. P. Agarwal, "On Nomerov's method for solving two point boundary value problems," *Utilitas Mathematica*, vol. 28, pp. 159–174, 1985.

- [41] R. P. Agarwal, *Difference Equations and Inequalities*, vol. 155 of *Monographs and Textbooks in Pure and Applied Mathematics*, Marcel Dekker, New York, NY, USA, 1992.
- [42] L. K. Bieniasz, "Use of the Numerov method to improve the accuracy of the spatial discretization in finite-difference electrochemical kinetic simulations," *Journal of Computational Chemistry*, vol. 26, pp. 633–644, 2002.
- [43] M. M. Chawla and P. N. Shivakumar, "Numerov's method for non-linear two-point boundary value problems," *International Journal of Computer Mathematics*, vol. 17, pp. 167–176, 1985.
- [44] L. Collatz, *The Numerical Treatment of Differential Equations*, Springer, Berlin, Germany, 3rd edition, 1960.
- [45] Y. Fang and X. Wu, "A trigonometrically fitted explicit Numerov-type method for second-order initial value problems with oscillating solutions," *Applied Numerical Mathematics*, vol. 58, no. 3, pp. 341–351, 2008.
- [46] T. E. Simos, "A Numerov-type method for the numerical solution of the radial Schrödinger equation," *Applied Numerical Mathematics*, vol. 7, no. 2, pp. 201–206, 1991.
- [47] Y.-M. Wang, "Monotone methods for a boundary value problem of second-order discrete equation," *Computers & Mathematics with Applications*, vol. 36, no. 6, pp. 77–92, 1998.
- [48] Y.-M. Wang, "Numerov's method for strongly nonlinear two-point boundary value problems," *Computers & Mathematics with Applications*, vol. 45, no. 4-5, pp. 759–763, 2003.
- [49] Y.-M. Wang, "The extrapolation of Numerov's scheme for nonlinear two-point boundary value problems," *Applied Numerical Mathematics*, vol. 57, no. 3, pp. 253–269, 2007.
- [50] Y.-M. Wang and B.-Y. Guo, "On Numerov scheme for nonlinear two-points boundary value problem," *Journal of Computational Mathematics*, vol. 16, no. 4, pp. 345–366, 1998.
- [51] Y.-M. Wang and R. P. Agarwal, "Monotone methods for solving a boundary value problem of second order discrete system," *Mathematical Problems in Engineering*, vol. 5, no. 4, pp. 291–315, 1999.
- [52] B. V. Numerov, "A method of extrapolation of perturbations," *Monthly Notices of the Royal Astronomical Society*, vol. 84, pp. 592–601, 1924.
- [53] R. P. Agarwal and Y.-M. Wang, "The recent developments of Numerov's method," *Computers & Mathematics with Applications*, vol. 42, no. 3–5, pp. 561–592, 2001.
- [54] C. V. Pao, *Nonlinear Parabolic and Elliptic Equations*, Plenum Press, New York, NY, USA, 1992.
- [55] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, vol. 2010, Article ID 408418, 22 pages, 2010.
- [56] 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.
- [57] M. Li, S. C. Lim, and S.-Y. 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.
- [58] 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.
- [59] 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.
- [60] S.-Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 480–487, 2011.
- [61] Y.-M. Wang, W.-J. Wu, and R. P. Agarwal, "A fourth-order compact finite difference method for nonlinear higher-order multi-point boundary value problems," *Computers & Mathematics with Applications*, vol. 61, no. 11, pp. 3226–3245, 2011.
- [62] R. S. Varga, *Matrix Iterative Analysis*, vol. 27 of *Springer Series in Computational Mathematics*, Springer, Berlin, Germany, 2nd edition, 2000.
- [63] L. Collatz, *Funktionalanalysis und Numerische Mathematik*, Springer, Berlin, Germany, 1964.

Research Article

An Efficient Approach to Solve the Large-Scale Semidefinite Programming Problems

**Yongbin Zheng,¹ Yuzhuang Yan,¹ Sheng Liu,² Xinsheng Huang,¹
and Wanying Xu¹**

¹ College of Mechatronics and Automation, National University of Defense Technology,
Changsha 410073, China

² College of Computer Science and Technology, Zhejiang University of Technology,
Hangzhou 310023, China

Correspondence should be addressed to Yongbin Zheng, zybnudt@nudt.edu.cn

Received 21 March 2011; Accepted 17 April 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Yongbin Zheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Solving the large-scale problems with semidefinite programming (SDP) constraints is of great importance in modeling and model reduction of complex system, dynamical system, optimal control, computer vision, and machine learning. However, existing SDP solvers are of large complexities and thus unavailable to deal with large-scale problems. In this paper, we solve SDP using matrix generation, which is an extension of the classical column generation. The exponentiated gradient algorithm is also used to solve the special structure subproblem of matrix generation. The numerical experiments show that our approach is efficient and scales very well with the problem dimension. Furthermore, the proposed algorithm is applied for a clustering problem. The experimental results on real datasets imply that the proposed approach outperforms the traditional interior-point SDP solvers in terms of efficiency and scalability.

1. Introduction

Semidefinite programming (SDP) is a technique widely used in modeling of complex systems and some important issues in computer vision and machine learning. Examples of application include model reduction [1], modeling of nonlinear systems [2, 3], optimal control [4], clustering [5–7], robust Euclidean embedding [8], kernel matrix learning [9], and metric learning [10]. It can be seen in [5, 7] that SDP relaxation can produce more accurate estimates than spectral methods. However, the SDP optimization suffers extremely high time complexity. Current SDP solvers utilizing interior-point methods scale as $O(n^{4.5})$ or worse

[1, 6, 10], where n is the number of rows (or columns) of the semidefinite matrix. As a result, SDP can only run on small datasets. It is meaningful to propose a scalable SDP solver.

Inspired by column generation [11], which is a classical technology in optimization literature for solving large scale problem, Shen et al. proposed matrix generation in [10] for solving a semidefinite metric learning problem. By using matrix generation, the particular semidefinite program in metric learning was converted into a sequence of linear programs, and it is possible to use the well-developed linear programming technology.

In this paper, we propose the matrix generation-based iteration approach to solve general SDP optimization problems. At each iteration, the exponentiated gradient (EG) algorithm [12] is applied to solve the special structure subproblem. Experiments are also carried out on some real datasets to evaluate the proposed method's efficacy and efficiency.

The method proposed here can be seen as an extension of column generation to solve SDP problems. The proposed matrix generation method also has the drawback of column generation. It converges slowly if one wants to achieve high accuracy. This is the so-called "tailing effect". In practice, for many applications, we do not need a very accurate solution. Typically, a moderately accurate solution suffices. On the other hand, the proposed method can also be considered as a generalization of EG to the matrix case. EG is used to solve problems with a vector optimization variable, and the vector must be on a simplex. Here, we generalize EG in the sense that the proposed method solves optimization with a semidefinite matrix whose eigenvalues are on a simplex.

We present our main results in the next section.

2. The Algorithm

2.1. Notation

The following notations will be used throughout the paper:

- (i) a bold lower-case letter (\mathbf{x}): a column vector,
- (ii) an upper-case letter (X): a matrix,
- (iii) $A \succcurlyeq 0$: a positive semidefinite (p.s.d.) matrix,
- (iv) $\mathbf{a} \succcurlyeq \mathbf{b}$: the component-wise inequality between two vectors,
- (v) $\mathbb{R}^{m \times n}$: the vector space of real matrices of size $m \times n$,
- (vi) \mathbb{S} : the space of real matrices,
- (vii) \mathbb{S}^n : the space of symmetric matrices of size $n \times n$,
- (viii) \mathbb{S}_+^n : the space of symmetric positive semidefinite matrices of size $n \times n$,
- (ix) $\langle A, B \rangle = \text{Tr}(A^\top B)$: the inner product defined on the above spaces,
- (x) $\text{Tr}(\cdot)$: the trace of a matrix,
- (xi) $\Delta_n := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \succcurlyeq 0, \mathbf{1}^\top \mathbf{x} = 1\}$: the n -dimensional simplex set for vectors,
- (xii) $\mathcal{D}_n := \{X \in \mathbb{S}_+^n \mid X \succcurlyeq 0, \text{Tr}(X) = 1\}$: the density matrix set,
- (xiii) $\mathcal{Q}_n := \{X \in \mathbb{S}_+^n \mid X \succcurlyeq 0, \text{Tr}(X) = 1, \text{rank}(X) = 1\}$: the dyad set of matrices.

2.2. Solving Optimization Problem with SDP Constraints Using Matrix Generation

We are interested in the following general convex problem:

$$\min_X f(X), \quad \text{s.t. } X \in \rho_n. \quad (2.1)$$

Here, $f(\cdot)$ is a convex function defined in \mathbb{S}_+ .

We need to extend the Fenchel conjugate [13] to matrices. For a function $f(\cdot) : \mathbb{R}^n \rightarrow (-\infty, \infty]$, the Fenchel conjugate $f^*(\cdot) : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is the lower semicontinuous and convex function

$$f^*(\mathbf{u}) = \sup_{\mathbf{x} \in \mathbb{R}^n} \{ \mathbf{u}^\top \mathbf{x} - f(\mathbf{x}) \}. \quad (2.2)$$

Now, we impose the following two conditions on $f(\cdot)$:

- (i) $f(\cdot)$ is differentiable everywhere,
- (ii) the gradient $f'(\mathbf{x})$ is continuous and monotonically increasing at each direction.

In other words, $f(\cdot)$ is differentiable and strictly convex. In this case, the Fenchel conjugate is also called Legendre transform and has the following important result:

$$\mathbf{u} = f'(\mathbf{x}). \quad (2.3)$$

By analogy, we can define a Fenchel conjugate for a matrix

$$f^*(U) = \sup_{X \in \mathbb{S}} \{ \langle U, X \rangle - f(X) \}, \quad (2.4)$$

as in the vector space \mathbb{R}^n . $f^*(\cdot)$ defined in (2.4) must be lower semicontinuous and convex, because $f^*(\cdot)$ is a supremum of linear continuous functions of U .

We are ready to reformulate our problem (2.1). It is proved in [10] that a p.s.d. matrix can always be decomposed as a linear convex combination of a set of rank-one matrices, we decompose $X \in \mathbb{S}_+$ into a convex combination of a set of dyads

$$X = \sum_{j=1}^J \theta_j Z_j, \quad (2.5)$$

with $\boldsymbol{\theta} = [\theta_1, \dots, \theta_J]^\top \in \Delta_J$, $Z_j \in \mathbb{Q}_n$, for all j . So, we can write (2.1) into

$$\begin{aligned} \min_{\boldsymbol{\theta}} f(X), \quad \text{s.t. } X &= \sum_{j=1}^J \theta_j Z_j, \\ \boldsymbol{\theta} &\succcurlyeq 0, \quad \mathbf{1}^\top \boldsymbol{\theta} = 1, \\ Z_j &\in \mathbb{Q}_n, \quad \forall j. \end{aligned} \quad (2.6)$$

Note that here, $\boldsymbol{\theta}$ and X are the optimization variables. Although X is redundant in (2.6), we need to keep it to arrive at the important Lagrange dual problem later.

This above problems is still very hard to solve, since it has nonconvex rank constraints, and the variable J is indefinite, because there are an indefinite number of rank-one matrices. However, if we somehow know matrices Z_j ($j = 1, \dots$) a priori, then all the constraints imposed on Z_j ($j = 1, \dots$) can be dropped, and the problem becomes a linear program.

We apply matrix generation to solve the problem. Matrix generation is an extension of column generation to nonpolyhedral semidefinite constraints for solving difficult large-scale optimization problems [10]. It is a method to avoid considering all variables of a problem explicitly, and only a small subset of the entire variable set is considered. Once the problem with the small subset variables is solved, the question of if there are any other variables that can be included to improve the solution should be answered. For convex programs, the primal variables correspond to the dual constraints. The dual can be used to solve the above subproblem.

The Lagrangian is

$$L(\theta, X, U, t, \mathbf{p}) = f(X) - \left\langle U, X - \sum_{j=1}^J \theta_j Z_j \right\rangle - \mathbf{p}^\top \theta - t(\mathbf{1}^\top \theta - 1), \quad (2.7)$$

with $\mathbf{p} \succeq 0$, and the dual is

$$\inf_{\theta, X} L = \inf_X (f(X) - \langle U, X \rangle) + \underbrace{(\boldsymbol{\varphi} - \mathbf{p} - t\mathbf{1})^\top \theta}_{\text{must be 0}} + t = t - f^*(U), \quad (2.8)$$

where we have defined $\boldsymbol{\varphi}$ with $\varphi_j = \langle U, Z_j \rangle$, $j = 1 \dots J$. The dual problem of (2.6) is

$$\max_{t, U} t - f^*(U), \quad \text{s.t. } \boldsymbol{\varphi} \succeq t\mathbf{1}. \quad (2.9)$$

Essentially the dual is

$$\max_{t, U} t - f^*(U), \quad \text{s.t. } \langle U, Z_j \rangle \geq t, \quad \forall j = 1 \dots J. \quad (2.10)$$

We now only consider a small subset of the variables in the primal; that is, only a subset of Z (denoted by \bar{Z}) is used. The LP solved using \bar{Z} is usually termed restricted master problem (RMP). Because the primal variables correspond to the dual constraints, solving RMP is equivalent to solve a relaxed version of the dual problem [10]. If all the constraints that we have not added to the dual problem are kept, solving the restricted problem is equivalent to solve the original problem. Otherwise, if there exists at least one constraint that is violated, the violated constraints correspond to variables in primal that are not in RMP. Adding these variables to RMP leads to a new RMP problem that needs to be reoptimized. Here, we have a iterative algorithm that either finds a new Z' such that

$$\langle U, Z' \rangle < t, \quad (2.11)$$

where t is the solution of the current restricted problem, or it guarantees that such a Z' does not exist. To make convergence fast, we find the one by solving the following optimization problem:

$$Z' = \arg \min_Z \{ \langle U, Z' \rangle \}, \quad \text{s.t. } Z \in Q_n. \quad (2.12)$$

The optimal value Z' is exactly vv^\top , where v is the eigenvector of U that corresponds to the smallest eigenvalue. For the time being, Z' is added to the dyads, and the only variable to estimate is θ . The primal can be written as

$$\min_{\theta} f(\theta), \quad \text{s.t. } \theta \in \Delta_J. \quad (2.13)$$

We generalize the EG algorithm to the matrix case and use it to solve the above problem.

At optimality, because of (2.3), we have the connection between the primal and dual variables

$$U^* = f'(X^*). \quad (2.14)$$

We also know that at least one of the dual constraints takes the equality at optimality, which enable us to obtain t^*

$$t^* = \min_{j=1 \dots J} \{\langle U^*, Z_j \rangle\}, \quad (2.15)$$

t^* can be used as the iteration stopping criteria.

2.3. The Algorithm Implementation

2.3.1. The Matrix Generation

We propose the implementation of the algorithm for the general convex problem (2.1). To be more general, we extend the SDP constraint $X \in \mathcal{D}_n = \{X \in \mathbb{S}_+^n \mid X \succcurlyeq 0, \text{Tr}(X) = 1\}$ to $X \in \{X \in \mathbb{S}_+^n \mid X \succcurlyeq 0, \text{Tr}(X) = k\}$, as shown in(2.16)

$$\min_X f(X), \quad \text{s.t. } X \succcurlyeq 0, \quad \text{Tr}(X) = k, \quad (2.16)$$

where k is a positive integer.

The detailed implementation of the algorithm is proposed in Algorithm 1. We can observe, at each iteration, that a new matrix is generated, hence the algorithm is named matrix generation. Besides, at each iteration, only the most significant eigenvalue and its corresponding eigenvector are needed. This makes the algorithm efficient.

2.3.2. The EG Algorithm

In [12], it is shown that EG style updates can converge very quickly compared to other methods. It has been successfully applied to structured prediction problems such as structured support vector machines and conditional random field. We generalize the EG algorithm to matrix to solve the special structure convex optimization problem in MG algorithm. We first give a brief introduction to the EG algorithm [12, 14, 15]. The EG algorithm efficiently solves the convex optimization problem [16]

$$\min_{\theta} f(\theta), \quad \text{s.t. } \theta = [\theta_1, \dots, \theta_J]^\top \in \Delta_J = \left\{ \theta \in \mathbb{R}^J : \mathbf{1}^\top \theta = k, \theta \geq 0 \right\}, \quad (2.17)$$

Initialize:

- (i) The maximum number of integrations J_{\max} .
- (ii) The pre-set tolerance value ε (e.g., 10^{-5}).
- (iii) $t_0^* = 0$.

Iteration 1:

- (1) randomly select $Z_1 \in Q_n$; $\theta = k$.
- (2) then compute: $X = \theta Z_1$, $U^* = f'(X)$, $t_1^* = \langle U, Z_1 \rangle$.

While iteration $J = 2, 3, \dots, J_{\max}$ **do**

- (1) If $\text{abs}(t_J^* - t_{J-1}^*) < \varepsilon$ then break (problem solved);
- (2) Find a new Z_J by finding the eigenvector v corresponding to the smallest eigenvalue of U^* (v is normalized by $v \leftarrow (v / \|v\|_2)$ and $Z_J = v * v^\top$).
- (3) Find a new $\theta = [\theta_1, \dots, \theta_J]^\top \in \Delta_J = \{\theta \in R^J : \mathbf{1}^\top \theta = k, \theta \succeq 0\}$, by solving the following spectral structure problem by EG algorithm as described in Section 2.3.3:

$$\min_{\theta} f(\theta), \quad \text{s.t. } \theta \in R^J, \quad \mathbf{1}^\top \theta = k, \quad \theta \succeq 0$$
 where $f(\theta)$ is obtained from $f(X)$ by substituting X with $\sum_{j=1}^J \theta_j Z_j$
- (4) Compute $X = \sum_{j=1}^J \theta_j Z_j$, $U^* = f'(X)$, $t_J^* = \min\{\langle U^*, Z_J \rangle\}$

Output: The p.s.d. matrix $X = \sum_{j=1}^J \theta_j Z_j$ and minimal value $f(X)$.

Algorithm 1: The matrix generation.

under the assumption that the object function $f(\cdot)$ is a convex Lipschitz continuous function with Lipschitz constant L_f w.r.t. a fixed given norm $\|\cdot\|$. The mathematical definition of L_f is that $|f(\mathbf{x}) - f(\mathbf{z})| \leq L_f \|\mathbf{x} - \mathbf{z}\|$ holds for any \mathbf{x}, \mathbf{z} in the domain of $f(\cdot)$. The detail of the EG algorithm is shown in Algorithm 2.

2.3.3. Evaluation of the Algorithm

In this section, we evaluate the convergence, running speed, and memory consumption of the algorithm by solving

$$\min_X \|K - X\|_F^2, \quad \text{s.t. } X \succeq 0, \quad X\mathbf{1}^\top = \mathbf{1}, \quad X = X^\top, \quad \text{Tr}(X) = k, \quad (2.18)$$

where $\|X\|_F = \sum_{ij} X_{ij}^2$ is the Frobenius norm, K is a $n \times n$ symmetric p.s.d. matrix, and k is a positive integer. This problem is important in the issue of the affinity matrix normalization of spectral clustering [17].

The above problem is approximate to

$$\min_X \left(\|K - X\|_F^2 + \delta \left\| X\mathbf{1}^\top - \mathbf{1} \right\|_2^2 \right), \quad \text{s.t. } X \succeq 0, \quad \text{Tr}(X) = k, \quad (2.19)$$

where δ is a scalar, for example, $\delta = 10^3$.

Before reporting the results, we compare the proposed algorithm with the convex optimization solver, namely, SDPT3 [18] which is used as internal solvers in the disciplined convex programming software CVX [19, 20]. The CVX toolbox can solve standard problems

Initialize:

(i) $\theta^0 \in$ the interior of $\Delta_J = \{\theta \in R^J : \mathbf{1}^T \theta = k, \theta \geq 0\}$.

(ii) The maximum number of integrations i_{\max}

While iteration $i = 1, 2, \dots, i_{\max}$ **do**

(1) Generate the sequence $\{\theta^i\}$ with:

$$\theta_j^i = \frac{\theta_j^{i-1} \exp[-\tau_i f'_j(\theta^{i-1})]}{\sum_{j=1}^J \theta_j^{i-1} \exp[-\tau_i f'_j(\theta^{i-1})]}$$

where τ_i is step-size, which can be determined by $\tau_i = (\sqrt{2 \log J / L_f})(1/\sqrt{i})$ following [15],
 $f'(\theta) = [f'_1(\theta), \dots, f'_J(\theta)]^T$ is the gradient of $f(\cdot)$.

(2) Stop if some stopping criteria are met.

Output: θ^i and $f(\theta)$.

Algorithm 2: The EG algorithm: solving the special structure problem shown in (2.17).

such as linear programs (LPs), quadratic programs (QPs), second-order cone programs (SOCPs), and semidefinite programs (SDPs).

(1) Convergence

We randomly generate a 50×50 p.s.d. matrix K in (2.18); let k be 5. In Figure 1, we plot the optimal value obtained by the proposed algorithm at each iteration (the blue curve). The red dash line shows the ground truth obtained by directly solving the original problem in (2.16) using the CVX toolbox. We can see that the algorithm converges to the near-optimal solutions quickly.

(2) Running Time

We solve the problem in (2.16) using the proposed algorithm and the CVX toolbox, respectively, and the running time is shown in Figure 2. The blue curve is the running time of the proposed algorithm versus the matrix dimension, and the red curve is the running time of CVX versus the matrix dimension. We can see that the running time of CVX increases quickly with the matrix dimension and the running time of the proposed algorithm is a fraction of the CVX and scales very well with the matrix dimension. The proposed algorithm has a clear advantage on computational efficiency. One reason is that at each iteration, only the most significant eigenvalue and its corresponding eigenvector are needed in the proposed algorithm.

(3) Memory Consumption

Figure 3 shows the memory consumption of the proposed algorithm and the CVX toolbox. The blue curve is the approximate memory consumption of the proposed algorithm versus the matrix dimension, and the red curve is the approximate memory consumption of CVX versus the matrix dimension. It can be seen that the memory consumption of CVX increases quickly with the matrix dimension n and the memory consumption of the proposed algorithm is a fraction of the CVX and scales very well with the dimension.

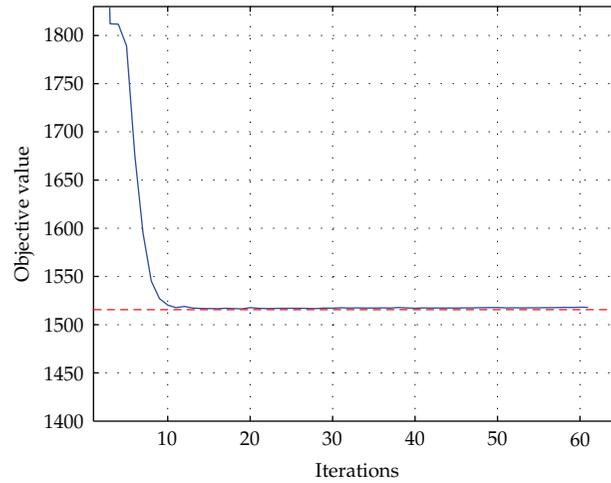


Figure 1: The convergence of the proposed algorithm. The blue curve shows the optimal value obtained by the proposed algorithm at each iteration, and the dash line shows the ground truth obtained by directly solving the original problem in (2.16) using the CVX toolbox.

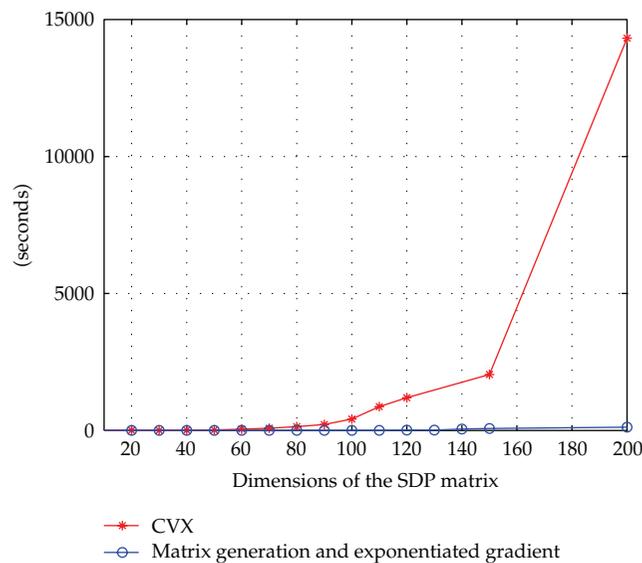


Figure 2: Running time of the proposed algorithm. The blue curve shows the running time of the proposed algorithm versus the matrix dimension, and the red curve shows the running time of CVX versus the matrix dimension.

In summary, the proposed algorithm can converge to the near-optimal solutions quickly and scale very well with the dimension, which are very important for solving large-scale SDP problems. As a result, the proposed method is potential to be applied to computer vision to deal with large matrix of 3D visual data [21] and clustering in data analysis. In next section, we will show how to apply the proposed method to clustering problems.

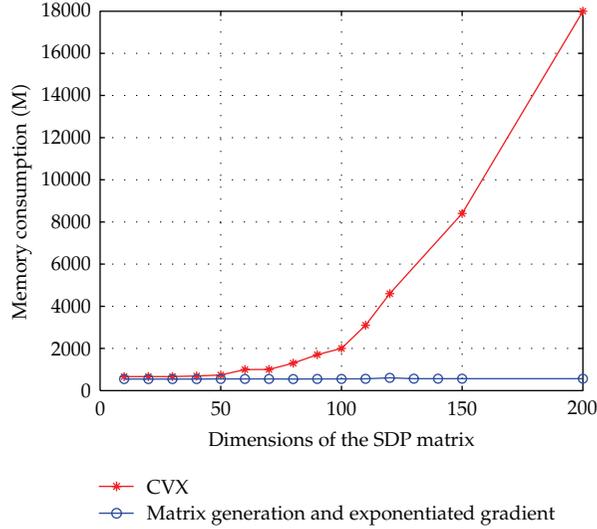


Figure 3: Memory consumption of the proposed algorithm. The blue curve shows the approximate memory consumption of the proposed algorithm versus the matrix dimension, and the red curve shows the approximate memory consumption of CVX versus the matrix dimension.

Table 1: UCI datasets used together with some characteristics and the clustering results achieved using the different methods.

