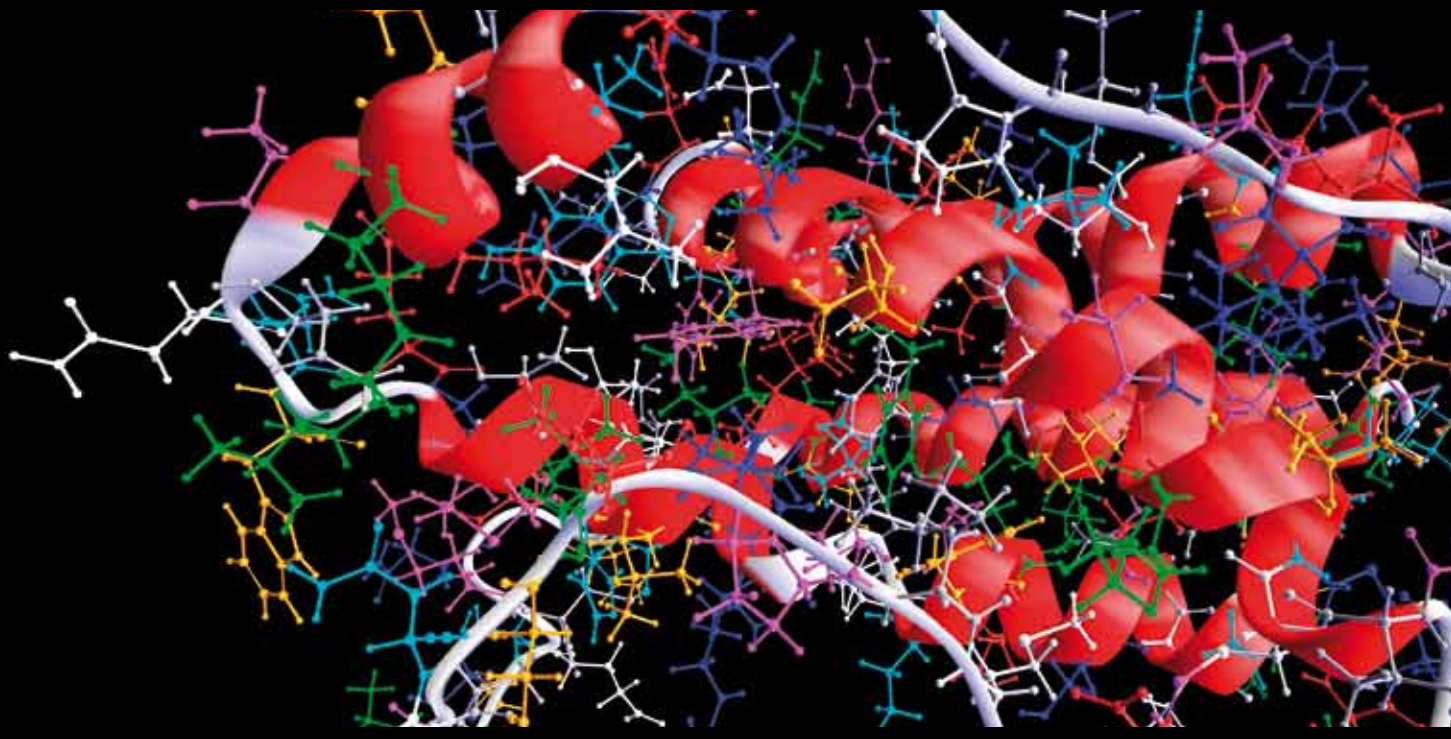
Computational and Mathematical Methods in Medicine

Biomedical Signal Processing and Modeling Complexity of Living Systems 2013

Guest Editors: Carlo Cattani, Radu Badea, Sheng-yong Chen, and Maria Crisan



Biomedical Signal Processing and Modeling Complexity of Living Systems 2013

Computational and Mathematical Methods in Medicine

Biomedical Signal Processing and Modeling Complexity of Living Systems 2013

Guest Editors: Carlo Cattani, Radu Badea, Sheng-yong Chen, and Maria Crisan

Editorial Board

Contents

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 173469, 2 pages
http://dx.doi.org/10.1155/2013/173469

Editorial

Biomedical Signal Processing and Modeling Complexity of Living Systems 2013

Carlo Cattani,[1] Radu Badea,[2] Sheng-Yong Chen,[3] and Maria Crisan[4]