Dataset	k	Size	Dims.	Accuracy			
				K-means	Spectral clustering	SDP relaxation Ours	CVX
Soybean	4	47	36	0.723	0.745	0.766	0.766
Iris	3	150	5	0.893	0.933	0.947	0.953
Wine	3	178	13	0.702	0.725	0.730	0.730
SPECTF heart	2	267	44	0.625	0.802	0.816	0.816

3. The Application

The problem of partitioning data points into a number of distinct sets, known as the clustering problem, is one of the most fundamental and significant topics in data analysis and machine learning. From an optimization viewpoint, clustering aims at the minimal sum-of-squares (MSSC) based on some definition of similarity or distance. But the variable is a 0-1 discrete indicator matrix $Y \in \mathbb{R}^{m \times n}$ (n is the number of samples, and m is the number of clusters), thus the optimization becomes a mixed integer programming with nonlinear objective, which is NP-hard. The computation time of NP-hard problems is too heavy to use in practical engineering applications and should be accelerated by some approximation methods. K-means and spectral clustering are the two most well-known algorithms to approximately solve the clustering problem. However, the former is merely able to achieve local optimum through an indirect way, and the later, which can obtain a global optimum through eigenvalue decomposition, is unfortunately a weak relaxation of the N-cut problem [7]. Recent research of semidefinite programming (SDP) relaxation on clustering proposed that the minimal sum-of-squares (MSSC) [22] is possible to be relaxed as SDP problems, which is more direct than K-means and tighter than spectral clustering [7].

Table 2: Time consumptions (seconds) of the proposed algorithm and CVX on clustering the selected UCI datasets.

	Soybean	Iris	Wine	SPECTF heart
Ours	1.34	2.879	4.22	21.29
CVX	1.21	4.629	6.95	28.00

In this section, we apply the proposed algorithm to solve the SDP relaxed clustering problem (0-1 SDP for clustering) [22]

$$\min_X \text{Tr}(W(I - X)), \quad \text{s.t. } X\mathbf{1} = \mathbf{1}, \text{Tr}(X) = m, X \geq 0, X \geq 0, \quad (3.1)$$

where W is the so-called affinity matrix whose entries represent the similarities or closeness among the entities in the dataset, and m is the class number of the dataset. According to [22], it is reasonable to relax the nonnegative constraint in (3.1)

$$\min_X \text{Tr}(W(I - X)), \quad \text{s.t. } X\mathbf{1} = \mathbf{1}, \text{Tr}(X) = m, X \geq 0. \quad (3.2)$$

We solve the problem in (3.2) using the algorithm described in Algorithm 1. To test the algorithm, we carry out some experiments on four real datasets collected from UCI machine learning repository (<http://www.ics.uci.edu/~mllearn/MLRepository.html>.) The datasets are listed in Table 1 together with some their characteristics. k , Size, and Dim are the class number, the number of instances, and the number of attributes of each dataset, respectively. All the experiments are done by using Matlab on a personal computer with a 2.93GHz processor and a 2G memory. The accuracy of clustering each dataset of our algorithm is compared with K-means, spectral clustering, and CVX (here CVX means the SDP problem in (3.2) is solved using the CVX toolbox). The results are shown in Table 1. We can see that

- (i) the proposed algorithm performs better than K-means,
- (ii) the proposed algorithm achieves higher accuracy than the spectral clustering that is because the proposed algorithm is the SDP relaxation of the clustering problem and is tighter than the spectral relaxation,
- (iii) the proposed algorithm is as good as CVX. This means it can solve the SDP problem as accurately as the CVX toolbox.

We also compare the speed of the proposed algorithm with CVX, and the results are shown in Table 2. It shows that the proposed algorithm is much faster than CVX.

As observed from the experiments, it can be found that the proposed algorithm performs very well on these datasets and can efficiently solve the 0-1 SDP problem for clustering. Both speed and accuracy are improved as compared with traditional interior-point SDP solvers. We plan to further extend the approach to more general problems and look for other applications such as modeling [23, 24] and stabilizing some kinds of dynamical systems [25] model reduction.

4. Conclusion

This paper proposed an approach to solve the optimization problem with SDP constraints using matrix generation and exponentiated gradient. The process is faster and more scalable than the traditional SDP approach. The results of numerical experiments show that the method scales very well with the problem dimensions. We also applied the approach to solve the SDP relaxation clustering problem. Experimental results on real datasets show that the algorithm outperforms much in both speed and accuracy as compared with interior-point SDP solvers.

References

- [1] A. Sootla, *Model reduction using semidefinite programming*, Licentiate Thesis, Department of Automatic Control, Lund University, Lund, Sweden, 2009.
- [2] Z. Mahmood, B. Bond, T. Moselhy, A. Megretski, and L. Daniel, "Passive reduced order modeling of multiport interconnects via semidefinite programming," in *Proceedings of the International Conference on Design, Automation and Test in Europe (DATE '10)*, pp. 622–625, 2010.
- [3] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, part 1, pp. 480–487, 2011.
- [4] P. S. Ansell, K. D. Glazebrook, I. Mitrani, and J. Niño-Mora, "Semidefinite programming approach to the optimal control of a single server queueing system with imposed second moment constraints," *Journal of the Operational Research Society*, vol. 50, no. 7, pp. 765–773, 1999.
- [5] E. P. Xing and M. I. Jordan, "On semidefinite relaxations for normalized k-cut and connections to spectral clustering," Tech. Rep., University of California at Berkeley, Berkeley, Calif, USA, 2003.
- [6] B. Kulis, A. C. Surendran, and J. C. Platt, "Fast low-rank semidefinite programming for embedding and clustering," in *Proceedings of the International Workshop on Artificial Intelligence and Statistics*, San Juan, Puerto Rico, 2007.
- [7] T. Zhou and D. Tao, "Fast gradient clustering," in *Proceedings of the Advances in Neural Information Processing Systems, Workshop on Discrete Optimization in Machine Learning*, pp. 1–6, 2009.
- [8] L. Cayton and S. Dasgupta, "Robust euclidean embedding," in *Proceedings of the International Conference on Machine Learning*, 2006.
- [9] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *Journal of Machine Learning Research (JMLR)*, vol. 5, pp. 27–72, 2003/04.
- [10] C. Shen, A. Welsh, and L. Wang, "PSDBoost: Matrix-generation linear programming for positive semidefinite matrices learning," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 1473–1480, Vancouver, Canada, 2008.
- [11] M. E. Lübbecke and J. Desrosiers, "Selected topics in column generation," *Operations Research*, vol. 53, no. 6, pp. 1007–1023, 2005.
- [12] A. Globerson, T. Y. Koo, X. Carreras, and M. Collins, "Exponentiated gradient algorithms for log-linear structured prediction," in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, vol. 227, pp. 305–312, 2007.
- [13] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
- [14] J. Kivinen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," *Information and Computation*, vol. 132, no. 1, pp. 1–64, 1997.
- [15] A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Operations Research Letters*, vol. 31, no. 3, pp. 167–175, 2003.
- [16] C. Shen, P. Wang, and H. Li, "LACBoost and FisherBoost: optimally building cascade classifiers," in *Proceedings of the 11th European Conference on Computer Vision (ECCV '10)*, vol. 2 of *Lecture Notes in Computer Science (LNCS) 6312*, pp. 608–621, Springer, Crete Island, Greece, September 2010.
- [17] R. Zass and A. Shashua, "Doubly stochastic normalization for spectral clustering," in *Proceedings of the Advances in Neural Information Processing Systems*, 2006.

- [18] R. H. Tütüncü, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming*, vol. 95, no. 2, pp. 189–217, 2003.
- [19] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," 2010, <http://cvxr.com/cvx>.
- [20] M. C. Grant and S. P. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds., vol. 371 of *Lecture Notes in Control and Information Sciences*, pp. 95–110, Springer, London, UK, 2008.
- [21] S. Y. Chen, Y. F. Li, and J. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
- [22] J. Peng and Y. Wei, "Approximating k -means-type clustering via semidefinite programming," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 186–205, 2007.
- [23] S. Y. Chen and Y. F. Li, "Vision sensor planning for 3-D model acquisition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 5, pp. 894–904, 2005.
- [24] B. N. Bond, Z. Mahmood, Y. Li et al., "Compact modeling of nonlinear analog circuits using system identification via semidefinite programming and incremental stability certification," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 29, no. 8, pp. 1149–1162, 2010.
- [25] 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.

Review Article

Visiting Power Laws in Cyber-Physical Networking Systems

Ming Li¹ and Wei Zhao²

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, Avenue Padre Tomas Pereira, Taipa 1356, Macau SAR, China

Correspondence should be addressed to Ming Li, ming_lihk@yahoo.com

Received 23 February 2011; Accepted 23 March 2011

Academic Editor: Carlo Cattani

Copyright © 2012 M. Li and W. Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cyber-physical networking systems (CPNSs) are made up of various physical systems that are heterogeneous in nature. Therefore, exploring universalities in CPNSs for either data or systems is desired in its fundamental theory. This paper is in the aspect of data, aiming at addressing that power laws may yet be a universality of data in CPNSs. The contributions of this paper are in triple folds. First, we provide a short tutorial about power laws. Then, we address the power laws related to some physical systems. Finally, we discuss that power-law-type data may be governed by stochastically differential equations of fractional order. As a side product, we present the point of view that the upper bound of data flow at large-time scaling and the small one also follows power laws.

1. Introduction

Cyber-physical networking systems (CPNSs) consist of computational and physical elements integrated towards specific tasks [1–3]. Generally, both data and systems in CPNSs are heterogeneous. For instance, teletraffic data are different from transportation traffic, letting along other data in CPNS, such as those in physiology. Therefore, one of the fundamental questions is what possible general laws are to meet CPNS in theory. The answer to that question should be in two folds. One is data. The other is systems that transmit data from sources to destinations within a predetermined restrict period of time according to a given quality of service (QoS).

In general, both data and systems in CPNS are multidimensional. For instance, data from sources to be transmitted may be from a set of sensors distributed in a certain area.

Destinations receiving data may be a set of actuators, for example, a set of cars distributed in a certain area. Systems to transmit data are generally distributed.

Denote by \mathbb{R}^n the n -dimensional Euclidean space. Denote data at sources and destinations, respectively, by $X(t)$ which is supposed to be n -dimensional and $Y(t)$ which is supposed to be m -dimensional. They are given by

$$X(t) = (x_1(t), \dots, x_n(t)), \quad (1.1)$$

$$Y(t) = (y_1(t), \dots, y_m(t)). \quad (1.2)$$

A stochastic equation describing an abstract relationship between $X(t)$ and $Y(t)$ may be expressed by

$$Y^T(t) = S(t) \otimes X^T(t) \oplus B^T(t), \quad (1.3)$$

where T implies the transposition, $S(t)$ is a servie matrix of $n \times m$ order of a system, and $B(t)$, which is a vector with the same dimension as that of $Y(t)$, may represent uncertainty for the operation of $S(t) \otimes X^T(t)$. The operations \otimes and \oplus are to be studied from a view of systems, and they are out of the scope of this paper.

Note that $X(t)$ is usually a random field, see for example the work of Chilés and Delfiner in [4] in geosciences, the work of Uhlig in [5] in telecommunications, the work of Messina et al. in [6] in power systems, the work of Muniandy and Stanslas in [7] in medical images, the work of Mason et al. in [8] in wind engineering, and of simply citing a few. The statistics of $X(t)$ is obviously crucial for the performance analysis of physical systems in CPNS. It is noted that the physical meaning of $X(t)$ is diverse. For example, it may represent a two-dimensional aeromagnetic data (Spector and Grant [9]), a medical image (Fortin et al. [10]), vegetation data (Myrhaug et al. [11]), surface crack in material science (Tanaka et al. [12]), and data in physiology (Werner [13], West [14]), DNA (Cattani [15]), data in stock markets (Rosenow et al. [16]), just mentioning a few. Therefore, seeking for possible universalities of $X(t)$ in CPNS is desired.

Without lose of generality, we rewrite (1.1) by

$$X(\mathbf{t}) = X(t_1, \dots, t_n), \quad (1.4)$$

where $\mathbf{t} = (t_1, \dots, t_n)$. The norm of \mathbf{t} is given by

$$\|\mathbf{t}\| = \sqrt{t_1^2 + \dots + t_n^2}. \quad (1.5)$$

The autocovariance function (ACF) of $X(\mathbf{t})$ is given, over the hyperrectangle $C = \prod_{i=1}^n [a_i, b_i]$ for $a_i, b_i \in \mathbb{R}$ (Adler [17]), by

$$C(\boldsymbol{\tau}) = E[X(\mathbf{t})X(\mathbf{t} + \boldsymbol{\tau})], \quad (1.6)$$

where E is the mean operator, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)$, and

$$\|\boldsymbol{\tau}\| = \sqrt{\tau_1^2 + \dots + \tau_n^2}. \quad (1.7)$$

The ACF $C(\boldsymbol{\tau})$ measures how $X(\mathbf{t})$ correlates to $X(\mathbf{t} + \boldsymbol{\tau})$.

From the point of view of applications of CPNS, we are interested in two asymptotic expressions of $C(\boldsymbol{\tau})$. One is $C(\boldsymbol{\tau})$ for $\|\boldsymbol{\tau}\| \rightarrow 0$. The other is $C(\boldsymbol{\tau})$ for $\|\boldsymbol{\tau}\| \rightarrow \infty$. The former characterizes the small scaling phenomenon of $X(\mathbf{t})$. The latter measures the large scaling one. It is quite natural for us to investigate two types of scaling phenomena. As a matter of fact, one may be interested in small scaling in some applications, for example, admission control in computer communication or monitoring sudden disaster in geoscience. On the other side, one may be interested in large scaling in applications, for example, long-term performance analysis of systems. Exact expression of $C(\boldsymbol{\tau})$ is certainly useful, but it may usually be application dependent. Consequently, we study possible generalities of $C(\boldsymbol{\tau})$ for $\|\boldsymbol{\tau}\| \rightarrow 0$ and $\|\boldsymbol{\tau}\| \rightarrow \infty$ instead of its exactly full expressions. The aim of this paper is to explain that both the small scaling described by $C(\boldsymbol{\tau})$ for $\|\boldsymbol{\tau}\| \rightarrow 0$ and the large scaling described by $C(\boldsymbol{\tau})$ for $\|\boldsymbol{\tau}\| \rightarrow \infty$, in some fields related to CPNS, ranging from geoscience to computer communications, follow power laws.

The rest of paper is organized as follows. Short tutorial about power laws is explained in Section 2. Some cases of power laws relating to computational and physical systems in CPNS are described in Section 3. Stochastically differential equations to govern power-law-type data are discussed in Section 4, which is followed by our conclusions.

2. Brief on Power Laws

Denote by (Ω, T, P) the probability space. Then, $x(t, \zeta)$ is said to be a stochastic process when the random variable x represents the value of the outcome of an experiment T for every time t , where Ω represents the sample space, T is the event space or sigma algebra, and P is the probability measure.

As usual, $x(t, \zeta)$ is simplified to be written as $x(t)$. That is, the event space is usually omitted. Denote by $P(x)$ the probability function of x . Then, one can define the general n th order, time varying, joint distribution function $P(x_1, \dots, x_n; t_1, \dots, t_n)$ for the random variables $x(t_1), \dots, x(t_n)$. The joint probability density function (pdf) is written by

$$p(x_1, \dots, x_n; t_1, \dots, t_n) = \frac{\partial^n P(x_1, \dots, x_n; t_1, \dots, t_n)}{\partial x_1, \dots, \partial x_n}. \quad (2.1)$$

For simplicity, we write $P(X) = P(x_1, \dots, x_n; t_1, \dots, t_n)$ and $p(X) = p(x_1, \dots, x_n; t_1, \dots, t_n)$. Then, the probability is given by

$$P(X_2) - P(X_1) = \text{Prob}[X_1 < \xi < X_2] = \int_{X_1}^{X_2} p(\xi) d\xi. \quad (2.2)$$

The mean and the ACF of X based on pdf are written by (2.3) and (2.4), respectively,

$$\mu_X = \int_{-\infty}^{\infty} Xp(X)dX, \quad (2.3)$$

$$C_{XX}(\tau) = \int_{-\infty}^{\infty} X(t)X(t + \tau)p(X)dX. \quad (2.4)$$

Let V_X be the variance of X . Then,

$$V_X = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (X - \mu_X)^2 p(X)dX. \quad (2.5)$$

The above expressions imply that the integrals in (2.3) and (2.5) are convergent in the domain of ordinary functions if $p(X)$ is light tailed, for example, exponentially decayed (Li et al. [18]). Light-tailed pdfs are not our interests. We are interested in heavy-tailed pdfs. By heavy tail we mean that $p(X)$ decays so slowly that (2.3) and (2.5) may be divergent. In the following subsections, we will describe power laws in probability space, ACF, and power spectrum density (PSD) function, respectively.

2.1. Power Law in pdf

A typical heavy-tailed case is the Pareto distribution. Denote by $p_{\text{Pareto}}(X)$ the pdf of the Pareto distribution. Then,

$$p_{\text{Pareto}}(X) = \frac{ab}{X^{a+1}}, \quad (2.6)$$

where a and b are parameters and $X \geq a$. The mean and variance of X that follows $p_{\text{Pareto}}(X)$ are given by (2.7) and (2.8), respectively,

$$\mu_{\text{Pareto}} = \frac{ab}{a-1}, \quad (2.7)$$

$$V_{\text{Pareto}} = \frac{ab^2}{(a-1)^2(a-2)}. \quad (2.8)$$

It is easily seen that μ_{Pareto} and V_{Pareto} do not exist if $a = 1$. Note that μ_X implies a global property of X while V_X represents a local property of X . Therefore, heavy-tailed pdfs imply that X is in wild randomness due to infinite or very large variance, see the work of Mandelbrot in [19] for the meaning of wild randomness.

Note 1. The Pareto distribution is an instance of power-law-type pdf.

2.2. Power Law in ACF

A consequence of a heavy-tailed random variable in ACF is that $C_{XX}(\tau)$ is slowly decayed. By slowly decayed we mean that $C_{XX}(\tau)$ decays hyperbolically in the power law given by (Adler et al. [20])

$$C_{XX}(\tau) \sim \tau^{-d}, \quad d \in \mathbb{R}. \quad (2.9)$$

The Taqqu theorem describes the relationship between a heavy-tailed pdf and hyperbolically decayed ACF (Abry et al. [21]).

2.3. Power Law in PSD

Denote by $S_{XX}(\omega)$ the PSD of X . Then,

$$S_{XX}(\omega) = \int_{-\infty}^{\infty} C_{XX}(\tau) e^{-j\omega\tau} d\tau, \quad j = \sqrt{-1}. \quad (2.10)$$

According to the theory of generalized functions (Kanwal [22]), one has

$$S_{XX}(\omega) \sim |\omega|^{d-1}. \quad (2.11)$$

Therefore, power law in PSD, which is usually termed $1/f$ noise, see the work of Wornell in [23], the work of Keshner in [24], the work of Ninness in [25], the work of Corsini and Saletti in [26], and the work of Li in [27].

2.4. Power Laws in Describing Scaling Phenomena

We now turn to scaling descriptions. Small scaling phenomenon may be investigated by $C_{XX}(\tau)$ for $\tau \rightarrow 0$ and large scaling for $\tau \rightarrow \infty$, respectively (Li and Zhao [28]).

On the one side, following Davies and Hall [29], if $C_{XX}(\tau)$ is sufficiently smooth on $(0, \infty)$ and if

$$C_{XX}(0) - C_{XX}(\tau) \sim c_1 |\tau|^\alpha \quad \text{for } |\tau| \rightarrow 0, \quad (2.12)$$

where c_1 is a constant and α is the fractal index of X , then the fractal dimension, denoted by D , of X is expressed by

$$D = 2 - \frac{\alpha}{2}. \quad (2.13)$$

Note 2. Fractal dimension is a parameter to characterize small scaling phenomenon (Mandelbrot [30], Gneiting and Schlather [31], Li [32]).

On the other side, if

$$C_{XX}(\tau) \sim |\tau|^{-\beta} \quad (\tau \rightarrow \infty), \quad (2.14)$$

then the parameter β is used to measure the statistical dependence of X . If $\beta > 1$, $C_{XX}(\tau)$ is integrable, and accordingly X is short-range dependent (SRD). If $0 < \beta < 1$, $C_{XX}(\tau)$ is nonintegrable and X is long-range dependent (LRD), see the work of Beran in [33]. Representing β by the Hurst parameter $H \in (0, 1)$ yields

$$H = 1 - \frac{\beta}{2}. \quad (2.15)$$

Note 3. Statistical dependence, either SRD or LRD, is a property for large scaling phenomenon.

3. Cases of Power Laws in CPNS

We address some application cases of power laws in CPNS in this section.

3.1. Power Laws in the Internet

Let $x(t)$ be the teletraffic time series. It may represent the packet size of teletraffic at time t . Denote the ACF of $x(t)$ by

$$R(\tau) = E[X(t + \tau)X(t)]. \quad (3.1)$$

Then, we have (Li and Lim [34])

$$R(\tau) \sim |\tau|^\alpha, \quad \tau \rightarrow 0, \quad (3.2)$$

$$R(\tau) \sim |\tau|^{-\beta}, \quad \tau \rightarrow \infty.$$

From (3.2), the fractal dimension D and the Hurst parameter H of teletraffic are, respectively, given by

$$D = 2 - \frac{\alpha}{2}, \quad (3.3)$$

$$H = 1 - \frac{\beta}{2}.$$

The above exhibits that both the small scaling and the large one follow power laws.

It is worth noting that the upper bounds of teletraffic also follow power laws. In fact, the amount of teletraffic accumulated in the interval $[0, t]$ is upper bounded by

$$\int_0^t x(u) du \leq \sigma + \rho t, \quad (3.4)$$

where σ and ρ are constants and $t > 0$ (Cruz [35]). Following Li and Zhao [28], we have the bounds of both the small-time scaling and the large one, respectively expressed by

$$\begin{aligned} \int_0^t x(u) du &\leq r^{2D-5} \sigma \text{ for small } t, \\ \int_0^t x(u) du &\leq a^{-H} \rho \text{ for large } t, \end{aligned} \quad (3.5)$$

where $r > 0$ is a small-scale factor and $a > 0$ is a large-scale factor. Therefore, we have the following theorem.

Theorem 3.1. *Both the small-scale factor and the large one of teletraffic obey power law, that is, r^{2D-5} and a^{-H} .*

Proof. Two scaling factors follow r^{2D-5} and a^{-H} , respectively. Thus, they obey power laws. This completes the proof. \square

In addition to teletraffic, others with respect to the Internet also follow power laws. Some are listed below.

Note 4. Barabasi and Albert [36] studied several large databases in the World Wide Web (WWW), where they defined vertices by HyperText Markup Language (HTML) documents. They inferred that the probability $P(k)$ that a vertex in the network interacts with k other vertices decays hyperbolically as $P(k) \sim k^{-\gamma}$ for $\gamma > 0$, hence, power law.

Note 5. Let $P_{\text{out}}(k)$ and $P_{\text{in}}(k)$ be the probabilities of a document to have k outgoing and incoming links, respectively. Then, $P_{\text{out}}(k)$ and $P_{\text{in}}(k)$ obey power laws (Albert [37]).

Note 6. The probability of web pages among sites is of power law (Huberman and Adamic [38]).

3.2. Power Laws in Geosciences

Let $(x, y, z) \in \mathbb{R}^3$ be a spatial point. The physical meaning of a random function $U(x, y, z)$ may be diverse in the field. For instance, it may represent prospected gold amount at (x, y, z) in a gold mine, or a value of pollution index for pollution alert at (x, y, z) in a city.

For simplicity, denote a vector by $l = (x, y, z)$. Let

$$\rho = |l| = \sqrt{x^2 + y^2 + z^2}. \quad (3.6)$$

Then, one may be interested in the covariance function of $U(\rho)$. Denote by $C(\tau)$ the covariance function of $U(\rho)$. Then,

$$C(\tau) = E\{[U(\rho) - EU(\rho)][U(\rho + \tau) - EU(\rho)]\}. \quad (3.7)$$

One of the commonly used models of covariance functions in geosciences is given by

$$C(\tau) = \frac{1}{1 + |\tau|^2}. \quad (3.8)$$

The above constant power is the case of the standard Cauchy process (Webster and Oliver [39]). It fits with some cases in geosciences, see, for example, the work of Wackernagel in [40]. We list some in the following notes.

Note 7. Let $C(s)$ be the covariance function between yield densities at any two points in a region, where s represents the distance difference between two points. Then,

$$C(s) \sim |s|^{-\nu} (\nu > 0) \text{ for large } \nu, \quad (3.9)$$

see the work of Whittle in [41].

Note 8. Sea-level fluctuations, river flow, and flood height follow power laws (Li et al. [42], Lefebvre [43], Lawrance and Kottegoda [44]).

Note 9. Urban growth obeys power laws (Makse et al. [45]).

3.3. Power Laws in Wind Engineering

Wind engineering is an important field relating to wind power generation and disaster preventions from a view of CPNS. In this field, studying fluctuations of wind speed is essential.

The PSD introduced by von Kármán [46], known as the von Kármán spectra (VKS), is widely used in the diverse fields, ranging from turbulence to acoustic wave propagation in random media, see for example, the work of Goedecke et al. in [47] and the work of Hui et al. in [48]. For the VKS expressed in (3.10), we use the term VKSW for short,

$$S_{\text{von}}(f) = \frac{4u_f^2 b_v w}{f(1 + 70.8w^2)^{5/6}}, \quad w = \frac{fL_u^x}{U}, \quad (3.10)$$

where f is frequency (Hz), L_u^x is turbulence integral scale, U is mean speed, u_f is friction velocity (ms^{-1}), and b_v is friction velocity coefficient such that the variance of wind speed $\sigma_u^2 = b_v u_f^2$. Equation (3.10) implies that VKSW obeys power law for $f \in (0, \infty)$.

Another famous PSD in wind engineering is the one introduced by Davenport [49], which is expressed by

$$\frac{f S_{\text{Dav}}(f)}{u_f^2} = 4 \frac{u^2}{(1 + u^2)^{4/3}}, \quad u = \frac{1200n}{z}, \quad (3.11)$$

where n is the normalized frequency (fz/U (10 m)), U (10 m) is the mean wind speed (ms^{-1}) measured at height 10 m, $U(z)$ is the mean wind speed (ms^{-1}) measured at height z . Davenport's PSD exhibits a power law of wind speed. Other forms of the PSDs of wind speed, such as those discussed by Kaimal [50], Antoniou et al. [51], and Hiriart et al. [52], all follow power laws, referring [50–52] for details.

4. Possible Equations for Power-Law-Type Data

The cases of power laws mentioned in the previous section are a few that people may be interested in from a view of CPNS. There are others that are essential in the field of CPNS, such as power laws in earthquake, see for example, the work of Pisarenko and Rodkin in [53]. Now, we turn to the discussions about the generality about the equations that may govern data of power law type.

Conventionally, a stationary random function $y(t)$ may be taken as a solution of a differential equation of integer order, which is driven by white noise $w(t)$. This equation may be written by

$$\sum_{i=0}^p a_i \frac{d^{p-i} y(t)}{dt^{p-i}} = w(t), \quad (4.1)$$

where p and i are integers.

Let $v > 0$ and $f(t)$ be piecewise continuous on $(0, \infty)$ and integrable on any finite subinterval of $[0, \infty)$. For $t > 0$, denote by ${}_0D_t^{-v}$ the Riemann-Liouville integral operator of order v [54–57]. Then,

$${}_0D_t^{-v} f(t) = \frac{1}{\Gamma(v)} \int_0^t (t-u)^{v-1} f(u) du, \quad (4.2)$$

where Γ is the Gamma function. For simplicity, we write ${}_0D_t^{-v}$ by D^{-v} below.

Let v_p, v_{p-1}, \dots, v_0 be a strictly decreasing sequence of nonnegative numbers. Then, for the constants a_i , we have

$$\sum_{i=0}^p a_{p-i} D^{v_i} y(t) = w(t). \quad (4.3)$$

The above is a stochastically fractional differential equation with constant coefficients of order v_p . This class of equations yield random functions with power laws (Li [27]). In the case of

random fields, (4.3) is extended to be a partial differential equation of fractional order given by

$$\sum_{i=0}^p a_{p-i} D^{v_i} \mathbf{y} = \mathbf{w}, \quad (4.4)$$

where both \mathbf{w} and \mathbf{y} are multidimensional and D is an operator of partial differentiation.

Another class of stochastically differential equations of fractional order is given by (Lim and Muniandy [58])

$$\left(\sum_{i=0}^p a_i \frac{d^{p-i} \mathbf{y}(t)}{dt^{p-i}} \right)^{\beta_1} = \mathbf{w}(t) \quad (\beta_1 > 0). \quad (4.5)$$

Note that (4.3), (4.4), or (4.5) should not be taken as a simple extension of conventional equation (4.1) from integer order to fractional one. As a matter of fact, there are challenging issues with respect to differential equations of fractional order. Since data of power-law-type may be with infinite variance (Samorodnitsky and Taqqu [59]), variance analysis which is a powerful tool in the analysis of conventional random functions fails to describe random data with infinite variance. Power-law type data may be LRD, which makes the stationarity test of data a tough issue, see for example, the work of Mandelbrot in [60], the work of Abry and Veitch in [61], the work of Li et al. in [62]. Owing to power laws, stability of systems that produce such a type of data becomes a critical issue in theory, see the work of Li et al. in [63] and the references therein. In addition, the prediction of data with power laws considerably differs from that of conventional data (M. Li and J. Y. Li [64], Hall and Yao [65]). Topics in power laws are paid attention to, see for example, the work of Kamoun in [66], the work of Ng et al. in [67], the work of Song et al. in [68], the work of Cattani et al. in [69–71], and in [72–79].

5. Conclusions

We have discussed the elements of power laws from both a mathematical point of view and with respect to applications to a number of fields in CPNS. The purpose of this paper is to exhibit that power laws may yet serve as a universality of data in CPNS. We believe that this point of view may be useful for data modeling and analysis in CPNS.

Acknowledgments

This work was supported partly by the National Natural Science Foundation of China under the project Grant nos 60873264, 61070214, and the 973 plan under the Project no. 2011CB302801/2011CB302802.

References

- [1] E. A. Lee, "Cyber physical systems: design challenges," Tech. Rep. UCB/EECS-2008-8, University of California, Berkeley, Calif, USA, 2008.

- [2] J. A. Stankovic, I. Lee, A. Mok, and R. Rajkumar, "Opportunities and obligations for physical computing systems," *Computer*, vol. 38, no. 11, pp. 23–31, 2005.
- [3] R. Alur, D. Thao, J. Esposito et al., "Hierarchical modeling and analysis of embedded system," *Proceedings of the IEEE*, vol. 91, no. 1, pp. 11–28, 2003.
- [4] 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.
- [5] S. Uhlig, "On the complexity of Internet traffic dynamics on its topology," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 167–180, 2010.
- [6] A. R. Messina, P. Esquivel, and F. Lezama, "Time-dependent statistical analysis of wide-area time-synchronized data," *Mathematical Problems in Engineering*, vol. 2010, Article ID 751659, 17 pages, 2010.
- [7] S. V. Muniandy and J. Stanslas, "Modelling of chromatin morphologies in breast cancer cells undergoing apoptosis using generalized Cauchy field," *Computerized Medical Imaging and Graphics*, vol. 32, no. 7, pp. 631–637, 2008.
- [8] M. S. Mason, D. F. Fletcher, and G. S. Wood, "Numerical simulation of idealised three-dimensional downburst wind fields," *Engineering Structures*, vol. 32, no. 11, pp. 3558–3570, 2010.
- [9] A. Spector and F. S. Grant, "Statistical methods for interpreting aeromagnetic data," *Geophysics*, vol. 35, no. 2, pp. 293–302, 1970.
- [10] C. Fortin, R. Kumaresan, W. Ohley, and S. Hofer, "Fractal dimension in the analysis of medical images," *IEEE Engineering in Medicine and Biology Magazine*, vol. 11, no. 2, pp. 65–71, 1992.
- [11] D. Myrhaug, L. E. Holmedal, and M. C. Ong, "Nonlinear random wave-induced drag force on a vegetation field," *Coastal Engineering*, vol. 56, no. 3, pp. 371–376, 2009.
- [12] M. Tanaka, R. Kato, Y. Kimura, and A. Kayama, "Automated image processing and analysis of fracture surface patterns formed during creep crack growth in austenitic heat-resisting steels with different microstructures," *ISIJ International*, vol. 42, no. 12, pp. 1412–1418, 2002.
- [13] G. Werner, "Fractals in the nervous system: conceptual implications for theoretical neuroscience," *Frontiers in Fractal Physiology*, vol. 1, article 15, 28 pages, 2010.
- [14] B. J. West, "Fractal physiology and the fractional calculus: a perspective," *Frontiers in Fractal Physiology*, vol. 1, article 12, 2010.
- [15] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.
- [16] B. Rosenow, P. Gopikrishnan, V. Plerou, and H. E. Stanley, "Dynamics of cross-correlations in the stock market," *Physica A*, vol. 324, no. 1-2, pp. 241–246, 2003.
- [17] R. J. Adler, *The Geometry of Random Fields*, Wiley Series in Probability and Mathematical Statistic, John Wiley & Sons, Chichester, UK, 1981.
- [18] M. Li, W. Zhao, and S.-Y. Chen, "mBm-based scalings of traffic propagated in Internet," *Mathematical Problems in Engineering*, vol. 2011, Article ID 389803, 21 pages, 2011.
- [19] B. B. Mandelbrot, *Multifractals and 1/f Noise*, Springer, New York, NY, USA, 1998.
- [20] 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.
- [21] 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.
- [22] R. P. Kanwal, *Generalized Functions: Theory and Applications*, Birkhäuser, Boston, Mass, USA, 3rd edition, 2004.
- [23] 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.
- [24] M. S. Keshner, "1/f noise," *Proceedings of the IEEE*, vol. 70, no. 3, pp. 212–218, 1982.
- [25] B. Ninness, "Estimation of 1/f noise," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 32–46, 1998.
- [26] G. Corsini and R. Saletti, "1/f γ power spectrum noise sequence generator," *IEEE Transactions on Instrumentation and Measurement*, vol. 37, no. 4, pp. 615–619, 1988.
- [27] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [28] 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.
- [29] S. Davies and P. Hall, "Fractal analysis of surface roughness by using spatial data," *Journal of the Royal Statistical Society. Series B*, vol. 61, no. 1, pp. 3–37, 1999.

- [30] B. B. Mandelbrot, *The Fractal Geometry of Nature*, Schriftenreihe für den Referenten, W. H. Freeman and Co., San Francisco, Calif, USA, 1982.
- [31] T. Gneiting and M. Schlather, "Stochastic models that separate fractal dimension and the Hurst effect," *SIAM Review*, vol. 46, no. 2, pp. 269–282, 2004.
- [32] M. Li, "A class of negatively fractal dimensional Gaussian random functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 291028, 18 pages, 2011.
- [33] J. Beran, *Statistics for Long-Memory Processes*, vol. 61 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York, NY, USA, 1994.
- [34] M. Li and S. C. Lim, "Modeling network traffic using generalized Cauchy process," *Physica A*, vol. 387, no. 11, pp. 2584–2594, 2008.
- [35] R. L. Cruz, "A calculus for network delay—part I: network elements in isolation—part II: network analysis," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 114–141, 1991.
- [36] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [37] R. Albert, H. Jeong, and A. L. Barabási, "Internet: diameter of the world-wide web," *Nature*, vol. 401, no. 6749, pp. 130–131, 1999.
- [38] B. A. Huberman and L. A. Adamic, "Internet: growth dynamics of the world-wide web," *Nature*, vol. 401, no. 6749, p. 131, 1999.
- [39] R. Webster and M. A. Oliver, *Geostatistics for Environmental Scientists*, John Wiley & Sons, 2007.
- [40] H. Wackernagel, *Multivariate Geostatistics: An Introduction with Applications*, Springer, 2005.
- [41] P. Whittle, "On the variation of yield variance with plot size," *Biometrika*, vol. 43, no. 3-4, pp. 337–343, 1962.
- [42] 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.
- [43] M. Lefebvre, "A one- and two-dimensional generalized Pareto model for a river flow," *Applied Mathematical Modelling*, vol. 30, no. 2, pp. 155–163, 2006.
- [44] A. J. Lawrance and N. T. Kottegoda, "Stochastic modelling of riverflow time series," *Journal of the Royal Statistical Society. Series A*, vol. 140, no. 1, pp. 1–47, 1977.
- [45] H. A. Makse, S. Havlin, and H. E. Stanley, "Modelling urban growth patterns," *Nature*, vol. 377, no. 6550, pp. 608–612, 1995.
- [46] T. von Kármán, "Progress in the statistical theory of turbulence," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 34, no. 11, pp. 530–539, 1948.
- [47] G. H. Goedecke, V. E. Ostashev, D. K. Wilson, and H. J. Auvermann, "Quasi-wavelet model of von Kármán spectrum of turbulent velocity fluctuations," *Boundary-Layer Meteorology*, vol. 112, no. 1, pp. 33–56, 2004.
- [48] M. C. H. Hui, A. Larsen, and H. F. Xiang, "Wind turbulence characteristics study at the Stonecutters Bridge site—part II: wind power spectra, integral length scales and coherences," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 97, no. 1, pp. 48–59, 2009.
- [49] A. G. Davenport, "The spectrum of horizontal gustiness near the ground in high winds," *Quarterly Journal of the Royal Meteorological Society*, vol. 87, no. 372, pp. 194–211, 1961.
- [50] J. C. Kaimal, J. C. Wyngaard, Y. Izumi, and O. R. Coté, "Spectral characteristics of surface-layer turbulence," *Quarterly Journal of the Royal Meteorological Society*, vol. 98, no. 417, pp. 563–589, 1972.
- [51] I. Antoniou, D. Asimakopoulos, A. Fragoulis, A. Kotronaros, D. P. Lalas, and I. Panourgias, "Turbulence measurements on top of a steep hill," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 39, no. 1–3, pp. 343–355, 1992.
- [52] D. Hiriart, J. L. Ochoa, and B. García, "Wind power spectrum measured at the San Pedro Mártir Sierra," *Revista Mexicana de Astronomía y Astrofísica*, vol. 37, no. 2, pp. 213–220, 2001.
- [53] V. Pisarenko and M. Rodkin, *Heavy-Tailed Distributions in Disaster Analysis*, vol. 30, Springer, 2010.
- [54] C. A. Monje, Y.-Q. Chen, B. M. Vinagre, D. Xue, and V. Feliu, *Fractional Order Systems and Controls—Fundamentals and Applications*, Springer, 2010.
- [55] M. D. Ortigueira, "An introduction to the fractional continuous-time linear systems: the 21st century systems," *IEEE Circuits and Systems Magazine*, vol. 8, no. 3, pp. 19–26, 2008.
- [56] Y. Q. Chen and K. L. Moore, "Discretization schemes for fractional-order differentiators and integrators," *IEEE Transactions on Circuits and Systems. I*, vol. 49, no. 3, pp. 363–367, 2002.
- [57] B. M. Vinagre, Y. Q. Chen, and I. Petráš, "Two direct Tustin discretization methods for fractional-order differentiator/integrator," *Journal of the Franklin Institute*, vol. 340, no. 5, pp. 349–362, 2003.
- [58] S. C. Lim and S. V. Muniandy, "Self-similar Gaussian processes for modeling anomalous diffusion," *Physical Review E*, vol. 66, no. 2, Article ID 021114, 14 pages, 2002.

- [59] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall, New York, NY, USA, 1994.
- [60] 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.
- [61] 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.
- [62] M. Li, W. S. Chen, and L. Han, "Correlation matching method for the weak stationarity test of LRD traffic," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 181–195, 2010.
- [63] M. Li, S. C. Lim, and S. Y. 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.
- [64] 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.
- [65] P. Hall and Q. Yao, "Inference in ARCH and GARCH models with heavy-tailed errors," *Econometrica*, vol. 71, no. 1, pp. 285–317, 2003.
- [66] F. Kamoun, "Performance analysis of a discrete-time queuing system with a correlated train arrival process," *Performance Evaluation*, vol. 63, no. 4-5, pp. 315–340, 2006.
- [67] J. K.-Y. Ng, S. Song, and W. Zhao, "Statistical delay analysis on an ATM switch with self-similar input traffic," *Information Processing Letters*, vol. 74, no. 3-4, pp. 163–173, 2000.
- [68] Z. Song, Y.-Q. Chen, C. R. Sastry, and N. C. Tas, *Optimal Observation for Cyber-Physical Systems*, Springer, 2009.
- [69] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.
- [70] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.
- [71] G. Mattioli, M. Scalia, and C. Cattani, "Analysis of large-amplitude pulses in short time intervals: application to neuron interactions," *Mathematical Problems in Engineering*, vol. 2010, Article ID 895785, 15 pages, 2010.
- [72] S. Y. Chen, Y. F. Li, and J. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
- [73] S. Y. Chen and Y. F. Li, "Vision sensor planning for 3-D model acquisition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 35, no. 5, pp. 894–904, 2005.
- [74] W. B. Mikhael and T. Yang, "A gradient-based optimum block adaptation ICA technique for interference suppression in highly dynamic communication channels," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 84057, 2006.
- [75] E. G. Bakhoun 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, 2010.
- [76] E. G. Bakhoun 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.
- [77] Z. Liao, S. Hu, and W. Chen, "Determining neighborhoods of image pixels automatically for adaptive image denoising using nonlinear time series analysis," *Mathematical Problems in Engineering*, vol. 2010, Article ID 914564, 2010.
- [78] Z.-W. Liao, S.-X. Hu, D. Sun, and W. F. 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.
- [79] J.-W. Yang, Z.-R. Chen, W.-S. Chen, and Y.-J. Chen, "Robust affine invariant descriptors," *Mathematical Problems in Engineering*, vol. 2011, Article ID 185303, 15 pages, 2011.

Research Article

Mixed Signature: An Invariant Descriptor for 3D Motion Trajectory Perception and Recognition

**Jianyu Yang,^{1,2,3} Y. F. Li,^{1,3} Keyi Wang,^{2,3} Yuan Wu,⁴
Giuseppe Altieri,⁵ and Massimo Scalia⁶**

¹ Department of MEEM, City University of Hong Kong, Hong Kong

² Department of PMPI, University of Science and Technology of China, Hefei 230027, China

³ Department of MEEM, USTC-CityU Joint Advanced Research Centre, Suzhou 215123, China

⁴ Department of CS, University of Science and Technology of China, Hefei 230027, China

⁵ Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy

⁶ Department of Mathematics, Sapienza University of Rome, P.le A. Moro 2, 00185 Rome, Italy

Correspondence should be addressed to Y. F. Li, meyfli@cityu.edu.hk

Received 30 March 2011; Accepted 27 April 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Jianyu Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motion trajectory contains plentiful motion information of moving objects, for example, human gestures and robot actions. Motion perception and recognition via trajectory are useful for characterizing them and a flexible descriptor of motion trajectory plays important role in motion analysis. However, in the existing tasks, trajectories were mostly used in raw data and effective descriptor is lacking. In this paper, we present a mixed invariant signature descriptor with global invariants for motion perception and recognition. The mixed signature is viewpoint invariant for local and global features. A reliable approximation of the mixed signature is proposed to reduce the noise in high-order derivatives. We use this descriptor for motion trajectory description and explore the motion perception with DTW algorithm for salient motion features. To achieve better accuracy, we modified the CDTW algorithm for trajectory matching in motion recognition. Furthermore, a controllable weight parameter is introduced to adjust the global features for tasks in different circumstances. The conducted experiments validated the proposed method.

1. Introduction

Motion trajectory is a sequence of positions of moving object by time series in spatiotemporal motion. For motions of human hand gestures, body movements, robot arm actions, and other complex long-term motions, we can record them by their trajectories. Motion trajectory contains dynamic information which is useful for characterizing and identifying motions.

Trajectory-based motion analysis is important for many applications, such as human behavior recognition [1], human intention perception with prediction, motion modeling for robot *Learning by Demonstration* (LbD) [2, 3], and other motion analysis. Most of these applications are useful for human-robot interaction in which motion analysis of human and robot actions are important [4]. As behaviors and activities are mostly performed in space, 3D trajectory analysis is much more important than 2D curve approaches. The information contained in trajectory can be extracted and used in motion perception and recognition. For example, in a surveillance application, the intentions of people need to be captured by perceiving their motions and predicting the further motions of people. Here, an effective trajectory descriptor for motion modeling which can extract adequate and overall motion information is valuable for motion description, perception, and recognition.

In the existing work, trajectory was mostly used in raw data directly [5]. However, raw data rely on the absolute positions of motion and are thus ineffective in computation and sensitive to noise. They are incapable of capturing shapes and detailed features under changes of viewpoints. Therefore, most information in 3D space cannot be captured directly from raw data, and a flexible and adaptable trajectory descriptor for 3D motions is needed. The flexibility and adaptability here can be measured with four criterions: (1) ability of capturing salient features, (2) accuracy of charactering motion trajectories, (3) efficiency in computation, and (4) invariant in different viewpoints and circumstances.

To get the salient features of motion, the concept of shape descriptor was developed. Most of shape descriptors lack adaptability for tasks under different environments, such as *Centroid-Contour Distance* (CCD) and R-S curve. Chain code [6] is a discrete method which makes the trajectory samples transformed to an apex point of the square around the sample position. Shape indexing approach [7] uses tokens to describe local shapes with their curvatures and directions via the M-tree structure. Another local shape description is the *Dominant Polygonal* (DP) [8], which uses polygonal curve to describe local dominant shapes. Contour functions [9] are used to describe local shapes, such as Cross-Section Function, Radius-Vector Function, Support Function, and Tangent-Angle Function. Curvature scale method is also used as curve description [10]. These methods are limited under spatial transformation and can only be used for simple shapes of regular human gestures, which is not appropriate for complex long term motions.

Mathematical descriptors were also used to describe motion trajectories, such as B-Spline [11], NURBS [12], and Bezier curve [13]. Trajectories and curves are represented by a group of control point parameters of the spline curve. To obtain the spline parameters for shape representation, data fitting of these mathematical descriptors is necessary. In this way, inaccuracy of them will be inevitably caused.

Transform functions were used in the existing work to describe the global features of curves and motion trajectories, such as *Fourier Descriptor* (FD) [14], wavelet coefficients [15, 16], and radon transform [17]. Shapes of curve and trajectory are represented as a whole by these transformation methods, where the local features were lost and the normalization process is always a problem. Therefore, the transform functions are not suitable as descriptors here.

Invariant descriptors are flexible for motion trajectory representation in different circumstances. Geometric invariants [18] are invariant at shapes of curve that can be used for the same shape in different viewpoints or transformation, as well as the algebraic invariants [19, 20]. The moment invariant function [21] is also an invariant description which abstracts the global features, but they are not capable for the complex and long-term trajectory. The method in [22] is viewpoint invariant for projected two-dimensional recognition rather than

real spatial data analysis. Several relevant articles are inspired in our work [23–25] for motion analysis.

Some new methods of shape description are used for motion analysis in the present study. Ogawara et al. [26] use motion density to detect repeated motion patterns, which is only useful for coincide motions. Prati et al. [27] use motion trajectory angles for description, Faria and Dias [28] use curvature and orientation to model hand motion trajectories, which are limited for simple shapes, and more than 3D coordinates will be needed. The method in [29] is viewpoint invariant, but it is only useful for analysis of the periodic motion. The context of motion is used for modeling trajectories in [30], but the composite of multitrajectory context are not flexible for single point in complex trajectory analysis.

Differential invariants in the previous signature descriptor are invariant in translation, rotation, scaling, and occlusion [31–33]. The signature descriptor is adaptable for motions under these spatial transform and is efficient in motion analysis. These differential invariants are good for describing local shapes, but global features are lost in this descriptor. Trajectories are usually characterized as a whole in motion analysis, and the global information of each point in a trajectory is also important for these tasks.

In this paper, we propose a new descriptor—mixed signature with not only differential invariants but also global invariants. The global invariants perform well in capturing global features which are necessary for trajectory perception and recognition. Apart from containing global information, the mixed signature inherits the advantage of the previous signature descriptor: the mixed signature is also invariant in spatial transformation, including translation, rotation, scaling, and occlusion. We use this descriptor to model motion trajectory for motion perception and recognition. A large database [34] is used in the experiment for testing the efficiency of our method.

The reminder of this paper is organized as follows. Section 2 presents the definition of the mixed signature descriptor. Theories of motion perception and recognition based on this descriptor are expatiated in Sections 3 and 4 separately. Section 5 presents experiments and result analysis. This paper is concluded in Section 6.

2. Mixed Signature for Trajectory Representation

A motion trajectory is a sequence of discrete points which represent the positions of the moving object in every frame. The 3D coordinates of these positions are raw data of trajectory, denoted as $\Gamma(t) = \{X(t), Y(t), Z(t) \mid t \in [1, N]\}$, where t is the index number of the frame sequence and N is the trajectory length. Figure 1 shows a piece of 3D motion trajectory.

The mixed signature \bar{S} of $\Gamma(t)$ is defined with differential invariants and global invariants as follows:

$$\bar{S} = \{k(t), k_s(t), \tau(t), \tau_s(t), s(t), r(t) \mid t \in [1, N]\}, \quad (2.1)$$

where

$$k(t) = \frac{\|\Gamma'(t) \times \Gamma''(t)\|}{\|\Gamma'(t)\|^3}, \quad (2.2)$$

$$\tau(t) = (\Gamma'(t) \times \Gamma''(t)) \cdot \frac{\Gamma'''(t)}{\|\Gamma'(t) \times \Gamma''(t)\|^2}, \quad (2.3)$$

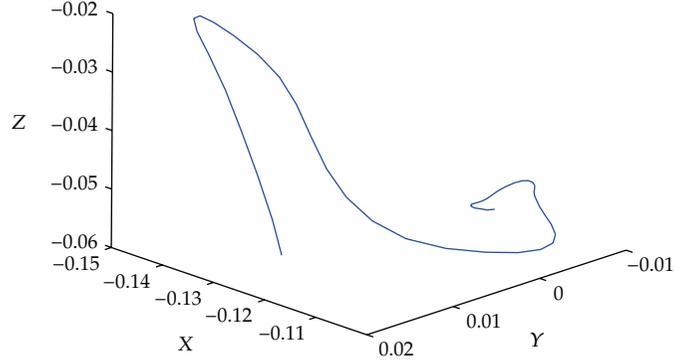


Figure 1: A piece of 3D motion trajectory.

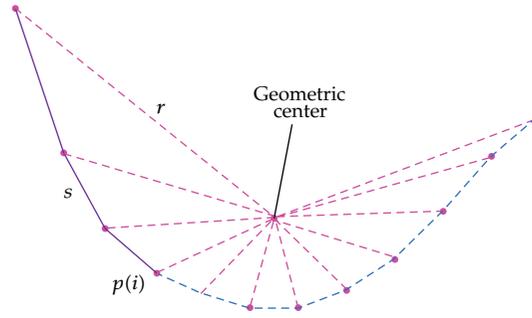


Figure 2: Global invariants r and s of mixed signature, where $p(i)$ is the present point.

$$k_s(t) = \frac{dk(t)}{ds} = \frac{dk(t)}{dt} \cdot \frac{dt}{ds} = \frac{k'(t)}{\|\Gamma'(t)\|}, \quad (2.4)$$

$$\tau_s(t) = \frac{d\tau(t)}{ds} = \frac{d\tau(t)}{dt} \cdot \frac{dt}{ds} = \frac{\tau'(t)}{\|\Gamma'(t)\|}, \quad (2.5)$$

$$s(t) = \int_1^t \|\Gamma'(t)\| dt, \quad (2.6)$$

$$r(t) = \|\Gamma(t) - c\| = \left\| \Gamma(t) - \frac{\int_1^N \Gamma(t) dt}{N} \right\|. \quad (2.7)$$

The parameters defined in (2.2)–(2.5) are the differential invariants: curvature k , torsion τ and their derivatives k_s and τ_s ; here, s is arc length defined in (2.6). As these differential invariants have been presented in the previous signature descriptor [31–33], we only discuss the global invariants in this paper. Parameter s denotes the arc length from the beginning of the trajectory to the present point, which represents different phases of the motion. The geometric centre of the trajectory is denoted by c , and parameter r represents the geometric distance from the present point to the center c . These two global invariants capture the relation between the present point and the whole trajectory, as illustrated in Figure 2.