[1] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy*
[2] *Department of Clinical Imaging Ultrasound, "Iuliu Hatieganu" University of Medicine and Pharmacy, 400000 Cluj-Napoca, Romania*
[3] *College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China*
[4] *Department of Histology, "Iuliu Hatieganu" University of Medicine and Pharmacy, 400000 Cluj-Napoca, Romania*

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

Received 7 November 2013; Accepted 7 November 2013

Biomedical signal processing aims to provide significant insights into the analysis of the information flows from physiological signals. As such, it can be understood as a specific interdisciplinary scientific discipline. In fact, biomedical signals extract information from complex biological models thus proposing challenging mathematical problems, whose solution has to be interpreted from a biological point of view. The focus of this special issue is the mathematical analysis and modeling of time series in living systems and biomedical signals. The main steps of the biomedical signals processing are as follows.

(1) Signal processing of biological data implies many different interesting problems dealing with signal acquisition, sampling, and quantization. The noise reduction and similar problems as image enhancement are a fundamental step in order to avoid significant errors in the analysis of data. Feature extraction is the most important part of the analysis of biological signals because of the importance which is clinically given to even the smallest singularity of the image (signal).

(2) Information flows from signals imply the modeling and analysis of spatial structures, self-organization, environmental interaction, behavior, and development. Usually this is related to the complexity analysis in the sense that the information flows come from complex systems so that signals show typical features, such as randomness, nowhere differentiability, fractal behavior, and self-similarity which characterize complex systems. As a consequence typical parameters of complexity such as entropy, power spectrum, randomness, and multifractality play a fundamental role, because their values can be used to detect the emergence of clinical pathologies.

(3) Physiological signals usually come as 1D time series or 2D images. The most known biosignals are based on sounds (ultrasounds), electromagnetic pulses (ECG, EEG, and MRI), radiation (X-ray and CT), images (microscopy), and many others. The clinical signal understanding of them follows from the correct, from a mathematical point of view, interpretation of the signal.

(4) Physiological signals are detected and measured by modern biomedical devices. Among others, one of the main problems is to optimize both the investigation methods and the device performances.

The papers selected for this special issue represent a good panel in recent challenges. They represent some of the most recent advances in many different clinical investigations devoted to the analysis of complexity in living systems, like, for example, network science, dynamical systems theory, dynamical complexity, pattern analysis, implementation, and algorithms. They cannot be exhaustive because of the rapid growing both of mathematical methods of signal analysis and of the technical performances of devices. However they aim

to offer a wide introduction on a multidisciplinary discipline and to give some of the more interesting and original solution of challenging problems. Among them the most fascinating is to understanding of the biological structure and organization, the intracellular exchange of information, the localization of information in cell nuclei, and in particular the unrevealing of the mathematical information (functionally related) content in DNA.

This special issue contains 23 papers. In the category of modeling dynamical complexity, L.-P. Tian et al. make complex analysis and parameter estimation of dynamic metabolic systems. M. Adib and E. Cretu present wavelet-based artifact identification and separation technique for EEG signals during galvanic vestibular stimulation. X. Wu and N. Wu use thresholded two-phase test sample representation for outlier rejection in biological recognition. Z. Ma et al. propose nonlinear Radon transform using Zernike moment for shape analysis. C.-Y. Liou et al. study structural complexity of DNA sequence. M. Li et al. investigate heavy-tailed prediction error in predicting biomedical signals of 1/f noise type. X. Wang et al. propose reliable RANSAC using a novel preprocessing model. J. Zheng et al. give fast discriminative stochastic neighbor embedding analysis.

In the category of methods for analysis of dynamical complexity, R. Schiavetti and G. Sannino give in vitro evaluation of ferrule effect and depth of post insertion on fracture resistance of fiber posts. G. Sannino and G. Vairo make comparative evaluation of osseointegrated dental implants based on platform-switching concept and find influence of diameter, length, thread shape, and in-bone positioning depth on stress-based performance. H.-T. Wu et al. use multiscale cross-approximate entropy analysis as a measure of complexity among the aged and diabetic. T. Kauppi et al. construct benchmark databases and protocols for medical image analysis with diabetic retinopathy. B. Zhu et al. present a novel automatic detection system for ECG arrhythmias using maximum margin clustering with an immune evolutionary algorithm. Y.-S. Juang et al. study optimization and implementation of scaling-free CORDIC-based direct digital frequency synthesizer for body care area network systems. Z. Bian et al. find the effect of Pilates training on alpha rhythm.

In the category of biomedical signal analysis, A. F. Badea et al. give fractal analysis of elastographic images for automatic detection of diffuse diseases of salivary glands. Q. Guan et al. present Bayes clustering and structural support vector machines for segmentation of carotid artery plaques in multicontrast MRI. J. Zhang et al. present self-adaptive image reconstruction inspired by insect compound eye mechanism. X. Zheng et al. improve spatial adaptivity of nonlocal means in low-dosed CT imaging using pointwise fractal dimension. N. Wu et al. study three-dimensional identification of microorganisms using a digital holographic microscope. Y. Tang et al. propose a computational approach to seasonal changes of living leaves. L. Lin et al. study plane-based sampling for a ray casting algorithm in sequential medical images. Y.-S. Juang et al. propose a rate-distortion-based merging algorithm for compressed image segmentation.

As already mentioned, the topics and papers are not an exhaustive representation of the area of biomedical signal processing and modeling complexity of living systems. However we believe that we have succeeded to collect some of the most significant papers in this area aiming to improve the scientific debate in the modern interdisciplinary field of biomedical signal processing.

Acknowledgments

We thank the authors for their excellent contributions and discussions on modern topics. The reviewers also deserve our special thanks for their useful comments on the papers that helped the authors to clarify some crucial points.

Carlo Cattani
Radu Badea
Sheng-Yong Chen
Maria Crisan

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 698341, 8 pages
http://dx.doi.org/10.1155/2013/698341

Research Article

Complexity Analysis and Parameter Estimation of Dynamic Metabolic Systems

Li-Ping Tian,[1] Zhong-Ke Shi,[2] and Fang-Xiang Wu[3,4]

[1] *School of Information, Beijing Wuzi University, Beijing 101149, China*
[2] *School of Atuomation, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China*
[3] *Department of Mechanical Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, Canada S7N 5A9*
[4] *Division of Biomedical Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK, Canada S7N 5A9*

Correspondence should be addressed to Fang-Xiang Wu; faw341@mail.usask.ca

Received 24 April 2013; Revised 18 August 2013; Accepted 5 September 2013

Academic Editor: Shengyong Chen

A metabolic system consists of a number of reactions transforming molecules of one kind into another to provide the energy that living cells need. Based on the biochemical reaction principles, dynamic metabolic systems can be modeled by a group of coupled differential equations which consists of parameters, states (concentration of molecules involved), and reaction rates. Reaction rates are typically either polynomials or rational functions in states and constant parameters. As a result, dynamic metabolic systems are a group of differential equations nonlinear and coupled in both parameters and states. Therefore, it is challenging to estimate parameters in complex dynamic metabolic systems. In this paper, we propose a method to analyze the complexity of dynamic metabolic systems for parameter estimation. As a result, the estimation of parameters in dynamic metabolic systems is reduced to the estimation of parameters in a group of decoupled rational functions plus polynomials (which we call improper rational functions) or in polynomials. Furthermore, by taking its special structure of improper rational functions, we develop an efficient algorithm to estimate parameters in improper rational functions. The proposed method is applied to the estimation of parameters in a dynamic metabolic system. The simulation results show the superior performance of the proposed method.

1. Introduction

Living cells require energy and material for maintaining their essential biological processes through metabolism, which is a highly organized process. Metabolic systems are defined by the enzymes dynamically converting molecules of one type into molecules of another type in a reversible or irreversible manner. Modeling and parameter estimation in dynamic metabolic systems provide new approaches towards the analysis of experimental data and properties of the systems, ultimately leading to a great understanding of the language of living cells and organisms. Moreover, these approaches can also provide systematic strategies for key issues in medicine, pharmaceutical, and biotechnological industries [1]. The formulation and identification of metabolic systems generally includes the building of the mathematical model of biological process and the estimating of system parameters. Because the components of a pathway interact not only with each other in the same pathway but also with those in different pathways, most (if not all) of mathematical models of metabolic systems are highly complex and nonlinear. The widely used approaches for modeling inter- and intracellular dynamic processes are based on mass action law [1–4]. By mass action law, the reaction rates are generally polynomials in concentrations of metabolites with reaction constants or rational functions which are a fraction and whose denominator and numerators are polynomials in concentrations of metabolites with reaction constants [1–4]. As a result, the mathematical model is nonlinear not only in the states but also in the parameters. Estimation of these parameters is crucial to construct a whole metabolic system [5–7].