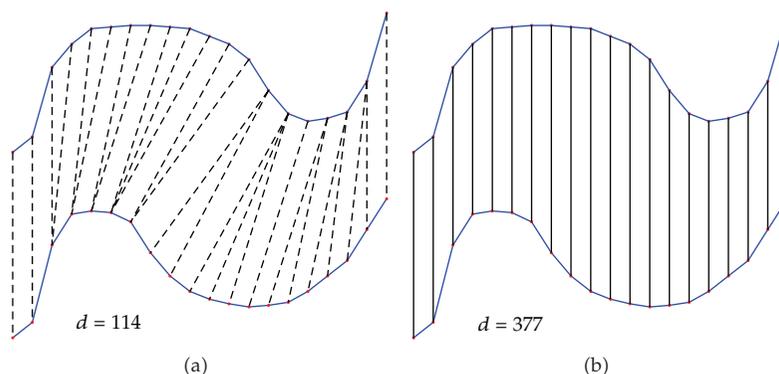


Figure 3: These two figures show the effect of global features. The result of matching gives small distance in (a) with only local features, but they are significantly different if the global features are not ignored in (b).

These two global features are necessary for the requirement of global information, especially when the motion is characterized as a whole. For example, when we demonstrate a task to a mobile robot as PbD learning, the distribution of different parts of the motion cannot be ignored for the integral structure of the task. Figure 3 shows a case of motion trajectory classification with *Dynamic Time Warping* (DTW) algorithm and Euclidean distance, respectively, where different distributions of similar local features will significantly alter the result. The parameter d in Figure 3 denotes the distance-degree of difference between the trajectories in matching. In this case, the trajectories will be wrongly classified by DTW with only local features which will lead to motion confusion. The raw data contain much more global information because it is represented by the absolute position of 3D space. Hence, the matching result in Figure 3(b) is significantly different from that in Figure 3(a). However, the raw data is inflexible that most motion information cannot be extracted and is not invariant in spatial transformation as the signature invariants. Therefore, the mixed signature is a tradeoff that preserves the particularity of invariant and captures the global information as much as possible.

The global invariants r and s illustrate the relation between $p(i)$ and the whole trajectory, which are independent from space transformation. No matter how the viewpoint is changed or the scale of trajectory altered, the value of the global invariants will be steady for every sample. In this way, the mixed signature inherits the invariant of translation, rotation, scaling, and occlusion of the previous signature with extra global information. Figure 4 shows the six invariants of the mixed signature in time index.

As the trajectory is a discrete sequence, the calculation of the high-order derivatives and integrals of the mixed signature will result in errors due to its sensitivity to noise. To avoid calculating the high-order derivatives directly, we replace the accurate invariants with an approximation based on several neighboring points of the sample. This approximation will reduce the sensitivity to noise with only the lowest-order derivatives [35]. The approximations of differential invariants have been discussed in [31–33] and we only present the approximations of global invariants in this paper.

As shown in Figure 2, we use line segments between points instead of the arc length, and the approximation s^* is the sum length from the beginning to the present point $p(i)$.

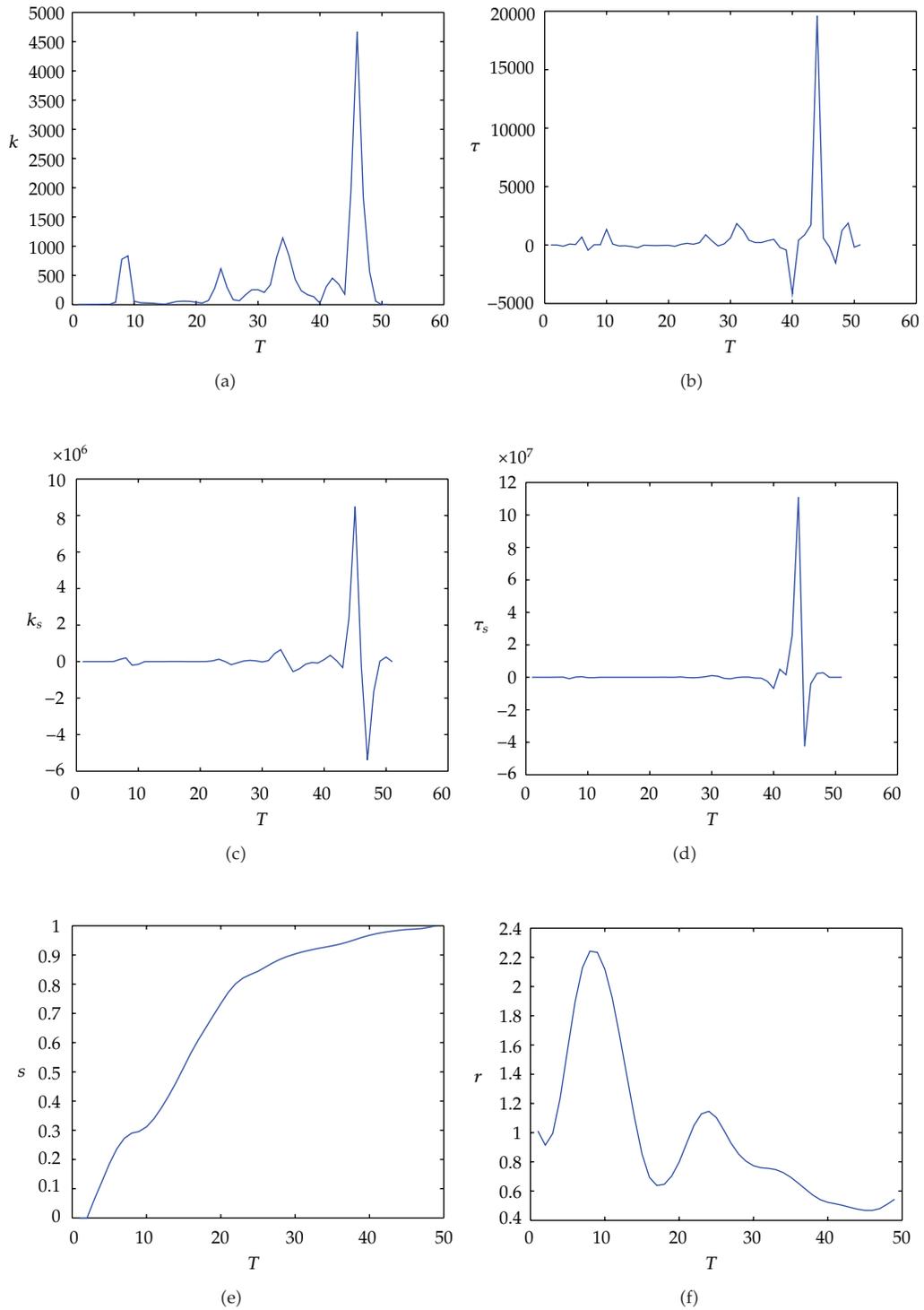


Figure 4: The mixed signature of the trajectory in Figure 1.

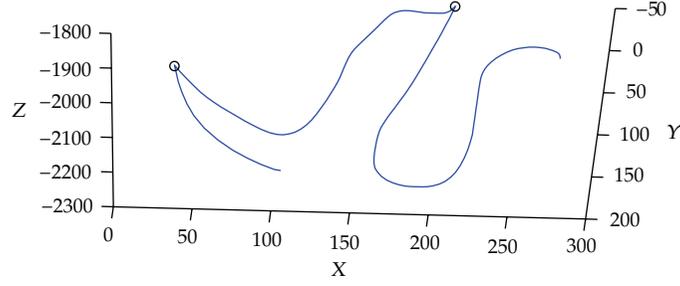


Figure 5: A trajectory from a dance fragment along which the zero speed moments are highlighted by circles.

Similarly, geometric center c^* is calculated by averaging all the points of trajectory. Equation (2.8) is definition of approximations s^* and r^* that normalized by the whole trajectory

$$s^*(t) = \frac{\sum_{i=2}^t \|p(i) - p(i-1)\|}{\sum_{i=2}^N \|p(i) - p(i-1)\|} = \frac{\sum_{i=2}^t \|(x_i, y_i, z_i) - (x_{i-1}, y_{i-1}, z_{i-1})\|}{\sum_{i=2}^N \|(x_i, y_i, z_i) - (x_{i-1}, y_{i-1}, z_{i-1})\|}, \quad (2.8)$$

$$r^*(t) = \frac{N \cdot \|p(i) - c^*\|}{\sum_{i=1}^N \|p(i) - c^*\|} = \frac{N \cdot \|(x_i, y_i, z_i) - c^*\|}{\sum_{i=1}^N \|(x_i, y_i, z_i) - c^*\|},$$

where

$$c^* = \sum_{i=1}^N \frac{p(i)}{N} = \sum_{i=1}^N \frac{(x_i, y_i, z_i)}{N}. \quad (2.9)$$

The mixed signature is calculated with the discrete samples of trajectory, and there maybe more than one sample at the same position for the condition that $\Gamma'(t) = 0$. We call this position the stationary point, and there are two types of stationary point in motion trajectory. One is the case that the motion alters its direction and then the object must have a moment that the speed is zero at the altered point. This case is shown in Figure 5. This point is important for the motion analysis in perception and recognition. The other case is the break of motion while the behavior accidentally holds on, and the position is recorded for several frames. This point is the noise in the motion analysis, and the repeatedly sampled points should be removed with only one sample staying in the trajectory.

The noise in trajectory will affect the calculation of the invariants, because the differential invariants are local parameters which depend only on several nearby samples. Trajectory smoothing is an important process to reduce the noise and vibration in a trajectory for accurate calculation of invariants. In our method, we use the wavelet smoother and it proved to be effective with acceptable shape deviations [31]. In this process of smoothing by the wavelet smoother, trajectory shape can also be preserved, as the decomposition level of wavelet smoother can be tuned according to the noise strength (see Figure 6).

In a real motion trajectory, there are usually some outlier samples which contain significant errors and will affect the calculation of invariants badly. The single distance

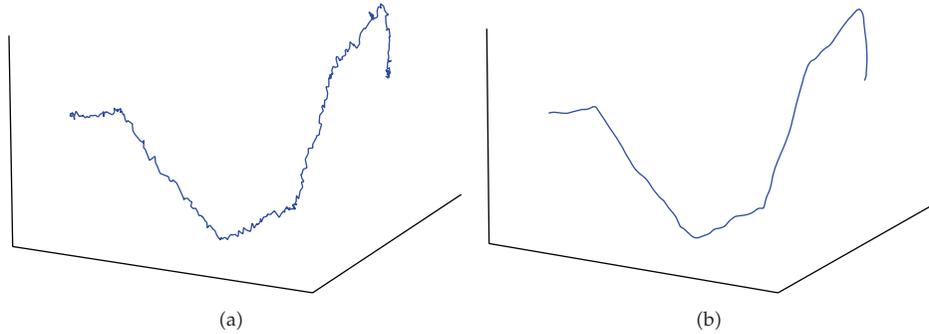


Figure 6: Smooth for noise trajectory by the wavelet smoother used in [30].

between outlier samples on separate trajectories is larger than the normal ones. In this paper, we set a threshold for single distance of every pair of corresponding points to filtrate the outlier samples. The threshold can be set in two ways: via local threshold and global threshold. The local one is calculated dynamically with the distances of nearby samples. The global one is calculated with the distances to a mean route which is calculated beforehand without any threshold. For general motion trajectories, the global one generally works well. However, in some cases where the motions are complex with different working situations, trajectories will suffer from varying covariance of distances. For this reason, the local threshold performs better than the global one.

3. Motion Perception

Motion perception is an important method in analysis of human and robot gestures by capturing their salient features. From these features, we can perceive the intention of motion or identify motions from database. Salient features can be captured by properties of motion, such as speed, symmetry, period, and feature shapes. These properties have been studied in existing tasks for motion analysis [29], and we discuss the symmetry and period to elaborate the mixed signature for motion perception.

3.1. Motion Symmetry Perception

As we all know, there is a symmetrical point in the center of the symmetrical part of the trajectory, denoted $p(i)$. All the corresponding points are symmetrical related to this point. The symmetrical part may be only a segment of a trajectory. We classify all the symmetrical conditions in 3D space into two basic classes. In the first class, the corresponding points have the same distance to the symmetrical plane. That is to say, this only plane is a mirror between the corresponding points and one point is the image of the other one. This case is shown in Figure 7, denoted mirror symmetry. In the other class, the corresponding points have the same distance to the center point. One part will cover the other part if rotate a certain angle around the center point. We denote this case central symmetry as shown in Figure 8. The methods to perceive these two classes of symmetrical motions are described in two theorems as follows.

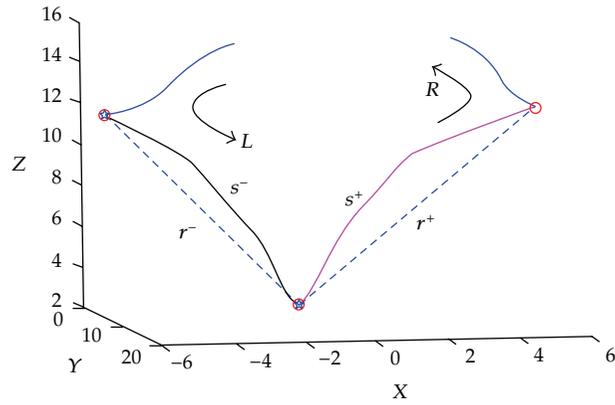


Figure 7: Mirror symmetry trajectory with Δr and Δs in corresponding segments. Arrows show the left-hand helix and right-hand helix.

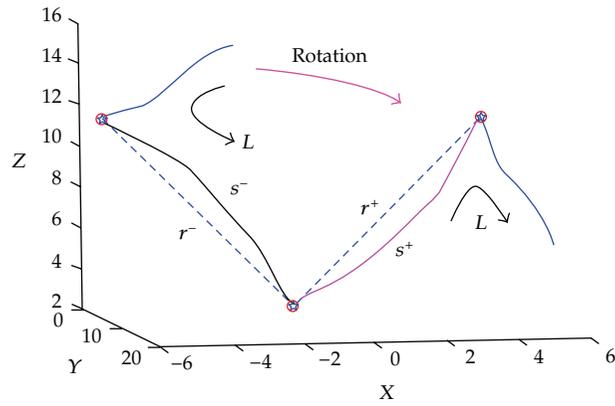


Figure 8: Central symmetry trajectory with Δr and Δs in corresponding segments. Arrows show the left-hand helix in both parts.

Theorem 3.1. As $p(i)$ is the symmetrical center, a pair of corresponding symmetrical points, for example, $p(i - 1)$ and $p(i + 1)$ hold the following relations:

$$k(i - 1) = k(i + 1), \tag{3.1}$$

$$k_s(i - 1) = -k_s(i + 1), \tag{3.2}$$

$$\tau(i - 1) = -\tau(i + 1), \tag{3.3}$$

$$\tau_s(i - 1) = \tau_s(i + 1), \tag{3.4}$$

$$s(i - 1) - s(i) = s(i) - s(i + 1), \tag{3.5}$$

$$\|r(i - 1) - r(i)\| = \|r(i) - r(i + 1)\|. \tag{3.6}$$

Proof. The symmetrical part composed with two corresponding segments which are the same in shape, so that $k(i-1) = k(i+1)$. A left-hand helix will turn to a right-hand helix via a mirror, which explains $\tau(i-1) = -\tau(i+1)$. As (3.1) proves that the function $k(i+x)$ is an even function, the derivatives of k with respect to s should be an odd function, and we have $k_s(i-1) = -k_s(i+1)$. Similarly, we have $\tau_s(i-1) = \tau_s(i+1)$. The distances and arc lengths from $p(i)$ to $p(i-1)$ and $p(i+1)$ should be the same, as shown in Figure 7. Hence, we have $s(i-1) - s(i) = s(i) - s(i+1)$ and $\|r(i-1) - r(i)\| = \|r(i) - r(i+1)\|$. \square

Theorem 3.2. *As one part will cover the corresponding part of the trajectory by rotating round the center point, the properties of corresponding points will be the same. Hence, a pair of corresponding symmetrical points, for example, $p(i-1)$ and $p(i+1)$, hold the following relations:*

$$k(i-1) = k(i+1), \quad (3.7)$$

$$k_s(i-1) = -k_s(i+1), \quad (3.8)$$

$$\tau(i-1) = \tau(i+1), \quad (3.9)$$

$$\tau_s(i-1) = -\tau_s(i+1), \quad (3.10)$$

$$s(i-1) - s(i) = s(i) - s(i+1), \quad (3.11)$$

$$\|r(i-1) - r(i)\| = \|r(i) - r(i+1)\|. \quad (3.12)$$

Proof. The curvature k and its derivative k_s are the same as those of Theorem 3.1, and (3.11)-(3.12) are the same as (3.5)-(3.6) as well. The torsion direction will not change while rotation, and the corresponding parts in Figure 8 are both left hand helix. Then, we have $\tau(i-1) = \tau(i+1)$ and $\tau_s(i-1) = -\tau_s(i+1)$. \square

We can perceive motion symmetry by detecting the relevant properties listed in these two theorems. The central point should be located first, according to $k_s(i) = 0$ and $\tau(i) = 0$ for mirror symmetry while $k_s(i) = 0$ and $\tau_s(i) = 0$ for central symmetry. Then, we check every pair of corresponding points via the equations in the theorems and confirm the length of the symmetry part. If the length of corresponding points is zero, this center point should be discarded. Equations (3.5)-(3.6) with global parameters are necessary in motion perception, which were not considered in previous signature descriptor. Due to the scaling invariant of differential invariants, the trajectory segments with similar shapes in different sizes will be erroneously perceived as symmetry without these global equations. We will discuss this condition in the implementations in Section 5.2.

For better analysis of motion symmetry, the coordinates of k/k_s and τ/τ_s are useful for extracting the properties. Figures 9 and 10 show the coordinates of the subsignature in the two classes of symmetry. From the figures we can figure out the center point and there maybe multiple center points in one trajectory. The symmetrical properties of the differential invariants are also observed in these figures.

3.2. Periodic Motion Perception

Periodic motion occurring in human and robot activity is usual, and motion analysis will capture more information in repeated tasks by perceiving this motion property. We can

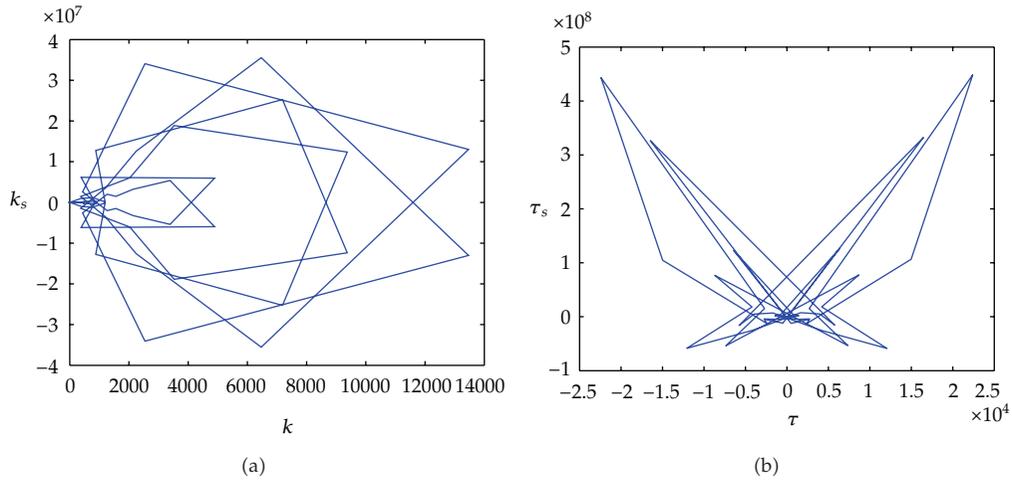


Figure 9: The subcoordinates of the mirror symmetrical motion trajectory. (a) is k/k_s and (b) is τ/τ_s .

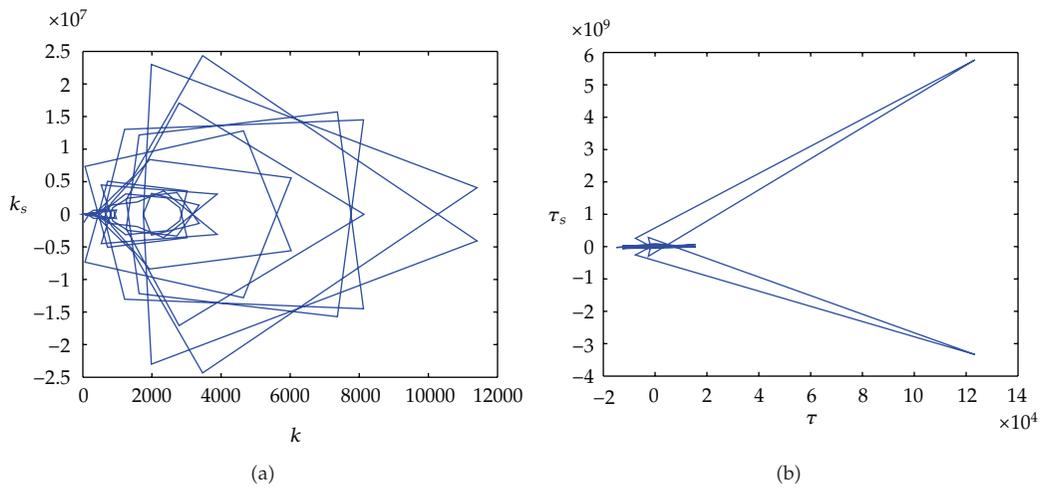


Figure 10: The subcoordinates of the central symmetrical motion trajectory. (a) is k/k_s and (b) is τ/τ_s .

perceive this feature according to the properties of periodic trajectory. Periodic motion is periodic in almost all of the motion features, such as speed, direction, shape, and displacement. All of these periodic features are indexed by the period T —the distance between neighboring periods, and they are generally expressed with the displacement function

$$f(t) = f(t + nT), \tag{3.13}$$

where f denotes motion features and n is an integer. The features can be well represented by the invariant signature under spatial transformation. The periodic property cannot be represented accurately with only differential invariants. Rather, geometric distance r and s

is more reliable here, which will ensure the features periodic in 3D space. Not only the local shapes but also the vectors are equal between neighboring periods as follows.

Theorem 3.3. *A pair of corresponding points $p(i)$ and $p(i + T)$ in different periods have the relations as follows:*

$$k(i) = k(i + T), \quad (3.14)$$

$$\tau(i) = \tau(i + T), \quad (3.15)$$

$$k_s(i) = k_s(i + T), \quad (3.16)$$

$$\tau_s(i) = \tau_s(i + T), \quad (3.17)$$

$$s(i) - s(i + T) = s(i + 1) - s(i + T + 1), \quad (3.18)$$

$$r(i) - r(i + T) = r(i + 1) - r(i + T + 1). \quad (3.19)$$

Proof. From (3.7), we can conclude the four equations of differential invariants directly. In 3D space, a segment of one period will be coincident with another period by translating along a vector. In this way, all the vectors between corresponding points are equal to this vector. Then, (3.18)-(3.19) can be inferred. \square

For periodic motion perception, we should confirm the period and starting point first. Then, we need to examine whether all the corresponding points in periodic trajectory satisfy the equations in Theorem 3.3. Therefore, we can perceive periodic motion only via a single trajectory without any database beforehand. The advantage of the mixed signature with global information will be illustrated in the experiments in Section 5.3.

3.3. Feature Perception via DTW Algorithm

Corresponding sample alignment and match is necessary in perception of symmetrical motion and periodic motion. We need to compare the invariants by the theorems between corresponding samples. However, as there will be a difference in sample rate or distribution of points that makes the samples not in the corresponding points, the comparison between corresponding samples should not directly use the equations in the theorems. The samples cannot be aligned one by one, and an appropriate method of alignment for matching samples is necessary. In this paper, we use a nonlinear alignment method—DTW algorithm [36], which can find the best alignment between corresponding segments according to the theorems. DTW is effective at similarity measurement that we can perceive motion by matching a segment of trajectory with the feature segments in a database. The segments in database indicate different motion features, and we can infer the intension of the motion by perceiving them.

DTW—Dynamic Time Warping algorithm—is to calculate the best correspondence of samples between two trajectory segments for the minimum distance (see Figure 11). This distance can be defined according to the demand of tasks to calculate the similarity. The alignment in every step of matching relies on the minimum sum distance

$$D(m, n) = \min\{D(m - 1, n), D(m, n - 1), D(m - 1, n - 1)\} + d(m, n), \quad (3.20)$$

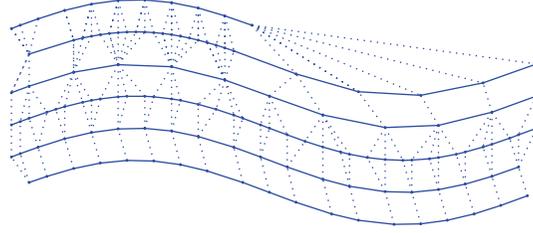


Figure 11: Trajectory sampling instances (in solid lines) and the DTW based nonlinear paths warping (in dotted lines).

where $d(m, n)$ denotes the distance of sample m and n in respective trajectory and $D(m, n)$ is the sum distance up to them.

The distance between corresponding samples in our work is defined with the descriptor of trajectory. For two trajectories A and B with respective lengths M and N , the distance between samples $A(m)$ and $B(n)$ is defined with the approximate mixed signature as follows:

$$d(m, n) = \Delta \bar{S}^{*m,n} = \frac{\Delta k^{*m,n} \cdot \Delta \tau^{*m,n} \cdot (1 + \lambda \cdot \Delta h^{*m,n})}{\sqrt{(S^{*m})^2} \cdot \sqrt{(S^{*n})^2}}, \quad (3.21)$$

where

$$\Delta k^{*m,n} = \|(k^{*m}, k_s^{*m}) - (k^{*n}, k_s^{*n})\|, \quad (3.22)$$

$$\Delta \tau^{*m,n} = \|(\tau^{*m}, \tau_s^{*m}) - (\tau^{*n}, \tau_s^{*n})\|, \quad (3.23)$$

$$(S^{*m})^2 = (k^{*m})^2 + (k_s^{*m})^2 + (\tau^{*m})^2 + (\tau_s^{*m})^2, \quad (3.24)$$

$$(S^{*n})^2 = (k^{*n})^2 + (k_s^{*n})^2 + (\tau^{*n})^2 + (\tau_s^{*n})^2, \quad (3.25)$$

$$\Delta h^{*m,n} = \|(s^*(m), r^*(m)) - (s^*(n), r^*(n))\|, \quad (3.26)$$

Equations (3.22)–(3.25) are the same as the definition in [31] with the previous differential signature S . Here, the parameter λ is the weight of global invariants in calculating the distance of trajectories. The weight of global information depends on two aspects: the circumstance for sampling and demand of tasks. The setting of this parameter will be presented in the next section.