In general, all algorithms for nonlinear parameter estimation can be used to estimate parameters in metabolic systems, for example, Gauss-Newton iteration method, and its variants such as Box-Kanemasu interpolation method, Levenberg damped least squares methods and Marquardt's

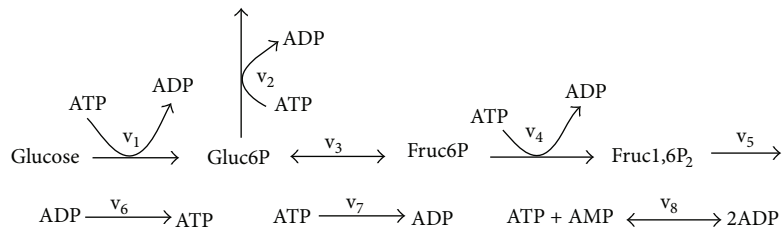


Figure 1: Schematic representation of the upper part of glycolysis [4].

method [8, 9]. However, these iteration methods are initial-sensitive. Another main shortcoming is that these methods may converge to the local minimum of the least squares cost function and thus cannot find the real values of parameters. Furthermore, because of their highly complexity and nonlinearity, Gauss-Newton iteration method and its variants cannot efficiently and accurately estimate the parameters in metabolic systems [5–7, 10, 11].

In this paper, we propose a systematic method for estimating parameters in dynamic metabolic systems. Typically mathematical model of dynamic metabolic systems consists of a group of nonlinear differential equations, some of which contains several rational functions in which parameters are nonlinear. In Section 2, we propose a method for model complexity analysis via the stoichiometric matrix. As a result, we obtain a group of equations, each of which contains only one-rational function plus polynomial functions which we called an improper rational function. Then, based on the observation that in the improper rational functions both the denominator and numerator are linear in parameters while polynomials are also linear in parameters, we develop an iterative linear least squares method for estimating parameters in dynamic metabolic systems in Section 3. The basic idea is to transfer optimizing a nonlinear least squares objective function into iteratively solving a sequence of linear least squares problems. In Section 4, we apply our developed method to estimate parameters in a metabolism system. Finally we give conclusions and some directions of future work along with this study in Section 5.

2. Model Complexity Analysis for Parameter Estimation

A dynamic metabolic system consists of k substances (molecules), and m reactions can be described by a system of differential equations as follows:

$$\frac{dx_i}{dt} = \sum_{j=1}^{m} c_{ij} r_j, \quad \text{for } i = 1, \ldots, k, \tag{1}$$

where x_i represents the concentrations of molecule i, r_j represents the reaction rate j, and c_{ij} represents the stoichiometric coefficient of molecule i in reaction j. The mass action law in biochemical kinetics [2–4, 12] states that the reaction rate is proportional to the probability of a collision of the reactants. This probability is in turn proportional to the concentration of reactants. Therefore, reaction rate r_j is a function of the concentrations of molecules involved in reaction j and proportion constants.

The stoichiometric coefficient c_{ij} assigned to molecule i and reaction j can be put into a so-called stoichiometric matrix $\mathbf{C} = [c_{ij}]_{k\times m}$. Let $X = [x_1, x_2, \ldots, x_k]^T$ and $\mathbf{r} = [r_1, r_2, \ldots, r_m]^T$, and let $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]^T$ represent the vector consisting of all independent proportion constants, and then (1) can be rewritten in the following vector-matrix format:

$$\frac{dX}{dt} = \mathbf{C}\mathbf{r}\left(X, \boldsymbol{\beta}\right). \tag{2}$$

In principle, the stoichiometric coefficient c_{ij} in matrix $\mathbf{C}$ is a constant integer and can be decided according to how molecule i is involved in reaction j. According to mass action law, the expression of reaction rates can be determined to be polynomials or rational functions with reaction constants [2–4, 12]. The challenge to build up the mathematic model of dynamic metabolic system (2) is to estimate the parameter vector $\boldsymbol{\beta}$, especially when some reaction rates are in the form of rational functions in which parameters are nonlinear.

If each differential equation in (2) contains one-rational function without or with polynomial functions, the parameters in model (2) can be estimated by algorithms in [13, 14] or a new algorithm proposed in the next section of this paper. Unfortunately, each differential equation contains a linear combination of several rational functions, which makes the parameter estimation in those coupled differential equations more difficult. The stoichiometric matrix contains very important information about the structure of the metabolic systems and is widely used to analyze the steady state and flux balance of metabolic systems [2–4]. In this paper, via the stoichiometric matrix, we propose a systematic method to transfer a system of differential equations (2) into another system of differential equations, in which each differential equation contains at most one-rational function.