From Figure 11, we can see that corresponding samples in different frame rate and distribution will be matched by DTW algorithm. This method considers the minimum distance of relative features, which is suitable for this task. The mixed signature is flexible and adaptable for this algorithm, because only the differential invariants in the previous signature descriptor are not enough. The matching in Figure 3(a) is a case of DTW matching by only differential invariants, and we can see that corresponding points are similar at local features. In this way, the distance between the two trajectories appears small while they are of different classes. To solve this problem, global information should be considered in distance calculation as (3.21). We will present the flexibility of mixed signature in the experiments by comparison with the previous signature.

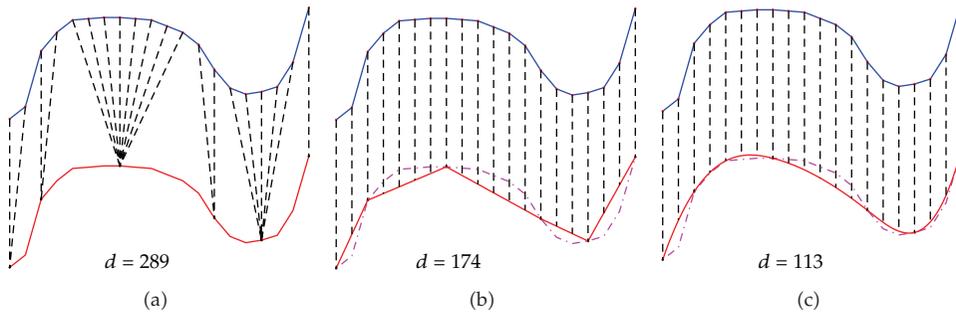


Figure 12: Matching two similar trajectories (whose distance should be 0) using DTW, CDTW, and our method. The distance of (c) is much more accurate with cubic polynomials interpolation. The curves of dashed line in (b) and (c) are the original trajectory.

4. Motion Recognition

Motion recognition is to classify motions with a database, where flexible descriptor and recognition algorithm are both crucial for good performance. The DTW algorithm is effective to overcome the diversity in motion speed and frame rate which will lead to different sample rate and distribution. However, this method has some disadvantages which will be inaccuracy in trajectory recognition. For example, if the sampling in one trajectory is sparse while the other is not (Figure 12(a)), the samples far from the corresponding position will cause a large distance due to the fact that DTW matches only discrete samples rather than continuous curves. Munich and Perona proposed a *Continuous Dynamic Time Warping* (CDTW) algorithm to explore a solution for this problem [37].

The alignment method in CDTW is suitable for curve matching rather than motion recognition because it does not consider the properties of kinematics. Figure 12 shows the results of matching by different algorithms. The trajectories in Figure 12 are the same motion under two sampling schemes. It shows that different match algorithms lead to different results. The linear interpolation of the trajectories in Figure 12(b) is not accurate for motion trajectory and the method in [37] is complex. Proper interpolation method considering motion properties is needed, and concise conditions will reduce the complexity of the algorithm.

In this section, we modified the CDTW algorithm for motion recognition with the mixed signature descriptor. The time warping method used in [37] classified the path of alignment between two trajectories into four matching conditions in the algorithm. Those four conditions are complex in calculation, and we simplify them into only two conditions in our algorithm. We use two subitems to express the distances of two conditions and calculate the minimum of them as follows:

$$D(m, n) = \min\{D(m, n-1) + d(m+\theta, n), D(m-1, n) + d(m, n+\theta)\}, \quad (4.1)$$

where m, n are sequence lengths of trajectory A and B , θ is a parameter between 0 and 1, $d(m, n)$ is the distance between samples $A(m)$ and $B(n)$, and $d(m+\theta, n)$ is the distance between samples $A(m+\theta)$ and $B(n)$. Here, $A(m+\theta)$ is a point moving on the trajectory A between samples $A(m)$ and $A(m+1)$ as shown in Figure 13.

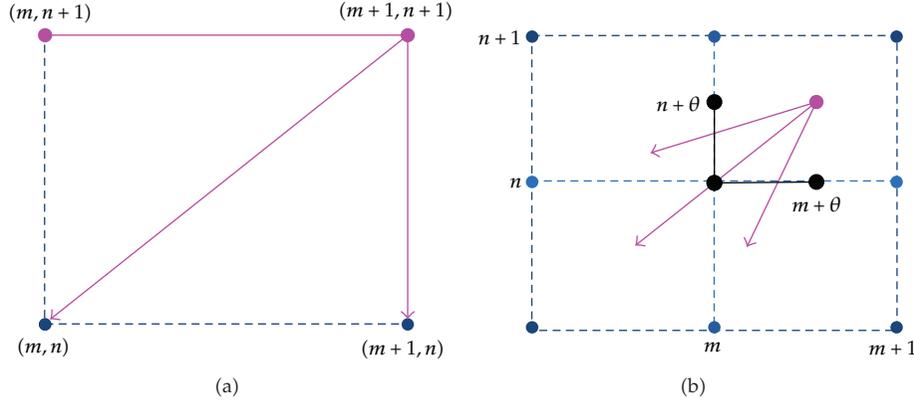


Figure 13: These two figures show the matching blocks of DTW in (a) and our method in (b). The matching in (a) can only choose 3 points: $(m, n+1)$, (m, n) , and $(m+1, n)$ while all the points on real lines can be matched in (b).

For the matching between $A(m)$ and $B(n)$ in DTW, the corresponding points can only be three positions: the three intersections (m, n) , $(m+1, n)$, $(m, n+1)$ in Figure 13(a). However, in our approach, the matching point can be anywhere on the two sides connecting the three points (see Figure 13(b)). In this way, the difference of samplings will not increase the distance between similar trajectories and the difference of trajectory lengths will not affect the matching either.

There are two warping conditions in every step of our algorithm, no matter which side the warping path go through in the previous step of matching and into the present “matching block” (see Figure 13(b)). As long as the warping path enters the block, it can only exit from the left side or bottom side, including the intersection of these sides. All these conditions are included in (4.1). When $\theta = 1$, this is the same condition as that in DTW algorithm.

In the warping algorithm, if the parameter θ is not zero, the corresponding point in one of the trajectories must between the adjacent samples and the position of the point is unknown. In the CDTW algorithm [37], the positions in x and y direction are calculated by linear interpolation separately. However, the linear interpolation is not accurate especially in motion trajectories, because not only the two adjacent samples decide the position between these samples but also the neighboring samples of them will affect the position of the unknown point as well.

As presented in the efficient prediction method Kalman Filtering [38], the prediction of the unknown point $x(m+\theta)$ depends and only depends on the present sample $x(m)$ and the previous sample $x(m-1)$. As the whole trajectory sample data are known in advance, the succeeding samples $x(m+1)$ and $x(m+2)$ are also useful in calculating the unknown point (see Figure 14). We can also use this theory by the feature of motion that the previous sample $x(m-1)$ will control the inertia of the unknown point by direction and speed as well as that the same property of the unknown point will also affect the succeeding sample $x(m+2)$.

In our method, the *cubic polynomials interpolation* is selected to calculate the coordinate of the unknown point with four samples: $x(m-1)$, $x(m)$, $x(m+1)$, and $x(m+2)$, because four samples can control a cubic curve. Then, we use the calculated coordinates and neighboring known samples to calculate the invariants of the point for the calculation of the matching distance. The two corresponding points $A(m)$ and $B(n)$ in trajectories A and B

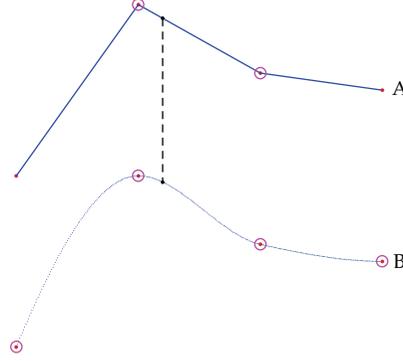


Figure 14: The Samples used in calculating the unknown points (linked by dashed) are denoted by circles. Linear interpolation uses only 2 points in A, while cubic polynomials interpolation uses 4 points in B.

(maybe $A(m + \theta)$ or $B(n + \theta)$) are represented by their mixed signature: $[k^m, k_s^m, \tau^m, \tau_s^m, s^m, r^m]$ and $[k^n, k_s^n, \tau^n, \tau_s^n, s^n, r^n]$. The distance between samples $A(m + \theta)$ and $B(n)$ is defined as follows:

$$d(m + \theta, n) = \min_{\theta} \frac{\Delta k^{*m+\theta, n} \cdot \Delta \tau^{*m+\theta, n}}{\sqrt{(s^{*m+\theta})^2} \cdot \sqrt{(s^{*n})^2}} \cdot (1 + \lambda \cdot \Delta h^{*m+\theta, n}), \quad (4.2)$$

where

$$\begin{aligned} \Delta k^{*m+\theta, n} &= \left\| (k^{*m+\theta}, k_s^{*m+\theta}) - (k^{*n}, k_s^{*n}) \right\|, \\ \Delta \tau^{*m+\theta, n} &= \left\| (\tau^{*m+\theta}, \tau_s^{*m+\theta}) - (\tau^{*n}, \tau_s^{*n}) \right\|, \\ (S^{*m+\theta})^2 &= (k^{*m+\theta})^2 + (k_s^{*m+\theta})^2 + (\tau^{*m+\theta})^2 + (\tau_s^{*m+\theta})^2, \\ (S^{*n})^2 &= (k^{*n})^2 + (k_s^{*n})^2 + (\tau^{*n})^2 + (\tau_s^{*n})^2, \\ \Delta h^{*m+\theta, n} &= \|(s^*(m + \theta), r^*(m + \theta)) - (s^*(n), r^*(n))\|. \end{aligned} \quad (4.3)$$

The definition of $d(m, n + \theta)$ is similar to that of $d(m + \theta, n)$.

The parameter λ is the same as the one used in motion perception via DTW. Trajectories in different circumstances will appear different in global distance and noise. Overload the global distance h will enlarge error in calculation, and lead to wrongly recognizing. For a database sampled under certain circumstance, we calculate the average distance with different λ . We set λ as zero and enlarge it with a certain step size in iterative calculation of average distance $D(i)$ until arrive the convergence condition. The subscript i here is the iteration index, and the convergence condition is a threshold of average distance: $D(i + 1) - D(i) < E$ (e.g., $E = 1e - 5$).

In another condition, a set of motion trajectories are classified with a standard beforehand, and we want the recognition engine to classify trajectories by this standard. Therefore, the weight λ should be trained by this standard database first. We adjust λ in

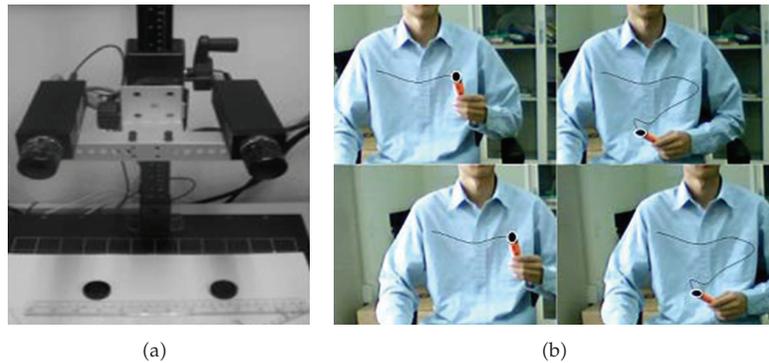


Figure 15: Stereo vision system setup and snapshots of the stereo tracking trajectory.

different value to classify this database, until the result of classification of this database is the same as the standard. Then, this λ satisfies this task.

5. Experiments

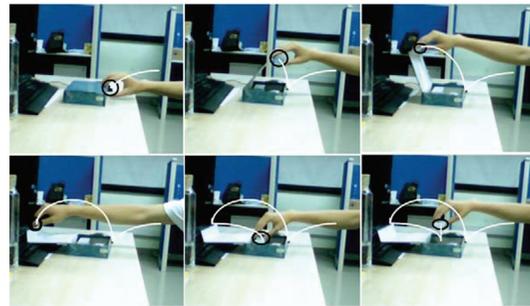
The aim of the experiments is to present the performance of the mixed signature in motion perception and motion recognition comparing with the previous methods. Motion perception and recognition are demonstrated based on the DTW and modified CDTW nonlinear matching algorithms respectively. We use the DTW algorithm in motion perception to show the flexibility of the mixed signature descriptor. The modified CDTW is used in motion recognition to improve the recognition accuracy.

Sign motion is an important sort of human action used for daily interaction. The signs are spatial symbols performed by human hands or other mode. We did implementations with several groups of sign data in order to illuminate the necessity of the global invariants in motion analysis. We used a stereo vision system to track sign motion trajectories of different people and recorded them in a PC (Figure 15). The 3D motion trajectory was calculated from the two image sequences captured by separate camera. We used these sign trajectories to test the properties of the mixed signature in motion perception and recognition compared with the previous method. But note that this descriptor is reliable in various tasks other than this given type of examples. A large trajectory database [34] was used in [31–33] for motion recognition, and we also used this database in our experiments for comparison.

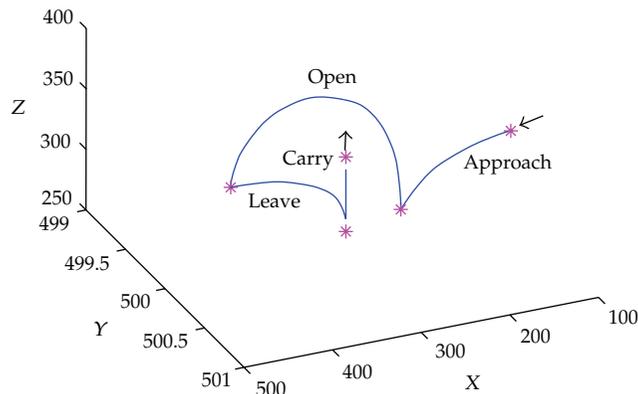
Beside the sign motion, daily behavior of people was tracked as well in the experiments. We recorded these motion trajectories for motion analysis, such as open a box, pure water, and other actions. In the trajectory acquisition, only the actions performed by one hand were tracked and we tracked the mark on hand instead of directly tracking the hand. The tracking was simplified by tracking a rigid object, and tracking other parts of body is out of the scope of this paper. Figure 16 shows the tracking of opening a box to carry an object.

5.1. Invariant Property in Motion Analysis

The mixed signature is a flexible descriptor in both local shapes and distribution as a whole for motion analysis. As an invariant signature, it is also spatial-temporal invariant for motion



(a)



(b)

Figure 16: (a) The action trajectory is shown in white with several snapshots. (b) The extracted trajectory data in 3D space.

trajectories in 3D space as the previous signature, including translation, rotation, scaling, speed and occlusion. This experiment demonstrates the intertrajectory perception between motion trajectories from the mixed signature based on the DTW alignment algorithm. Figures 17–20 show several cases of trajectory path matching, which give an intuitive perception of the invariant properties.

From the instance in Figure 17(a), the trajectories of the same action in different positions were tracked. Both the viewpoint and motion position were translated from one position to another position. Figure 17(b) is the interalignment matching between the trajectories which illustrate the invariant of the mixed signature under translation. It is observed that the corresponding points of respective trajectory are matched via DTW algorithm in the figure. We should notice that the bottle in A-1 and B-1 are the same one; hence, the trajectories A-1 and B-1 are the same in size. The trajectories in Figure 18(a) represents the same action rotated to different directions and were tracked in different viewpoints. The abstracted trajectories A-2 and B-2 are matched in Figure 18(b) which demonstrates the invariant property under rotation as well. Figure 19 is the case that the similar signs in different sizes which are matched via DTW. We can observe that A-3 and B-3 are similar in shape with different sizes in the same coordinate system. This difference is not resulted by the distance from the vision system to the object, but the real difference in scale. This case verified the scaling invariant of the mixed signature. In general motion

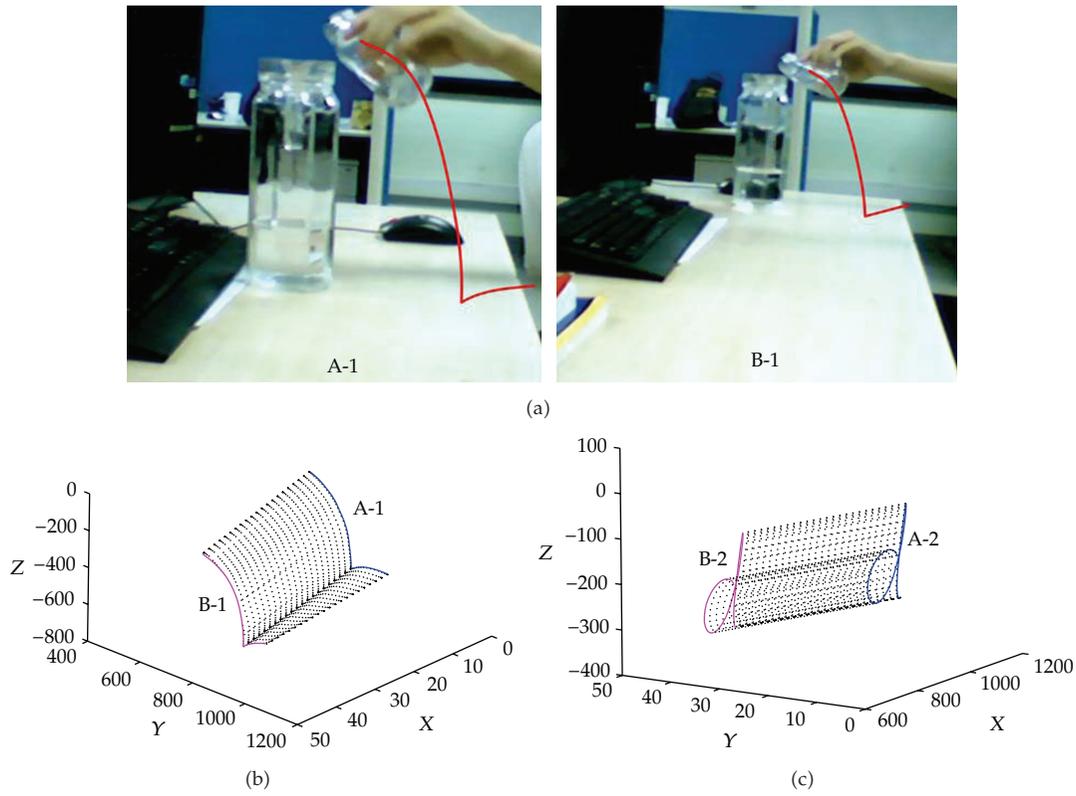


Figure 17: (a) The same action translated to another place. (b) The matching of the motion trajectories in the same coordinate system. (c) Perception of translation invariant in sign motions.

instances, we should notice that there may be not only one spatial transformation between two similar actions. The transformation between two motion trajectories is probably a mixture of translation, rotation, and scaling. For example, there is also translation between the actions under rotation in Figure 18, which shows the invariant for complex transformation.

The instance in Figure 20 is two trajectories of similar signs in different speed. We tracked these signs in the same frame rate and the sampling of B-5 is much denser than that of A-5, which shows the speed of A-5 is faster than that of B-5. In this way, the samples in A-5 are warping to more than one corresponding samples in B-5 via alignment, which demonstrates the invariant in different speed. The occlusion in tracking of motion is another important factor in motion analysis and the mixed signature is invariant for this as well. However, as the global invariants are calculated with all samples of a trajectory, the mixed signature cannot be directly used in this condition. There should be preprocessing of motion trajectory under the occlusion condition. Furthermore, a high level alignment method of trajectories should be introduced instead of DTW/CDTW, and this is out of the scope of this paper.

5.2. Perception of Motion Symmetry

In this experiment, we generated a group of sign motion trajectories for motion symmetry perception via the mixed signature compared with the previous signature. Symmetrical

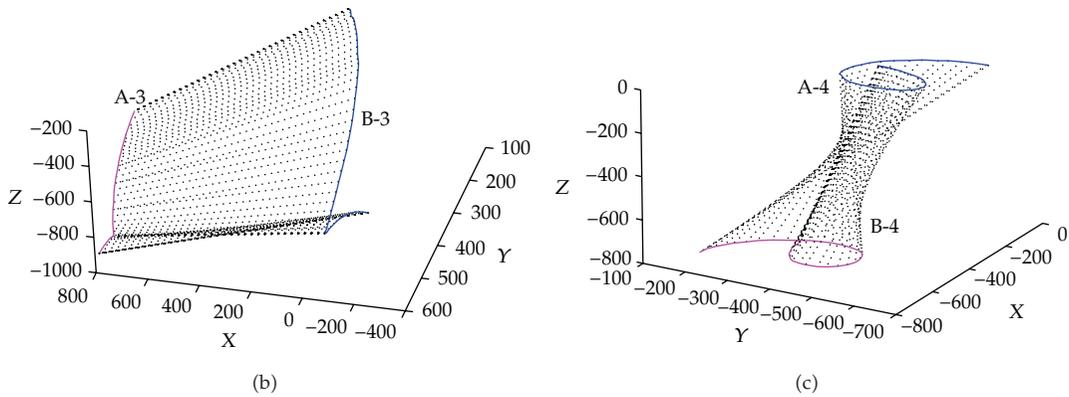
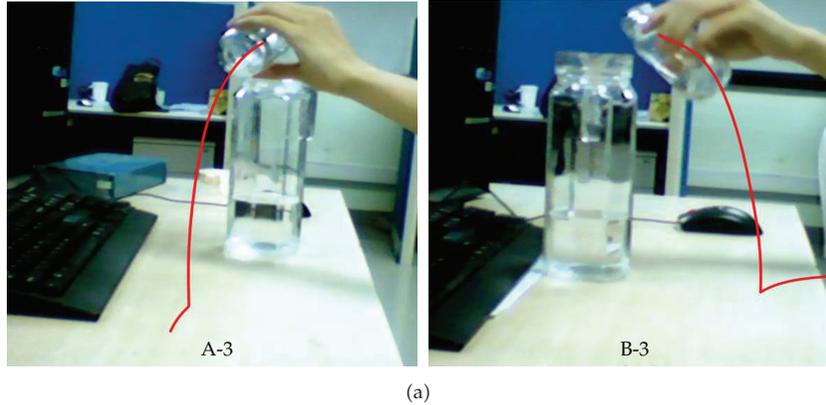


Figure 18: (a) The same action rotated to different direction. (b) The matching of the motion trajectories in the same coordinate system. (c) Perception of rotation invariant in sign motions.

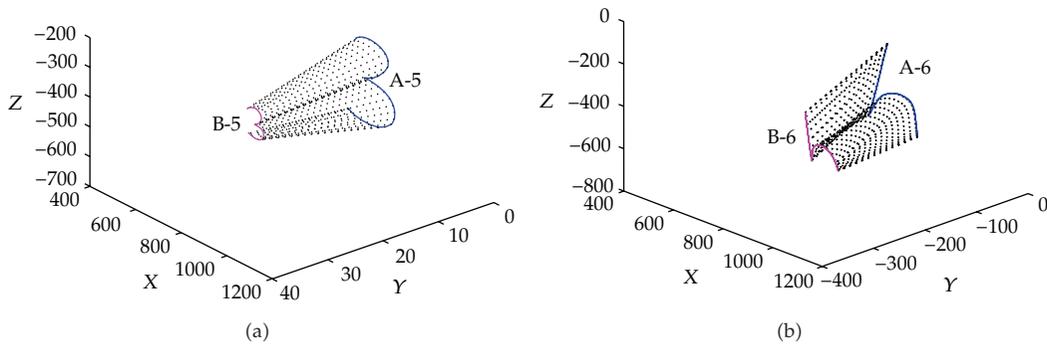


Figure 19: Matching of trajectories in different sizes.

segments of a trajectory were perceived excluding the dissymmetrical points. There might be more than two pairs of symmetrical segments in a trajectory, and some of them maybe overlap each other. These conditions can be perceived in our method as shown in Figure 21. There is one pair of symmetrical segments in Figure 21(a), and the central point is noted with a black star. Those points which cannot be matched by Theorem 3.1 will not be perceived.

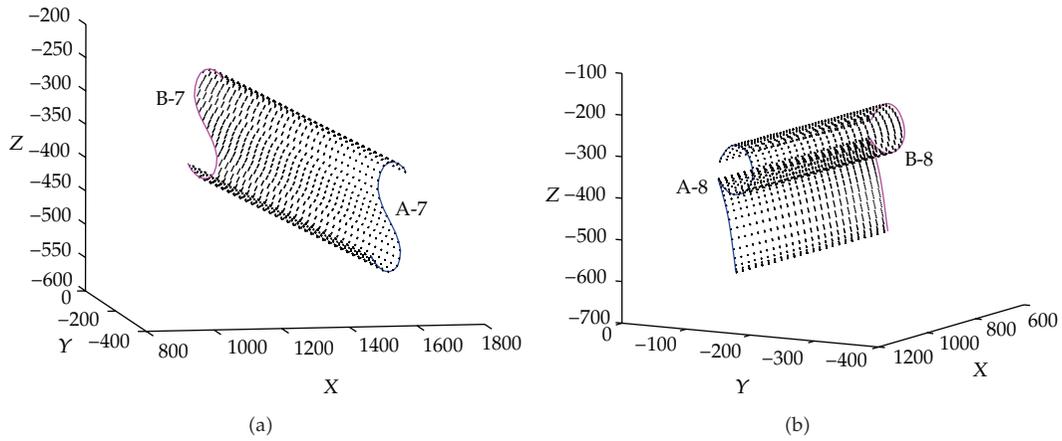


Figure 20: Matching of trajectories in different speed.

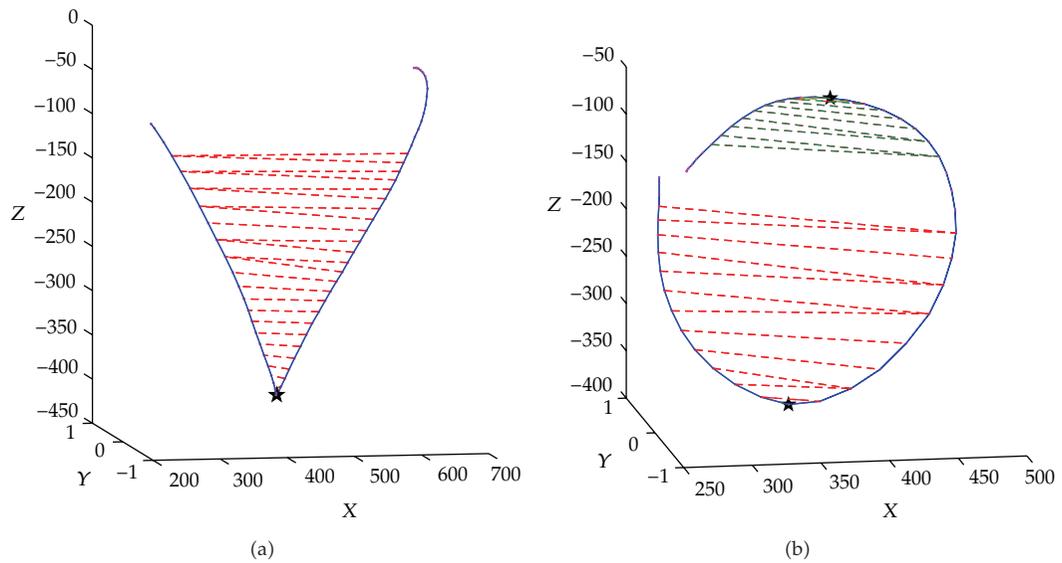


Figure 21: Mirror symmetry of trajectories are perceived by mixed signature. Corresponding points are matched by DTW according to Theorem 3.1.

Similarly, both of the two pairs of the symmetrical segments in Figure 21(b) can be perceived with respective central point. The cases in Figure 21 are of mirror symmetry, and a case of central symmetry is shown in Figure 22.

As the previous signature is invariant for scaling, two segments of a trajectory in different scale will have the same signature if they are similar in local shape. That is to say, if a part of symmetry motion is scaled, this motion will be error perceived as symmetry motion as well. If the global information is considered, this error will never occur. The mixed signature is also invariant in space transform, but not invariant for different parts within

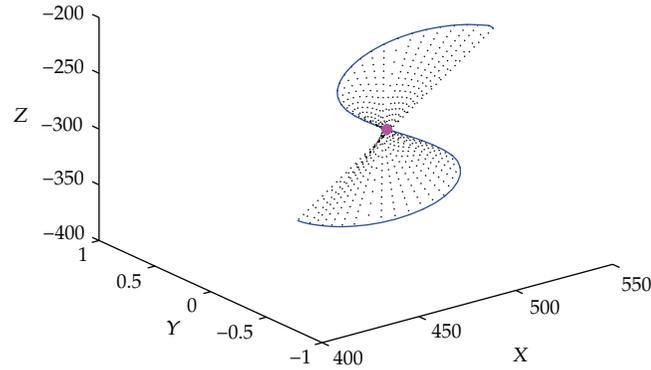


Figure 22: Central symmetry of sign trajectory is perceived by mixed signature. Corresponding points are matched by DTW according to Theorem 3.2.

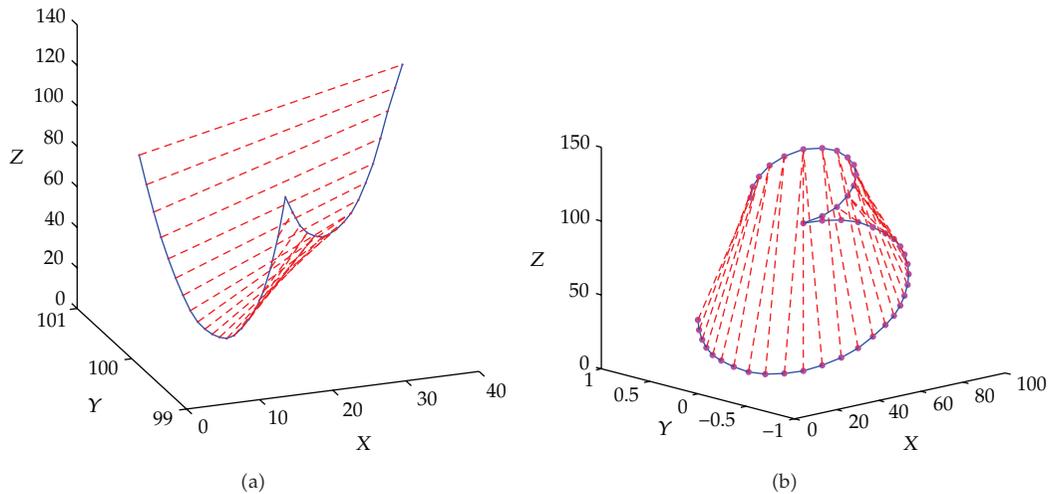


Figure 23: Dissymmetry trajectories are error perceived by differential invariants via DTW matching algorithm.

a single trajectory. Some words signed in this condition were error perceived by separate trajectories. Figure 23 illustrates this problem.

From Figure 23, we can see that two segments of the trajectories are similar in shape but different in scale. As scaling will not change the differential invariants, the corresponding points satisfy the theorem of symmetry in the previous signature. In this way, these signs were error perceived as symmetrical trajectories by that method [31]. However, the global invariants of the corresponding points are different at all. For example, the arc lengths from the central point to the corresponding points are different. Hence, these signs will not be perceived as symmetrical trajectories by the mixed signature. The theorem of mixed signature is more accurate and strict in symmetrical motion perception than the previous one. The same condition will also occur in the central symmetry perception.

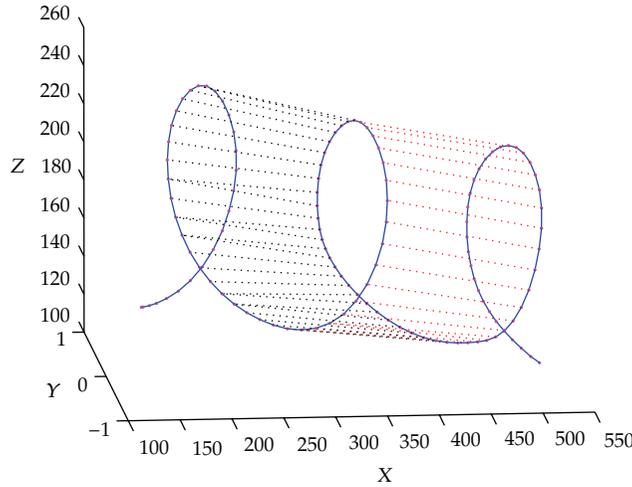


Figure 24: Perception of periodic motion trajectory.

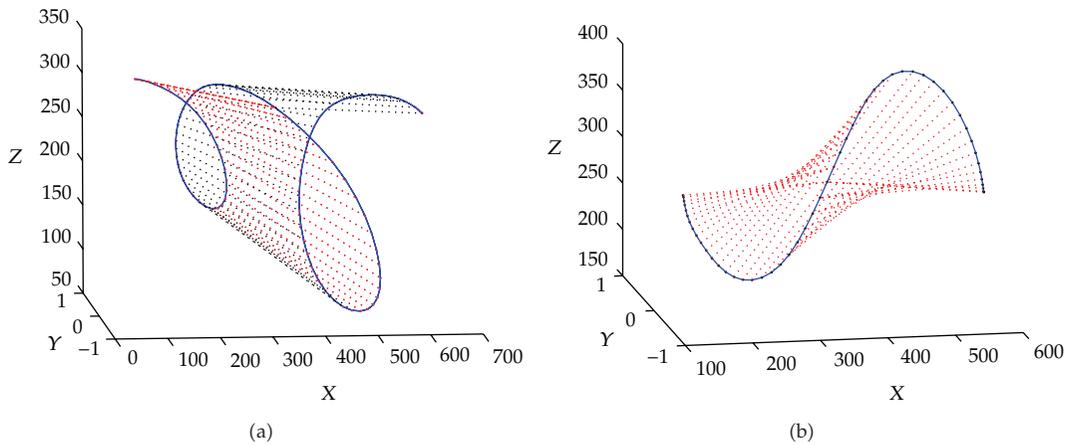


Figure 25: Error perception of periodic motion trajectories.

5.3. Perception of Periodic Motion

We sampled several groups of periodic motions by different people and perceived the periodic properties of them. Figure 24 shows a case of periodic trajectory with 3 periods. Some of the corresponding points are linked with dash lines.

The similar condition also occurs in the perception of periodic motion by the previous signature as occurred in symmetrical motion perception. Periodic motion perception suffers much more from the periodic theorem of differential invariants than that in symmetrical motion perception. Not only the scaling invariant will lead to error perception while a trajectory is not of periodic property, but also the rotation invariant will cause this error perception as well. If a periodic segment of a trajectory is scaled or rotating to another form, this trajectory will be error perceived by the previous signature as shown in Figure 25.

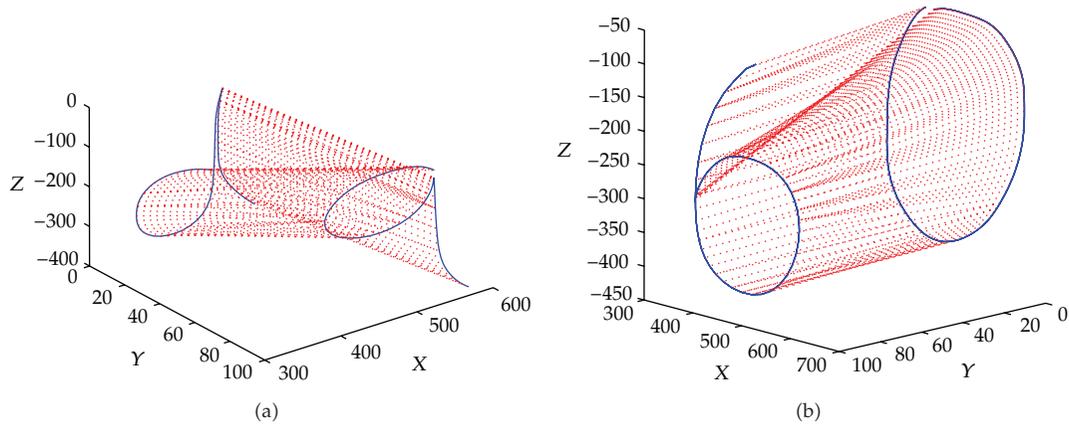


Figure 26: Matching of similar local shapes by the previous signature. Corresponding points were linked by dash lines.

A trajectory with two similar segments in different scale in Figure 25(a) was erroneously perceived as periodic trajectory by the previous signature. Another one in Figure 25(b) was also erroneously perceived while two segments rotate to different directions. However, the global parameters Δs and Δr are different between the corresponding points. In this way, these trajectories would not be error perceived as periodic motions by the mixed signature.

5.4. Sign Motion Recognition

We implemented sign motion recognition via the mixed signature comparing with the previous signature in order to test the performance of our method. We captured sign motion trajectories from the vision system in this experiment to test the characteristics of our method for motion recognition, and we performed the statistics implementation with a large 3D database in next subsection.

Two groups of sign motion trajectories in different classes were sampled by several different signers. There are similar shapes between different classes of these signs and hard to distinguish. Every two signs in different classes in these groups are similar in their local shapes. As the differential invariants are similar between the corresponding points in these similar local shapes, these signs will be wrongly classified by the previous signature descriptor. Figure 26 shows this condition with two cases: $d-q$ and $0-6$. We can see the similar local shapes were matched by the local features in separate figure. Hence, their distances were largely decreased.

We matched all these signs of different classes by previous descriptor and new descriptor separately. The results in comparison are listed in Table 1. The data in this table are distances by separate descriptor, and 100 pairs of sign data were tested for every subgroup. We calculated the average values and the extreme values from the 100 data and listed them in the table. From these data we can see that the matching of $d-q$ by the previous signature is not clearly different from $d-d$ and $q-q$, with even some of the extreme values overlapped. In this way, d and q cannot be classified accurately by the previous descriptor and will be wrongly accepted or rejected for the boundary confusion. In contrast, we can see that the results of

Table 1: Sign matching of “*d*” and “*q*”.

Match group	Previous signature			Mixed signature		
	Average	Min	Max	Average	Min	Max
<i>d-d</i>	22.3	20.5	24.8	28.2	21.0	33.8
<i>q-q</i>	20.1	17.7	23.6	17.8	14.6	21.6
<i>d-q</i>	26.3	23.8	29.4	94.7	71.3	127

Table 2: Sign matching of “0” and “6”.

Match group	Previous signature			Mixed signature		
	Average	Min	Max	Average	Min	Max
0-0	24.1	21.6	26.8	32.3	28.1	34.7
6-6	24.8	21.4	27.1	24.4	18.5	29.9
0-6	26.6	23.0	31.7	69.2	57.6	97.9

Table 3: Sign matching of two types of “4”.

Match group	Previous signature			Mixed signature		
	Average	Min	Max	Average	Min	Max
Type A-A	12.5	11.0	14.2	23.3	21.1	24.6
Type B-B	11.5	10.2	12.3	21.9	18.1	25.3
Type A-B	14.7	12.6	18.9	32.6	25.0	37.4

d-q are obviously different from *d-d* and *q-q* under the mixed signature descriptor and the boundaries are distinguished. The same condition occurs in the matching between 0 and 6 as well. Table 2 lists the experimental results of this group. The experiments in [33] also suffer the confusion of 0 and 6 which is presented in that paper.

We also tested the same word by different fonts (just like the Experiment 5.1–5.3 in [31], see Figure 27). Recognition with the previous descriptor can only classify the words in different classes but cannot distinguish the same word in different fonts. In some cases, similar words in different classes cannot be correctly classified either. However, the mixed signature can solve these confusions in our experiments. We tested 4 and 9 which were signed in different fonts, just like Figure 27. The results of classifying the two fonts of number 4 in Figure 27(a) (Type A and Type B) are listed in Table 3.

5.5. Motion Recognition with Large Database

In this experiment, we tested the mixed signature descriptor by recognizing 3D trajectories of different classes in a large database [34]. The database was used in [31–33] for experiments and we used the same database for comparison. Two instances of the sign words “crazy” and “name” are shown in Figure 28. There are 95 classes in the database and 29 samples for each class. We used half of the samples for training and the other half for testing. Several classes of samples were randomly selected and recognized in our method, and we repeated this test 50 times. The average ratios of correct recognition are listed in Table 4 for different number of classes.

The experimental results in Table 4 show that the new method by mixed signature achieves higher recognition rate in matching within 2, 4, and 8 classes separately.

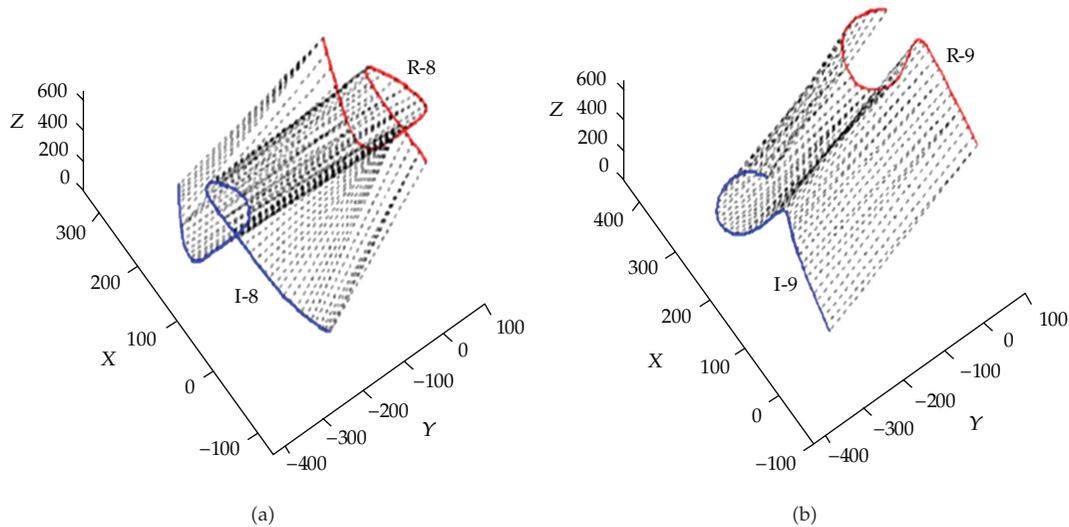


Figure 27: Matching of the same word signed in different fonts referred from Figure 16 in [31].

Table 4: Average ratio of correct recognition (%).

Number of classes	2	4	8
Previous signature	92.03	87.52	81.19
Mixed signature	94.30	90.15	86.71
Fourier descriptor	87.98	75.74	63.85

Data of the previous signature in this table is referred the experimental results of [31].

Table 5: Average ratio of correct recognition (%).

Number of classes	2	4
Previous descriptor	32.06	18.92
Mixed signature	54.51	31.14

Furthermore, the correct ratio of mixed signature in two classes is 2.27% more than that of the previous method. However, this difference of correct ratio in four and eight classes recognition is enlarged to 2.63% and 5.52%, respectively. Hence, as the number of classes increases, the recognition rate of the new method outperforms much more than the previous one. That is to say, our method is more flexible for multiclass recognition in large database.

We also used another dataset from the same database [34] including 95 classes and each class has 70 samples signed by five people. Two and Four people were selected randomly every time and half of their signatures were used for training and half for testing. This experiment was repeated over 100 times in all these 95 classes, and the average recognition performance of these tests is listed in Table 5. Figure 29 shows the confusion degree among the motions from these five signers with the intensity image. The grey levels of the intensity image denote the degree of confusion which is the ratio of error classification between two classes.

From Table 5, we can see that the recognition rate of our method is higher than the previous one. Some fonts between signers are hard to distinguish that these fonts of the two

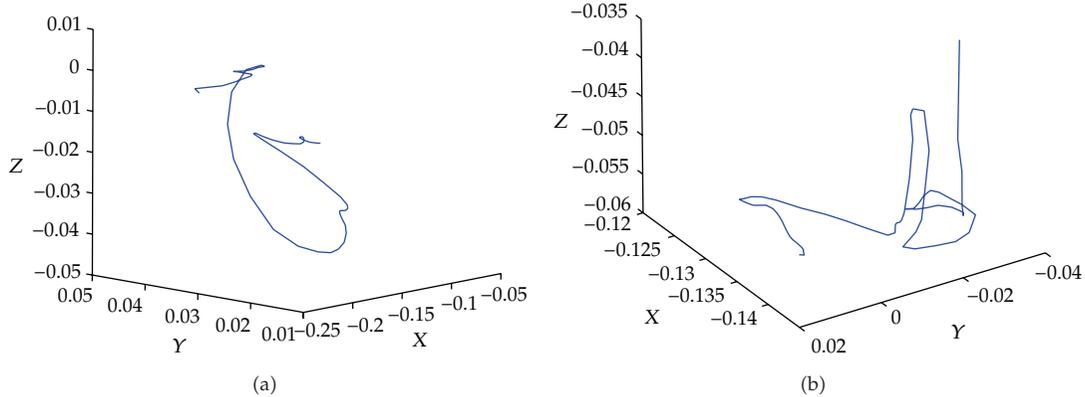


Figure 28: Sign samples of the words “crazy” in (a) and “name” in (b).

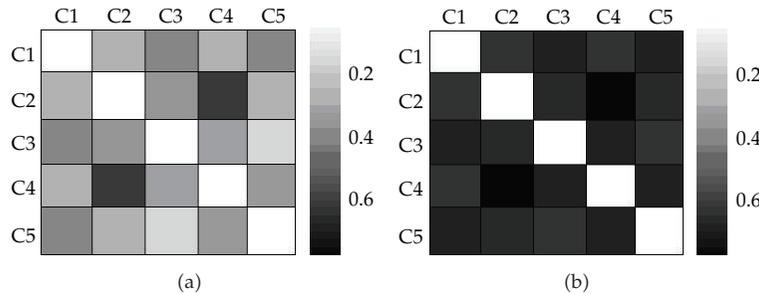


Figure 29: Confusion degree among 5 people under our new method in (a) and the previous method in (b).

people are very similar. For example, the recognition between C2 and C4 are hard to classify by either method (see Figure 29). For this reason, the average recognition rate is much lower than classification among different words. However, the result can also support our method which has performance in this aspect. These comparisons show that the proposed method of mixed signature with global invariants is more capable for motion trajectory recognition under various circumstances.

6. Conclusions

A new invariant descriptor—mixed signature is presented for 3D motion perception and recognition via the trajectory. This new descriptor is based on differential invariants but uses extra parameters containing global information which was not included in the previous study. An effective alignment algorithm CDTW is modified and used in our method for trajectory matching. We modified the CDTW algorithm for matching trajectories with the *cubic polynomials interpolation*. Our new method is flexible and adaptable for different tasks by adjustable λ . Experimental results show the advantage of this method.

We compare the performance of classifying different classes of trajectories by our new descriptor with the performance of the previous descriptor. Our method shows better performance especially in distinguishing motions with similar shapes. We also test these

methods by trajectories which were signed by different persons, and our method outperforms previous methods. To increase the computational efficiency of the CDTW algorithm for high speed implementation, some efficient methods need to be developed for computing the invariants. Furthermore, we will apply this method for motion analysis in the biology and human health area [39].

Acknowledgment

This work was supported by a grant from City University of Hong Kong (Project no. 7002511).

References

- [1] A. Psarrou, S. Gong, and M. Walter, "Recognition of human gestures and behaviour based on motion trajectories," *Image and Vision Computing*, vol. 20, no. 5-6, pp. 349–358, 2002.
- [2] A. Billard and R. Siegwart, "Robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 47, no. 2-3, pp. 65–67, 2004.
- [3] K. Vloker, L. H. Dennis, B. Sanmohan, U. Ales, and K. Danica, "Learning actions from observations," *IEEE Robotics and Automation Magazine*, vol. 17, no. 2, pp. 30–43, 2010.
- [4] K. K. Lee, M. Yu, and Y. Xu, "Modeling of human walking trajectories for surveillance," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1554–1559, 2003.
- [5] J. Martin, D. Hall, and J. L. Crowley, "Statistical gesture recognition through modeling of parameter trajectories," *Lecture Notes in Computer Science*, vol. 1739, pp. 129–140, 1999.
- [6] E. Bribiesca, "A chain code for representing 3D curves," *Pattern Recognition*, vol. 33, no. 5, pp. 755–765, 2000.
- [7] S. Berretti, A. D. Bimbo, and P. Pala, "Retrieval by shape similarity with perceptual distance and effective indexing," *IEEE Transactions on Multimedia*, vol. 2, no. 4, pp. 225–239, 2000.
- [8] A. Masood, "Optimized polygonal approximation by dominant point deletion," *Pattern Recognition*, vol. 41, no. 1, pp. 227–239, 2008.
- [9] V. V. Kindratenko, "On using functions to describe the shape," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 33, pp. 225–245, 2003.
- [10] M. J. Black and A. D. Jepson, "A probabilistic framework for matching temporal trajectories: condensation-based recognition of gestures and expressions," in *Proceedings of the European Conference on Computer Vision*, vol. 1, pp. 909–924, Freiburg, Germany, 1998.
- [11] F. S. Cohen, Z. Huang, and Z. Yang, "Invariant matching and identification of curves using B-splines curve representation," *IEEE Transactions on Image Processing*, vol. 4, no. 1, pp. 1–10, 1995.
- [12] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 480–487, 2011.
- [13] M. C. Shin, L. V. Tsap, and D. B. Goldgof, "Gesture recognition using Bezier curves for visualization navigation from registered 3-D data," *Pattern Recognition*, vol. 37, no. 5, pp. 1011–1024, 2004.
- [14] P. R. G. Harding and T. J. Ellis, "Recognizing hand gesture using Fourier descriptors," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, pp. 286–289, August 2004.
- [15] G. C. H. Chuang and C. C. J. Kuo, "Wavelet descriptor of planar curves: theory and applications," *IEEE Transactions on Image Processing*, vol. 5, no. 1, pp. 56–70, 1996.
- [16] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, vol. 2010, Article ID 408418, 22 pages, 2010.
- [17] S. Tabbone, L. Wendling, and J. P. Salmon, "A new shape descriptor defined on the Radon transform," *Computer Vision and Image Understanding*, vol. 102, no. 1, pp. 42–51, 2006.
- [18] I. Weiss, "Geometric invariants and object recognition," *International Journal of Computer Vision*, vol. 10, no. 3, pp. 207–231, 1993.
- [19] D. Forsyth, J. L. Mundy, A. Zisserman, and C. M. Brown, "Projectively invariant representations using implicit algebraic curves," *Image and Vision Computing*, vol. 9, no. 2, pp. 130–136, 1991.

- [20] S. Manay, D. Cremers, B. W. Hong, A. J. Yezzi, and S. Soatto, "Integral invariants for shape matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1602–1617, 2006.
- [21] S. Y. Chen, J. Zhang, Q. Guan, and S. Liu, "Detection and amendment of shape distortions based on moment invariants for active shape models," *IET Image Processing*, vol. 5, no. 3, pp. 273–285, 2011.
- [22] Y. Caspi, D. Simakov, and M. Irani, "Feature-based sequence-to-sequence matching," *International Journal of Computer Vision*, vol. 68, no. 1, pp. 53–64, 2006.
- [23] 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.
- [24] M. Scalia, G. Mattioli, and C. Cattani, "Analysis of large-amplitude pulses in short time intervals: application to neuron interactions," *Mathematical Problems in Engineering*, vol. 2010, Article ID 895785, 15 pages, 2010.
- [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, p. 21, 2011.
- [26] K. Ogawara, Y. Tanabe, R. Kurazume, and T. Hasegawa, "Detecting repeated motion patterns via dynamic programming using motion density," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1743–1749, 2009.
- [27] A. Prati, S. Calderara, and R. Cucchiara, "Using circular statistics for trajectory shape analysis," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, June 2008.
- [28] D. R. Faria and J. Dias, "3D hand trajectory segmentation by curvatures and hand orientation for classification through a probabilistic approach," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '09)*, pp. 1284–1289, December 2009.
- [29] E. Ribnick and N. Papanikolopoulos, "View-invariant analysis of periodic motion," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '09)*, pp. 1903–1908, October 2009.
- [30] J. Sun, X. Wu, S. Yan, L. -F. Cheong, T. -S. Chua, and J. Li, "Hierarchical spatio-temporal context modeling for action recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 2004–2011, 2009.
- [31] S. D. Wu and Y. F. Li, "Flexible signature descriptions for adaptive motion trajectory representation, perception and recognition," *Pattern Recognition*, vol. 42, no. 1, pp. 194–214, 2009.
- [32] J. Y. Yang, Y. F. Li, and K. Y. Wang, "Mixed signature descriptor with global invariants for 3D motion trajectory perception and recognition," in *Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 1952–1956, 2010.
- [33] S. Wu and Y. F. Li, "On signature invariants for effective motion trajectory recognition," *International Journal of Robotics Research*, vol. 27, no. 8, pp. 895–917, 2008.
- [34] UCI KDD ASL Archive, <http://kdd.ics.uci.edu/databases/auslan2/auslan.html>.
- [35] E. Calabi, P. J. Olver, C. Shakiban, A. Tannenbaum, and S. Haker, "Differential and numerically invariant signature curves applied to object recognition," *International Journal of Computer Vision*, vol. 26, no. 2, pp. 107–135, 1998.
- [36] L. R. Rabiner and B. H. Huang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
- [37] M. Munich and P. Perona, "Continuous dynamic time warping for translation-invariant curve alignment with applications to signature verification," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 1, pp. 108–115, 1999.
- [38] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transaction of ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.
- [39] S. Chen, J. Zhang, H. Zhang et al., "Myocardial motion analysis for determination of tei-index of human heart," *Sensors*, vol. 10, no. 12, pp. 11428–11439, 2010.

Research Article

Radial Basis Functional Link Network and Hamilton Jacobi Issacs for Force/Position Control in Robotic Manipulation

Shuhuan Wen,¹ Junying Yuan,² and Jinghai Zhu²

¹ *Institute of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China*

² *Key Lab of Industrial Computer Control Engineering of Hebei Province, Yanshan University, Qinhuangdao 066004, China*

Correspondence should be addressed to Shuhuan Wen, wenshuhuan@sohu.com

Received 14 April 2011; Revised 18 May 2011; Accepted 29 May 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Shuhuan Wen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper works on hybrid force/position control in robotic manipulation and proposes an improved radial basis functional (RBF) neural network, which is a robust relying on the Hamilton Jacobi Issacs principle of the force control loop. The method compensates uncertainties in a robot system by using the property of RBF neural network. The error approximation of neural network is regarded as an external interference of the system, and it is eliminated by the robust control method. Since the conventionally fixed structure of RBF network is not optimal, resource allocating network (RAN) is proposed in this paper to adjust the network structure in time and avoid the underfit. Finally the advantage of system stability and transient performance is demonstrated by the numerical simulations.

1. Introduction

During robotic operation, the end-effectors may perform tactile contact with the environment, which consists of a force interaction between the end-effector and the environment. In addition to robot's position control, the force control is more necessary in order to fulfill its tasks better. Raibert and Craig [1] firstly introduce such an idea in 1981. After then many other researchers have proposed and explored new hybrid control strategies, for example, by combination with visual information [2–4].

Due to uncertainties of the robot model, the system's performance becomes greatly weaken or even unstable, so robust control methods for robots are widely concerned. By using a fixed controller structure, the method has the advantage of eliminating the impact of the uncertainty, ensuring the stability of the system during its operation.

The main assumption of the method is the fact that only the upper bound of uncertainty is known. However, the upper bound is difficult to be measured which is the limitation of the robust control method. To overpass this limitation, the radial basis functional (RBF) neural networks (RBFNNs) approximate the function to compensate for the lack of robust control. RBFNN has a compact topology structure and rapidly convergence, and its structural parameters can be learned separately [5–8]. Because of its fixed or a more complex structure, RBF will lead to the result that learning time is too long or wasting network resources. Therefore, we use resource allocation network (RAN) in this paper. The RAN method replaces the sampling point by the biggest error sampling point and by doing so, the network can perform self-learning and its complexity is reduced [9, 10]. Radial Basis Functional Link Network (RBFLN) increases the weight from input to output; therefore, RBFNN not only includes RBF advantage, but also compensates for the slow response of RBF.

2. Manipulator Dynamics

The dynamic equation of the n-link manipulator in joint-space coordinates is given by

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) = \tau_p + \tau_f + J^T f + \omega(q, \dot{q}, t), \quad (2.1)$$

where the vector $q \in R^n$ is the joint angle, the vector $\dot{q} \in R^n$ denotes the joint angular velocity, the vector $\ddot{q} \in R^n$ is the joint angular acceleration, $M(q) \in R^{n \times n}$ is the symmetric positive definite inertia matrix, $C(q, \dot{q})\dot{q} \in R^n$ denotes the vector of Coriolis and centrifugal forces, $G(q) \in R^n$ denotes the gravitational vector, τ_p is the vector of joint actuator torques in position control loop, τ_f is the vector of joint actuator torques in force control loop, $f \in R^n$ is the force between the end-effector and the environment, $J \in R^{n \times n}$ denotes the Jacobian matrix, and $\omega(q, \dot{q}, t)$ represents the vector of external disturbance joint torques and unmodeled dynamics.

In the position loop, the simplest PD controller can be expressed as

$$\tau_p = -K_p e_p - K_v \dot{e}_p, \quad (2.2)$$

where K_p, K_v are the constant matrixes,

$$e_p = q - q_d, \quad \dot{e}_p = \dot{q} - \dot{q}_d. \quad (2.3)$$

In the force loop, the dynamic equation should be transferred from joint-space to Cartesian-space [1, 7]. Based on $\dot{r} = J\dot{q}$, $\ddot{r} = \dot{J}\dot{q} + J\ddot{q}$, (2.1) can be derived as follows:

$$M_r \ddot{r} + C_r \dot{r} + G_r = U + \omega_r + f, \quad (2.4)$$

where

$$\begin{aligned} M_r &= J^{-T} M(q) J^{-1}, & C_r &= J^{-T} (C(q, \dot{q}) - M(q) J^{-1} \dot{J}) J^{-1}, \\ G_r &= J^{-T} G(q), & U &= J^{-T} \tau_f, & \omega_r &= J^{-T} \omega(q, \dot{q}, t) \end{aligned} \quad (2.5)$$

Equation (2.4) which is showed by the Cartesian coordinate has the following important quality.

Assume that r_d is the desired trajectory and f_d is the desired force. The force between the end-effector and the environment is given by the following expression [1, 8]:

$$f = G_e(r - r_e), \quad e_f = f - f_d, \quad (2.6)$$

where G_e is the environment stiffness, r_e is reference position of environment, so $\dot{r} = G_e^{-1}(\dot{e}_f + \dot{f}_d)$, $\ddot{r} = G_e^{-1}(\ddot{e}_f + \ddot{f}_d)$.

We can obtain the error equation as follows:

$$M_r G_e^{-1} \ddot{e}_f + C_r G_e^{-1} \dot{e}_f - e_f + \Delta = U + \omega_r, \quad (2.7)$$

where $\Delta = M_r G_e^{-1} \ddot{f}_d + C_r G_e^{-1} \dot{f}_d + G_r - f_d$. State variables can be defined as $x_1 = e_f$, $x_2 = \dot{e}_f + \alpha e_f$, where α is a given positive number, (2.7) can be derived as

$$\begin{aligned} \dot{x}_1 &= x_2 - \alpha x_1, \\ M_r G_e^{-1} \dot{x}_2 &= -C_r G_e^{-1} x_2 + x_1 - \Delta + U + \omega_r + \omega, \end{aligned} \quad (2.8)$$

where $\omega = M_r G_e^{-1} \alpha \dot{e}_f + C_r G_e^{-1} \alpha e_f$.

3. Design of Control Law

In order to obtain the control law, we introduce a theorem in this section. Assume that there is a system with disturbance as follows:

$$\begin{aligned} \dot{x} &= f(x) + g(x)d, \\ z &= h(x), \end{aligned} \quad (3.1)$$

where d is the disturbance and z is the signal of evaluation.

For the force control loop, the operation space of robot is transformed. Because $\dot{r} = J\dot{q}$, $\ddot{r} = \dot{J}\dot{q} + J\ddot{q}$, (2.1) is written as

$$M_r \ddot{r} + C_r \dot{r} + G_r = U + \omega_r + f, \quad (3.2)$$

where $M_r = J^{-T} M(q) J^{-1}$, $G_r = J^{-T} G(q)$, $U = J^{-T} \tau_i$, $C_r = J^{-T} (C(q, \dot{q}) - M(q) J^{-1} \dot{J}) J^{-1}$, and $\omega_r = J^{-T} \omega(q, \dot{q}, t)$.

Suppose r_d is the desired position and f_d is the desired force. Then $f = G_e(r - r_e)$, $e_f = f - f_d = G_e(r - r_e) - f_d$. G_e is the rigidity matrix and r_e is reference position of the environment. $\dot{r} = G_e^{-1}(\dot{e}_f + \dot{f}_d)$ and $\ddot{r} = G_e^{-1}(\ddot{e}_f + \ddot{f}_d)$. Then error equation is

$$M_r G_e^{-1} \ddot{e}_f + C_r G_e^{-1} \dot{e}_f - e_f + \Delta = U + \omega_r, \quad (3.3)$$

where $\Delta = M_r G_e^{-1} \ddot{f}_d + C_r G_e^{-1} \dot{f}_d + G_r - f_d$. Then it is transformed into the state space. $x_1 = e_f, x_2 = \dot{e}_f + \alpha e_f$ and α is a positive number. Equation (3.3) becomes

$$\begin{aligned} \dot{x}_1 &= x_2 - \alpha x_1, \\ M_r G_e^{-1} \dot{x}_2 &= -C_r G_e^{-1} x_2 + x_1 - \Delta + U + \omega_r + \omega. \end{aligned} \quad (3.4)$$

where $\omega = M_r G_e^{-1} \alpha \dot{e}_f + C_r G_e^{-1} \alpha e_f$.

The improved RAN network approaches ω_r . ε_f is the approaching error of the network. $\omega_r = P_f W_f + X V_f + \varepsilon_f$, P_f is output matrix of the hidden layer, W_f is the weight matrix from hidden layer to output layer. X is input matrix, V_f is the weight from input layer to output layer. $P_f W_f$ is the contribution from hidden layer to output layer. $X V_f$ is contribution from input layer to output layer. Equation (3.4) can be derived as

$$\begin{aligned} \dot{x}_1 &= x_2 - \alpha x_1, \\ M_r G_e^{-1} \dot{x}_2 &= -C_r G_e^{-1} x_2 + x_1 - \Delta + U + P_f W_f + X V_f + \varepsilon_f + \omega \end{aligned} \quad (3.5)$$

ε_f is regarded as interfere and its evaluation signal is $z = 2c\varepsilon_f = 2cx_1$, then L_2 gain is $J = \sup_{\|\varepsilon_f\| \neq 0} (\|z\|_2 / \|\varepsilon_f\|_2)$.

Theorem 3.1. For (3.5) if the study law of network is given by the following equation:

$$\begin{aligned} \dot{W}_f &= -\eta W_f, \\ \dot{V}_f &= -\lambda V_f, \end{aligned} \quad (\lambda, \eta > 0) \quad (3.6)$$

the following controller is expressed for the force loop:

$$U = -(G_e + I)x_1 + \Delta - P_f W_f - X V_f - \omega - \left(\frac{1}{2\gamma^2} + \theta \right) (x_2^T G_e^{-1})^T \quad (3.7)$$

and c in $z = 2cx_1$ must meet to (3.8)

$$\alpha - 2c^2 = \beta, \quad (3.8)$$

where β and θ are given positive numbers, then the L_2 gain of closed-loop system (3.5) and (3.7) is less than γ .

Proof. For (3.5), the Lyapunov function is defined as

$$V = \frac{1}{2} x_1^T x_1 + \frac{1}{2} x_2^T G_e^{-1} M_r G_e^{-1} x_2 + \frac{1}{2} W_f^T W_f + \frac{1}{2} V_f^T V_f. \quad (3.9)$$

Then

$$\begin{aligned}
\dot{V} &= x_1^T \dot{x}_1 + \frac{1}{2} x_2^T G_e^{-1} \dot{M}_r G_e^{-1} x_2 + x_2^T G_e^{-1} M_r G_e^{-1} \dot{x}_2 + W_f^T \dot{W}_f + V_f^T \dot{V}_f \\
&= x_1^T (x_2 - \alpha x_1) + \frac{1}{2} (G_e^{-1} x_2)^T (M_r - 2C_r) (G_e^{-1} x_2) \\
&\quad + x_2^T G_e^{-1} (x_1 - \Delta + U + P_f W_f + X V_f + \varepsilon_f + \omega) + V_f^T \dot{V}_f + W_f^T \dot{W}_f.
\end{aligned} \tag{3.10}$$

Substituting (3.6) into the above equality, we have

$$\begin{aligned}
\dot{V} &= x_1^T (x_2 - \alpha x_1) + x_2^T G_e^{-1} (x_1 - \Delta + U + P_f W_f + X V_f + \varepsilon_f + \omega) \\
&\quad - \eta W_f^T W_f - \lambda V_f^T V_f \\
&= -\alpha x_1^T x_1 - \eta W_f^T W_f - \lambda V_f^T V_f \\
&\quad + x_2^T G_e^{-1} (G_e x_1 + x_1 - \Delta + U + P_f W_f + X V_f + \varepsilon_f + \omega).
\end{aligned} \tag{3.11}$$

According to HJI, we get

$$H = \dot{V} - \frac{1}{2} \gamma^2 \|\varepsilon_f\|^2 + \frac{1}{2} \|z\|^2. \tag{3.12}$$

Then

$$\begin{aligned}
H &= -\alpha x_1^T x_1 - \eta W_f^T W_f - \lambda V_f^T V_f - \frac{1}{2} \gamma^2 \|\varepsilon_f\|^2 + 2c^2 \|x_1\|^2 \\
&\quad + x_2^T G_e^{-1} (G_e x_1 + x_1 - \Delta + U + P_f W_f + X V_f + \varepsilon_f + \omega) \\
&= -(\alpha - 2c^2) \|x_1\|^2 - \eta \|W_f\|^2 - \lambda \|V_f\|^2 + x_2^T G_e^{-1} \varepsilon_f - \frac{1}{2} \gamma^2 \|\varepsilon_f\|^2 \\
&\quad \times x_2^T G_e^{-1} (G_e x_1 + x_1 - \Delta + U + P_f W_f + X V_f + \omega).
\end{aligned} \tag{3.13}$$

Due to

$$\begin{aligned}
x_2^T G_e^{-1} \varepsilon_f - \frac{1}{2} \gamma^2 \|\varepsilon_f\|^2 &= -\left(-x_2^T G_e^{-1} \varepsilon_f + \frac{1}{2} \gamma^2 \|\varepsilon_f\|^2 + \frac{1}{2\gamma^2} \|x_2^T G_e^{-1}\|^2 - \frac{1}{2\gamma^2} \|x_2^T G_e^{-1}\|^2 \right) \\
&= -\frac{1}{2} \left\| \frac{1}{\gamma} x_2^T G_e^{-1} - \gamma \varepsilon_f \right\|^2 + \frac{1}{2\gamma^2} \|x_2^T G_e^{-1}\|^2 \\
&\leq \frac{1}{2\gamma^2} \|x_2^T G_e^{-1}\|^2,
\end{aligned} \tag{3.14}$$

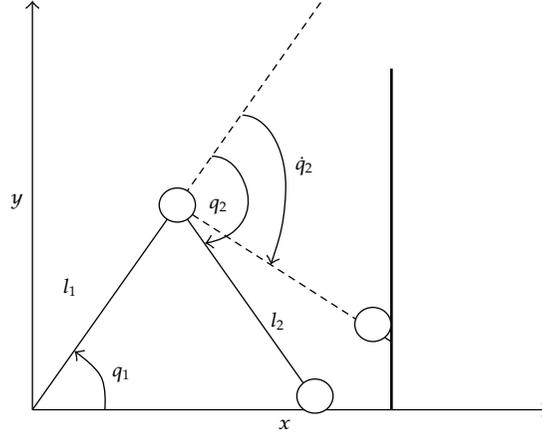


Figure 1: A two-link manipulator with constraint surface.

we get

$$H \leq -\beta \|x_1\|^2 - \eta \|W_f\|^2 - \lambda \|V_f\|^2 + x_2^T G_e^{-1} \times \left(G_e x_1 + x_1 - \Delta + U + P_f W_f + X V_f + \omega + \frac{1}{2\gamma^2} (x_2^T G_e^{-1})^T \right). \quad (3.15)$$

Substituting (3.7) into the above inequality, we have

$$H \leq -\beta \|x_1\|^2 - \eta \|W_f\|^2 - \lambda \|V_f\|^2 - \theta \|x_2^T G_e^{-1}\|^2 \leq 0. \quad (3.16)$$

So the system meets $\dot{V} \leq (1/2)\gamma^2 \|\varepsilon_f\|^2 - (1/2)\|z\|^2$. □

4. Experiments and Results

To verify the effectiveness of the proposed control strategy, we made some software simulation by using methods [11–16]. Here the model is based on two-link manipulator, which is shown in Figure 1.

In the simulation, we took a horizontal plane as the work space: $r = [x \ y]^T$ and describe the constraint surface as $X = 1.6$, the desired trajectory is $y_d = 0.007t + 0.5$, $t \in [0, 10]$, the desired force is $f_d = 5\text{N}$. Assume that the initial position of the manipulator end effector is $r_0 = [1.5 \ 0]^T$ and initial velocity is $dr = [0 \ 0]^T$. In order to analyze comparatively, we use PD control and robust neural network control, respectively, in the force control loop. First the model is controlled by PD controller. The PD parameters are determined by output result. $P = 57$, $D = 1.3$.

We adopt MATLAB Simulink and S-functions to design control system, the parameters are set $\alpha = 18.1$, $\beta = 0.1$, $\theta = 0.1$, $\gamma = 0.05$, $c = 3$, and $\eta = 0.1$, $\lambda = 0.05$. The simulation results

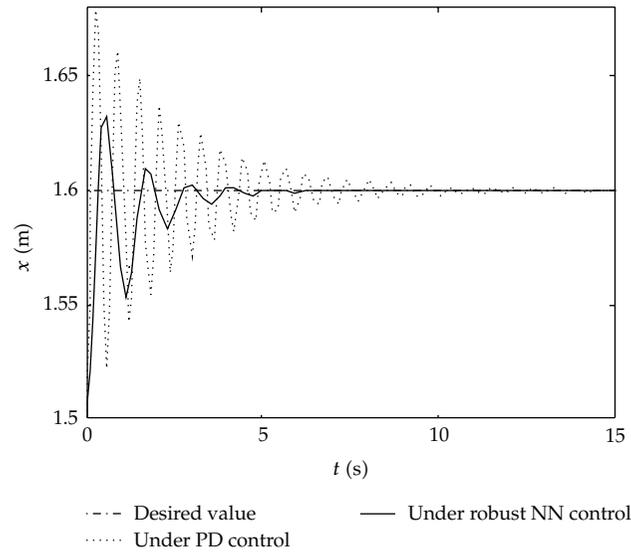


Figure 2: Tracking the location along x -axis under robust neural networks control and PD control.

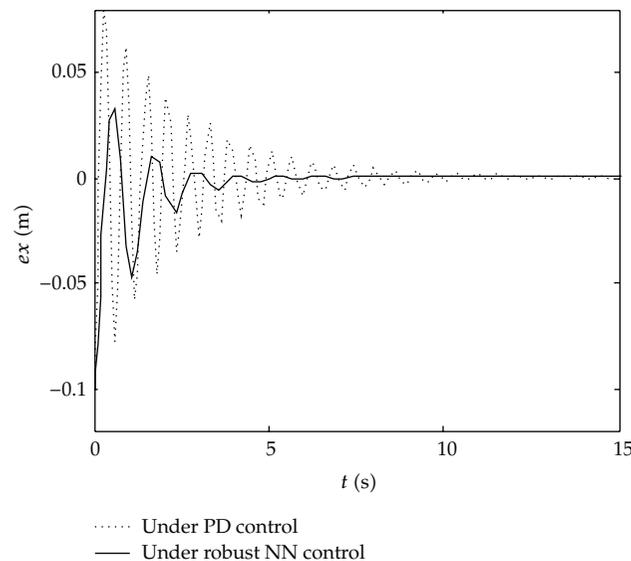


Figure 3: Tracking the location's error along x -axis under robust neural networks control and PD control.

are shown in Figures 2–6, among which Figures 2–5 give the tracking results of position and position error and Figure 6 gives force tracking results.

Figures 2 and 3 show that the control effect along x -axis is unlikeliness. The robust NN control result is superior to conventional PD along x -axis. Figures 4 and 5 show that there's no obvious difference along y -axis.

Figure 6 shows that the methods under robust neural networks control and PD control can make force convergence desired value. But the effort of the two methods has great difference. The oscillation is severe, and convergence speed is slow under PD control

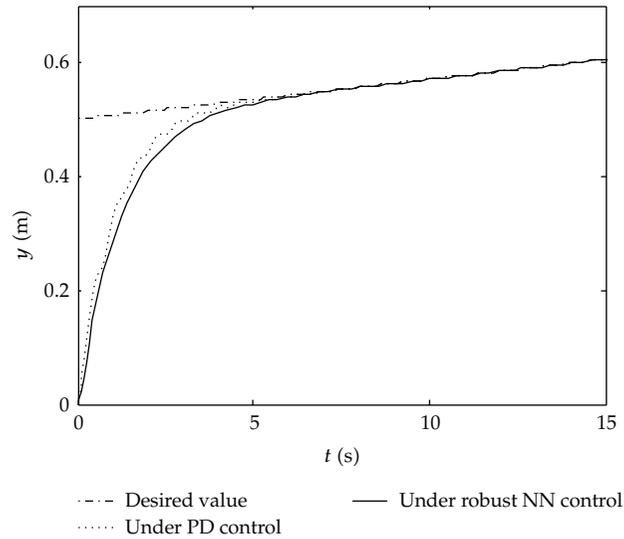


Figure 4: Tracking the location along y -axis under robust neural networks control and PD control.

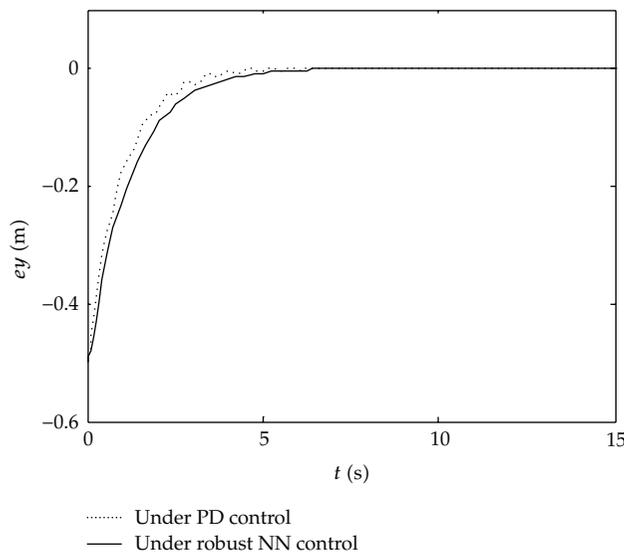


Figure 5: Tracking the location's error along y -axis under robust neural networks control and PD control.

method. The oscillation and convergence speed are improved under RAN NN control method. The stability and transient performance are greatly superior to the effect under PD control.

From the simulation results, we know that the improved RBF neural network robust control method can decrease the dramatic oscillation and improve the convergence speed. The stability and transient performance of the system are much better than the PD control, and therefore it is an effective control method.

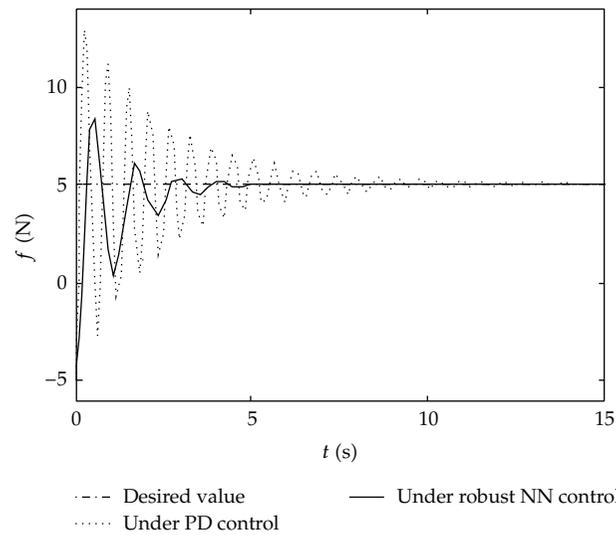


Figure 6: The force under robust neural networks control and PD control.

5. Conclusion

An improved RAN NN controller has been designed in this paper for robot. In case of difficulty measuring an external disturbance, the upper bound of uncertainty cannot be obtained. The controller can make the system's uncertainty significantly reduced without obtaining the upper bound of uncertainty. It is found that the system can obtain good transient performance and strong adaptability. For the force and position control, it has good robustness and tracking ability. For future study, a simulation platform is constructed in the paper to intuitively demonstrate the control process.

References

- [1] M. H. Raibert and J. J. Craig, "Hybrid position/force control of manipulators," *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, vol. 103, no. 2, pp. 126–133, 1981.
- [2] 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.
- [3] S. Y. Chen, Y. F. Li, and J. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
- [4] 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.
- [5] C.-H. Kao, C.-F. Hsu, C.-H. Wang, and H. S. Don, "Chaos synchronization using adaptive dynamic neural network controller with variable learning rates," *Mathematical Problems in Engineering*, vol. 2011, Article ID 701671, 20 pages, 2011.
- [6] L. A. Mozelli and R. M. Palhares, "Less conservative H_∞ fuzzy control for discrete-time takagi-sugeno systems," *Mathematical Problems in Engineering*, vol. 2011, Article ID 361640, 21 pages, 2011.
- [7] M. M. Belhaouane, M. F. Ghariani, H. Belkhiria Ayadi, and N. B. Braiek, "Improved results on robust stability analysis and stabilization for a class of uncertain nonlinear systems," *Mathematical Problems in Engineering*, vol. 2010, Article ID 724563, 24 pages, 2010.
- [8] Z.-Y. Xing, Y. Qin, X.-M. Pang, L.-M. Jia, and Y. Zhang, "Modelling of the automatic depth control electrohydraulic system using RBF neural network and genetic algorithm," *Mathematical Problems in Engineering*, vol. 2010, Article ID 124014, 16 pages, 2010.

- [9] A. L. Yu, "Research on the dynamic modeling based on genetic wavelet neural network for the robot wrist force sensor," *Acta Physica Sinica*, vol. 57, no. 6, pp. 3385–3390, 2008.
- [10] L. JinKun, *The Design and MATLAB Simulation of Robot Control System*, Tsinghua University, Beijing, China, 2008.
- [11] C. G. Looney, "Radial basis functional link nets and fuzzy reasoning," *Neurocomputing*, vol. 48, pp. 489–509, 2002.
- [12] C. Z. Xing, *Robotics*, Tsinghua University, Beijing, China, 2000.
- [13] 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*. In press.
- [14] Y. Zhao and C. C. Cheah, "Hybrid vision-force control for robot with uncertainties," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 261–266, May 2004.
- [15] S. Chiaverini and L. Sciavicco, "Force/position regulation of compliant robotmanipulators," *IEEE Transactions on Robotics and Automation*, vol. 9, no. 4, pp. 361–373, 1993.
- [16] Z. Doulgeri and S. Arimoto, "A position/force control for a robot finger with soft tip and uncertain kinematics," *Journal of Robotic Systems*, vol. 19, no. 3, pp. 115–131, 2002.

Research Article

Incorporating a Local Binary Fitting Model into a Maximum Regional Difference Model for Extracting Microscopic Information under Complex Conditions

Chunyan Yao,^{1,2} Jianwei Zhang,² Min Chen,³ Qiu Guan,³ and Massimo Scalia⁴

¹ Key Laboratory of E&M, Zhejiang University of Technology, Ministry of Education of China, 310014 Hangzhou, China

² Department of Informatics, University of Hamburg, 22527 Hamburg, Germany

³ College of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China

⁴ Department of Mathematics, Sapienza University of Rome, P.le A. Moro 2, 00185 Rome, Italy

Correspondence should be addressed to Chunyan Yao, yao@informatik.uni-hamburg.de

Received 30 March 2011; Accepted 15 May 2011

Academic Editor: Shengyong Chen

Copyright © 2012 Chunyan Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a novel region-based method for extracting useful information from microscopic images under complex conditions. It is especially used for blood cell segmentation and statistical analysis. The active model detects several inner and outer contours of an object from its background. The method incorporates a local binary fitting model into a maximum regional difference model. It utilizes both local and global intensity information as the driving forces of the contour model on the principle of the largest regional difference. The local and global fitting forces ensure that local dissimilarities can be captured and globally different areas can be segmented, respectively. By combining the advantages of local and global information, the motion of the contour is driven by the mixed fitting force, which is composed of the local and global fitting term in the energy function. Experiments are carried out in the laboratory, and results show that the novel model can yield good performances for microscopic image analysis.

1. Introduction

Observing and analyzing microscopic information has become one of the most important ways for exploring the microcosmic world for mankind. One typical issue is cell image analysis, which is at the forefront of clinics and bioengineering. In recent years, biological discovery and its translation into new clinical therapies have advanced rapidly. Pathologists use histopathological images of biopsy samples removed from patients and examine them

under a microscope. While examining such images, a pathologist typically assesses the deviations in the cell structures and the change in the proportion between nucleolus and enchylema across the tissue under examination [1]. It is crucial to discover an automatic method to detect the contour of the cell and its nucleolus [2]. Cell detection is critical to disease diagnosis and prevention.

Dealing with cell images, there may exist many complex conditions such as large density of many cell objects, overlapping, and complex nucleus structures. This paper proposes a hybrid contour model for segmentation of the cell nucleolus and membranes. The proposed algorithm can handle cell overlapping and calculate a number of morphological parameters as well. By adopting the affinity propagation algorithm, a classifier can be designed for cell classification. Local and global intensity information is used as the driving forces of the active contour model under the guidance of the maximum regional difference. The local and global fitting forces ensure that local dissimilarities can be captured and global different areas can be segmented, respectively. This hybrid contour can capture the image details in the nucleus area relative to the traditional contour model, which is important to study the morphological nucleus parameters.

A tremendous amount of research work has been conducted in the area of microscopic image processing [3]. There are some challenges mentioned in these studies. The noise pollution is the most frequently occurring problem. It arises from staining the biopsy samples, and uneven distribution of the stain usually causes problems in processing the stained material. Besides noise, a strong shading effect is apparent in the images so that the cell images do not appear smooth. Accurate cell segmentation often precedes other analyses such as that of the morphologic process and behavior. Active contour models have been among the most successful methods facing the challenge mentioned above [4, 5]. Techniques based on active contour models have the potential to produce better estimates of cell morphologies.

The existing active contour models can be categorized into two classes: edge-based models [5–7] and region-based models [8–10]. The edge-based model directly uses intensity gradient information to attract the contour toward the object boundaries. Therefore, this kind of model has a worse performance for weak object boundaries since the cell image is very fuzzy due to low contrast at the location of the cell membrane. The region-based model aims to identify each region of interest by using a certain region descriptor. It guides the motion of the contour and is to some extent less sensitive to the location of initial contours [11, 12]. It is therefore more suitable for cell segmentation [13].

The Chan-Vese model [14] is one of the most popular region-based active contour models. This model has been successfully used for segmenting images. However, the regions of interest in images are often not statistically smooth and, therefore, the models are not applicable to cell images. Li et al. mentioned the intensity inhomogeneity as the biggest problem in low-signal-noise-ratio images such as biopsy samples and medical images [15].

In this paper, a novel model based on the region-based active contour model is proposed, which is able to detect the local unclear boundaries and segment the main differing regions successfully according to the principle of the maximum interregional difference. An energy function with two main terms, the local fitting term and the global fitting term, is defined in this model. The local fitting term produces a strong force to attract the contours and makes them stop at the boundaries although the object boundary is not clear or even blurry. The global fitting term ensures that the curve extracts the main object regions on the principle of the maximal regional difference. In addition, a strategic weight parameter, which is defined using the image intensity gradient information, is introduced to make the

two forces work together as a hybrid model. This parametric model induces a mixed force comprising local and global information as driving force.

2. Maximum Regional Difference Model

Chan and Vese [14] proposed an active contour model that segments an image into two sets of possibly disjoint regions, by minimizing a simplified Mumford-Shah function. Assume that $\Omega \subset \mathfrak{R}^2$ are the image domain and $I : \Omega \rightarrow \mathfrak{R}$ are the given image. Mumford and Shah consider segmentation as a problem of seeking an optimal contour C that divides the image domain into two approximated piecewise-constant regions with intensities u_i and u_o . Denote C as its boundary. Thus, the global data fitting term in the Chan-Vese model is defined as follows:

$$E^{CV}(c_1, c_2) = \int_{\bar{\Omega}} (I - c_1)^2 dx dy + \int_{\Omega} (I - c_2)^2 dx dy, \quad (2.1)$$

where Ω and $\bar{\Omega}$ represent the regions outside and inside the contour C , respectively, and c_1 and c_2 are two constants that fit the image intensities outside and inside C .

This model considers pixels within the same region as having the most similarity and makes up for the shortcomings of the edge detector. When the contour accurately represents the object boundary, the two fitting terms minimize the fitting energy value. In each segmented area, the clustered pixels' mean value approximately equals c_1 and c_2 , respectively. Thus the fitting terms with respect to c_1 and c_2 are the driving forces that evolve the curve motion on the principle of inner region homogeneity.

Since the regional difference is the guideline in image segmentation, the interregional differences should be considered as the model's driving force as follows:

$$E = -\frac{1}{2} \cdot (c_1 - c_2)^2. \quad (2.2)$$

This kind of region-based active contour model energy is characterized by the maximum dissimilarity between regions. To minimize the energy E in (2.2) is the same as maximizing the difference between different regions. Equation (2.2) formulates the global instructive guidance term. This paper revises the guidance term which is defined as follows:

$$E = -\frac{A}{2} \cdot (c_1 - c_2)^2, \quad (2.3)$$

where A represents the entire image area. The curve evolution of the contour C is defined as

$$E^G(c_1, c_2) = A \cdot (c_1 - c_2) \left(\frac{I - c_1}{A_{c1}} + \frac{I - c_2}{A_{c2}} \right) \cdot N, \quad (2.4)$$

where $A_{c1} = \int_{\bar{\Omega}} dx dy$ and $A_{c2} = \int_{\Omega} dx dy$, denoting the areas outside of the contour and inside of the contour, respectively, and N denotes the outward unit normal of C . In order to show a stable value of (2.4) rather than an oscillating one when the contour divides

the area with a small value, that is, to avoid the denominator in (2.4) from being zero, we approximately evolve the E^G as follows:

$$E^G(c_1, c_2) = (c_1 - c_2)(I - c_1p_2 - c_2p_1) \cdot N, \quad (2.5)$$

where p_1 represents the ratio between A_{c_1} and A and p_2 represents the ratio between A_{c_2} and A .

3. Incorporation of the Local Binary Fitting Model

Though the E^G in (2.5) regards the regional difference as the guideline of curve evolution, the region-based model has little ability to deal with the intensity inhomogeneity or shading effect caused in the imaging process. The region-based model considers the global difference as the driving force rather than local detail information, and thus it needs a term of local fitting energy to improve the performance of segmentation in the inhomogeneous area.

Li et al. proposed a local binary fitting (LBF) model, which utilizes the local intensity information [15, 16]. An important parameter called the kernel function is introduced to denote the range of the local area. Additionally, f_1 and f_2 as the two spatially varying fitting functions are used to approximate the local intensities [17, 18]. In the LBF model, the local fitting term is defined as

$$E^{\text{LBF}}(\phi, f_1, f_2) = \lambda_1 \int \left[\int K_\sigma(x - y) |I(y) - f_1(x)|^2 H(\phi(y)) dy \right] dx \\ + \lambda_2 \int \left[\int K_\sigma(x - y) |I(y) - f_2(x)|^2 (1 - H(\phi(y))) dy \right] dx, \quad (3.1)$$

where H is the Heaviside function and K_σ is a Gaussian kernel that $K_\sigma(x)$ decreases and approaches zero as $|x|$ increases. The function $f(x)$ calculates the fitting degree around point x , and x can be considered as the center point of the local area. One of the greatest advantages of this LBF model is that it discovers the object boundaries more precisely than the Chan-Vese model.

The proposed method makes use of the advantages of the region-based model, as in (2.5), and the local binary fitting model by taking the local and global intensity terms into account. The entire energy model is defined as

$$E^M(\phi, c_1, c_2, f_1, f_2) = (1 - \lambda_g) \cdot E^{\text{LBF}}(\phi, f_1, f_2) + \lambda_g E^G(\phi, c_1, c_2), \quad (3.2)$$

where λ_g is the strategic weight parameter, $E^{\text{LBF}}(\phi, f_1, f_2)$ is defined in (3.1), and $E^G(\phi, c_1, c_2)$ is defined in (2.5). The strategic weight parameter λ_g is defined as

$$\lambda_g(x, y) = \frac{1}{(1 + \alpha |\nabla I(x, y)|)}, \quad (3.3)$$

where $|\nabla I(x, y)|$ represents the gradient information in image I and α is a positive constant.

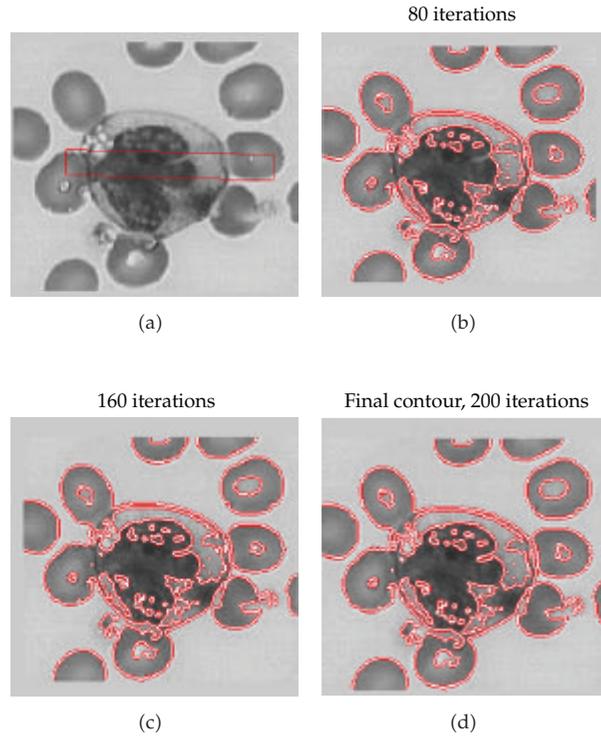


Figure 1: Neutrophil segmentation and level set function ϕ .

This strategic weight parameter realizes the trade-off into a hybrid model [19, 20]. Working with the mixed energy restriction terms, the curve's driving force appears strategically. When the local image area is flat or far away from the object boundaries, the value of $|\nabla I(x, y)|$ is relatively small therefore the right side of (3.3) has a value approximating one. The hybrid model is guided to maximize the interregional difference. $|\nabla I(x, y)|$ would gain a big value when the point (x, y) is located in the region with intensity inhomogeneity where the object boundaries exist, thus the value of $\lambda_g(x, y)$ is close to zero. The model possesses a potential force to extract local boundaries. It combines the advantages of the local fitting model and the global regional difference model by the use of intensity gradients as its strategic parameter.

By minimizing the energy function in (3.2), we face solving a partially differential equation. The gradient descent algorithm is chosen to give a numerical solution with respect to the PDE. In addition, we introduce a fitting term mentioned in [5] to regularize the level set function for accurate computation by penalizing the deviation of the level set function ϕ from a signed distance function. It is defined as

$$P(\phi) = \int \frac{1}{2} (|\nabla \phi(x)| - 1)^2 dx. \quad (3.4)$$

Besides, we introduce a length restriction term that has been used in the Chan-Vese model to

smooth the zero level set contour,

$$L(\phi) = \int |\nabla H(\phi(x))| dx. \quad (3.5)$$

Now, the entire model proposed in this paper is defined as

$$\begin{aligned} E^E(\phi, c_1, c_2, f_1, f_2) \\ = E^M(\phi, c_1, c_2, f_1, f_2) + P(\phi) + L(\phi). \end{aligned} \quad (3.6)$$

The fitting functions f_1 and f_2 are given by

$$\begin{aligned} f_1(x) &= \frac{K_\sigma^*(H(\phi) \cdot I)}{K_\sigma^*H(\phi)}, \\ f_2(x) &= \frac{K_\sigma^*[(1 - H(\phi)) \cdot I]}{K_\sigma^*[1 - H(\phi)]}. \end{aligned} \quad (3.7)$$

c_1 and c_2 are constants defined as in [14].

4. Experiments and Results

4.1. Implementation Issues

To minimize the energy function with respect to ϕ defined in (3.6), we give the solution using a gradient descent algorithm which is defined as

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= \lambda_g \cdot \delta(\phi) \cdot (c_1 - c_2)(I - c_1 p_2 - c_2 p_1) + (1 - \lambda_g) \cdot \delta(\phi) \cdot [\lambda_1 e_1 - \lambda_2 e_2] \\ &+ v \cdot \delta(\phi) \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \mu \left(\nabla^2 \phi - \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \right), \end{aligned} \quad (4.1)$$

where λ_1, λ_2, v , and μ are constants and $\delta(\cdot)$ is the Dirac delta function. e_1 and e_2 are given by

$$e_i(x) = \int K_\sigma(y - x) |I(x) - f_i(y)|^2 dy, \quad i = 1, 2. \quad (4.2)$$

The iteration scheme is used by discretizing the PDE (4.1). The level set function ϕ ultimately sets the image object boundaries at $\phi = 0$. Furthermore, we set the parameters $\lambda_1 = \lambda_2 = 1$, $\mu = 1$, and $v = 60$. α equals three in the expression of $\lambda_g(x, y) \cdot \sigma$ in the Gaussian kernel K_σ equals two.

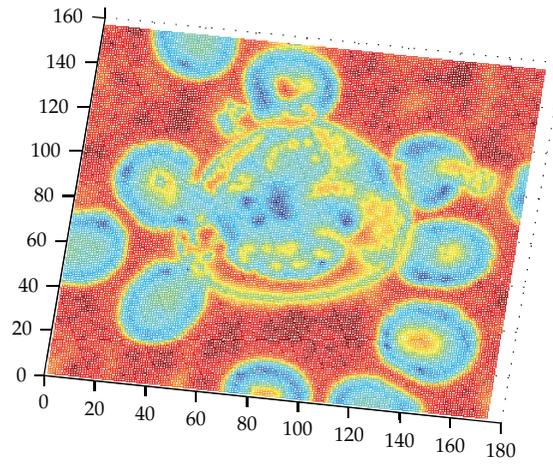


Figure 2: Level set function ϕ after 200 iterations.

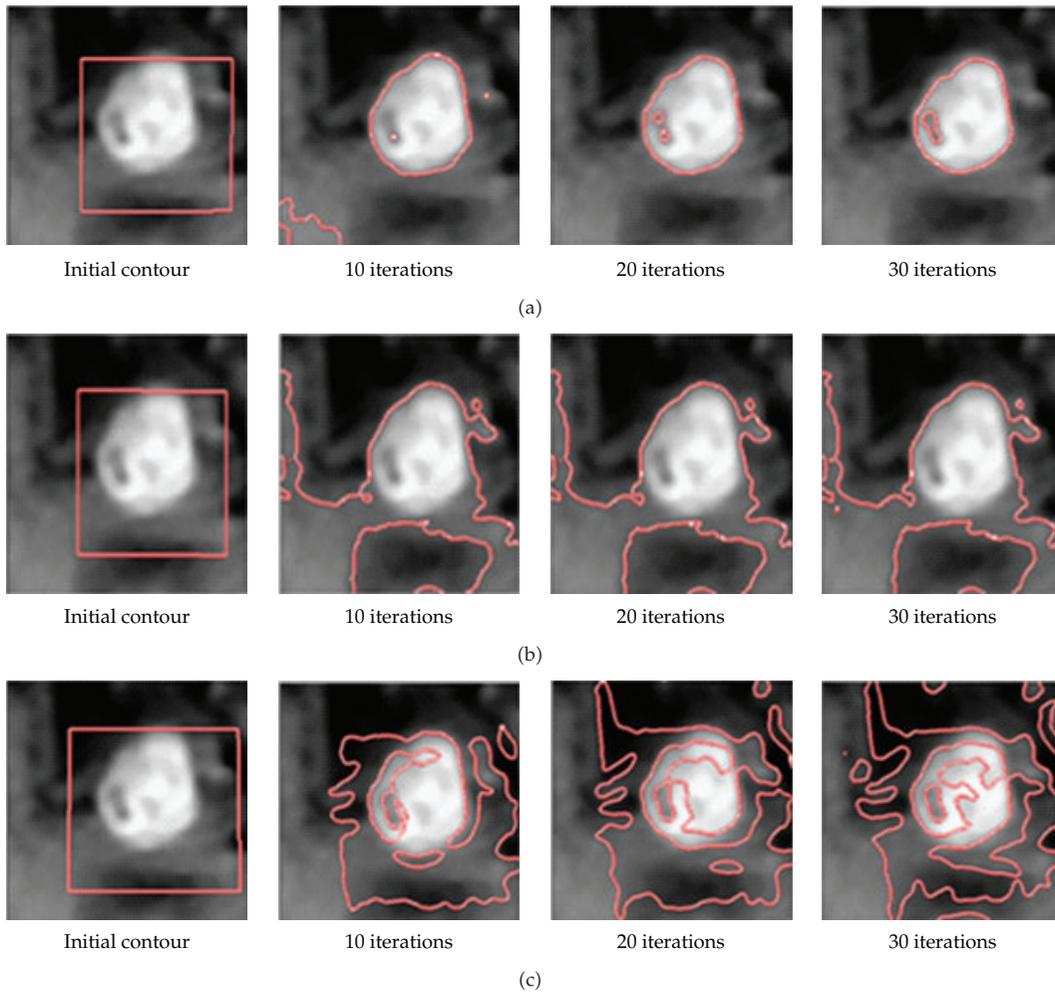


Figure 3: Comparison of different models.

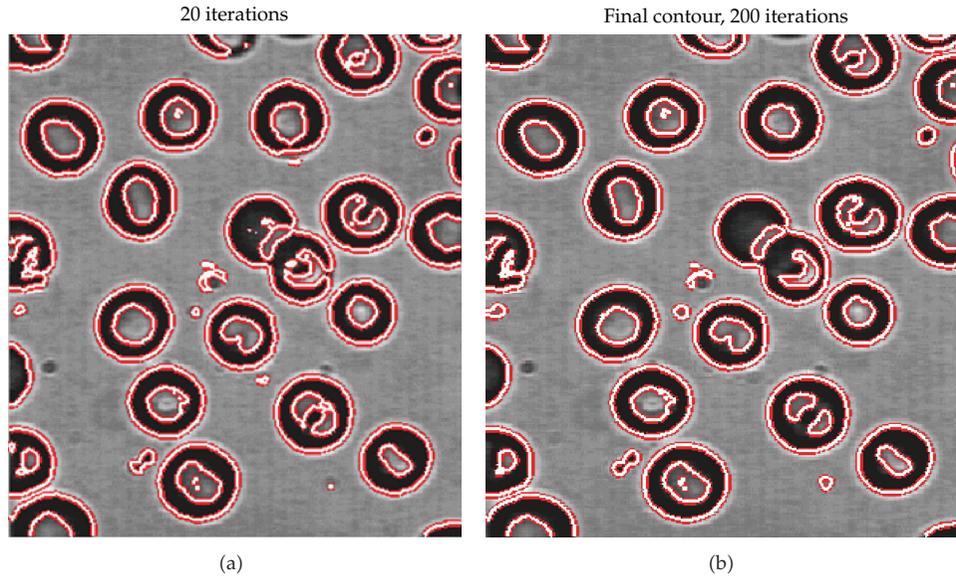


Figure 4: Another example for cell segmentation using the proposed model.

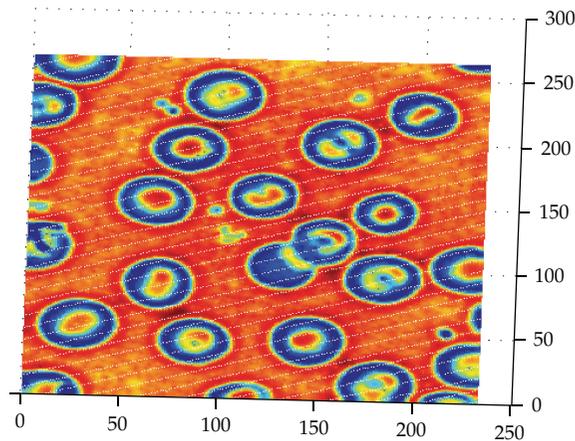


Figure 5: Corresponding function ϕ after 200 iterations.

4.2. Experimental Results

In this section, we use some cell images to demonstrate the mixed driving force in the proposed method for verifying its performance in detecting the contours of cell nucleolus and membrane. Figure 1 shows the process for segmenting neutrophil images. In Figure 1(a), the upper left picture depicts the initial contour and the Figures 1(b), 1(c), and 1(d) show how the curve evolves. It can easily establish the main cell nucleolus boundary and the weak object boundary of the cell membrane. The images resulting after 200 iterations of ϕ are shown in Figure 2. Due to its strong ability to capture local difference, the model can detect the weak boundary of the neutrophil membrane while it accurately distinguishes the nucleolus from the cytoplasm. The global driving force ensures that the model to segment the object has the

maximum regional difference and that failure cases are avoided. Moreover, a comparison has also been carried out in our experiments with the Chan-Vese model and the LBF model, using a tumor lymphocyte cell image as test images. The results are shown in Figure 3. Figure 3(a) is the result by our proposed model. Figures 3(b) and 3(c) show the results of the Chan-Vese and LBF model, respectively. The tumor cell in an intensity inhomogeneous condition is detected by the zero level set function. The model successfully extracts the nucleolus, whose size relative to the area of the cell is a significant parameter in biology.

Figure 4 illustrates another example for cell segmentation using the proposed model, where the parameters are $\alpha = 3$, $\nu = 0$, $\lambda_1 = 1$, $\lambda_2 = 2$, and $\sigma = 3$. Figure 5 shows the corresponding function ϕ after 200 iterations.

5. Conclusion

The model proposed in this paper yields a relatively more desirable performance than the traditional models, although the results still feature over segmentation in the nucleolus area. However, this over segmentation problem can be overcome by a postmorphological process. For example, we find the locations where function $\phi < 0$ and $\phi > 0$ ($\phi = 0$ denotes the contour of the object) and substitute them with a binary value of zero and one, respectively. Then, we give a postprocess by setting a threshold value of the area size in the region of the neutrophil nucleolus area to eliminate the oversegmentation.

From the experimental results, it follows that the proposed hybrid model can well capture the weak edge boundary by using the local fitting terms, while it observes the rule of regional difference and uses it as guideline to form the fitting terms that drive the contour to accurately capture the boundaries. The experiments demonstrate that it performs well in segmenting the nucleolus and cell membrane.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51075367, 60870002) and the Science and Technology Department of Zhejiang Province (R1110679, 2010R10006, and 2010C33095).

References

- [1] S.-J. Chern and P.-C. Huang, "On the existence of a weak solution of a half-cell model for PEM fuel cells," *Mathematical Problems in Engineering*, Article ID 701096, 16 pages, 2010.
- [2] 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.
- [3] C. Gunduz, B. Yener, and S. H. Gultekin, "The cell graphs of cancer," *Bioinformatics*, vol. 20, no. 1, pp. 145–151, 2004.
- [4] C. Li, J. Liu, and M. D. Fox, "Segmentation of external force field for automatic initialization and splitting of snakes," *Pattern Recognition*, vol. 38, no. 11, pp. 1947–1960, 2005.
- [5] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: a new variational formulation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 430–436, 2005.
- [6] A. Vasilevskiy and K. Siddiqi, "Flux maximizing geometric flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1565–1578, 2002.
- [7] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: a level set approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 158–175, 1995.

- [8] S. Y. Chen, Y. F. Li, and J. Zhang, "Realtime structured light vision with the principle of unique color codes," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '07)*, pp. 429–434, Roma, Italy, April 2007.
- [9] A. Tsai, A. Yezzi, and A. S. Willsky, "Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification," *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1169–1186, 2001.
- [10] L. Vese and T. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 271–293, 2002.
- [11] 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*, Article ID 416963, 12 pages, 2011.
- [12] S. Y. Chen and Y. F. Li, "Vision sensor planning for 3-D model acquisition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 5, pp. 894–904, 2005.
- [13] M. Chen, S. Chen, and Q. Guan, "Hybrid contour model for segmentation of cell nucleolus and membranes," in *Proceedings of the 2nd International Conference on Biomedical Engineering and Informatics (BMEI '09)*, Tianjin, China, October 2009.
- [14] T. Chan and L. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [15] C. Li, C. Y. Kao, J. C. Gore, and Z. Ding, "Implicit active contours driven by local binary fitting energy," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, June 2007.
- [16] C. Li, C.-Y. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1940–1949, 2008.
- [17] S. Y. Chen, "Cardiac deformation mechanics from 4D images," *Electronics Letters*, vol. 43, no. 11, pp. 609–611, 2007.
- [18] C.-G. Zhu and R.-H. Wang, "Least squares fitting of piecewise algebraic curves," *Mathematical Problems in Engineering*, Article ID 78702, 11 pages, 2007.
- [19] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, Article ID 157264, 26 pages, 2010.
- [20] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, Article ID 408418, 22 pages, 2010.