Running Example. To illustrate the proposed method, we use the upper part of glycolysis system as a running example, showing how the method is applied to this system step after step. The schematic representation of this system is shown in Figure 1. The model for this metabolic system is described by the system of differential equations (2) as follows:

$$\begin{aligned}
\frac{d}{dt}\text{Gluc6P} &= r_1 - r_2 - r_3,\\
\frac{d}{dt}\text{Fruc6P} &= r_3 - r_4,\\
\frac{d}{dt}\text{Fruc1,6P}_2 &= r_4 - r_5,
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt}\mathrm{ATP} &= -r_1 - r_2 - r_4 + r_6 - r_7 - r_8,\\
\frac{d}{dt}\mathrm{ADP} &= r_1 + r_2 + r_4 - r_6 + r_7 + 2r_8,\\
\frac{d}{dt}\mathrm{AMP} &= -r_8.
\end{aligned} \tag{3}$$

Based on the mass action law, the individual reaction rates can be expressed as

$$\begin{aligned}
r_1 &= \frac{V_{\max,2}\mathrm{ATP}(t)}{K_{\mathrm{ATP},1} + \mathrm{ATP}(t)},\\
r_2 &= k_2\mathrm{ATP}(t)\cdot\mathrm{Gluc6P}(t),\\
r_3 &= \left(\frac{V^f_{\max,3}}{K_{\mathrm{Gluc6P},3}}\mathrm{Gluc6P}(t) - \frac{V^r_{\max,3}}{K_{\mathrm{Fruc6P},3}}\mathrm{Fruc6P}(t)\right)\\
&\quad\times\left(1 + \left(\frac{\mathrm{Gluc6P}(t)}{K_{\mathrm{Gluc6P},3}}\right) + \frac{\mathrm{Fruc6P}(t)}{K_{\mathrm{Fruc6P},3}}\right)^{-1},\\
r_4 &= \frac{V_{\max,4}(\mathrm{Fruc6P}(t))^2}{K_{\mathrm{Fruc6P},4}\left(1 + \kappa(\mathrm{ATP}(t)/\mathrm{AMP}(t))^2\right) + (\mathrm{Fruc6P}(t))^2},\\
r_5 &= k_5\mathrm{Fruc1,6P_2}(t),\\
r_6 &= k_6\mathrm{ADP}(t),\\
r_7 &= k_7\mathrm{ATP}(t),\\
r_8 &= k_{8f}\mathrm{ATP}(t)\cdot\mathrm{AMP}(t) - k_{8r}(\mathrm{ADP}(t))^2.
\end{aligned} \tag{4}$$

Model (3) has six ordinary differential equations (ODEs) and 15 parameters contained in eight reaction rates, three out of which are rational functions. Some ODEs contain more than one rational reaction rates, which makes the parameter more difficult.

Comparing (3) to (2) we have the state vector: $\mathbf{X}$ = [Gluc6P; Fruc6P; Fruc1,6P$_2$; ATP, ADP, AMP] and stoichiometric matrix:

$$\mathbf{C} = \begin{bmatrix}
1 & -1 & -1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0\\
-1 & -1 & 0 & -1 & 0 & 1 & -1 & -1\\
1 & 1 & 0 & 1 & 0 & -1 & 1 & 2\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1
\end{bmatrix}. \tag{5}$$

In the following, we describe our proposed method to analyze the complexity of model (2) through the running example.

Step 1. Collect the columns in the stoichiometric matrix corresponding to the rational reaction rates in model (2) to construct a submatrix $\mathbf{C}_r$ and collect other columns (corresponding to polynomial reaction rates) to construct a submatrix $\mathbf{C}_p$. Therefore, we have

$$\frac{dX}{dt} = \mathbf{C}\mathbf{r}(X,\boldsymbol{\beta}) = \mathbf{C}_r\mathbf{r}_r(X,\boldsymbol{\beta}) + \mathbf{C}_p\mathbf{r}_p(X,\boldsymbol{\beta}), \tag{6}$$

where $\mathbf{r}_r$ is the subvector of $\mathbf{r}$ and consists of all rational reaction rates while $\mathbf{r}_p$ is another subvector of $\mathbf{r}$ and consists of all polynomial reaction rates. In this step, we should make sure that the rank of matrix $\mathbf{C}_r$ equals the number of rational reaction rates. If the rank of matrix $\mathbf{C}_r$ does not equal the number of rational reaction rates, it means that some rational reaction rates are not independent. Then we combine dependent rational reaction rates together to create a new reaction rate such that all resulted rational reaction rates should be linearly independent [14]. As a result, the rank of matrix $\mathbf{C}_r$ will equal the number of rational reaction rates.

For the running example, we have

$$\begin{aligned}
\mathbf{C}_r &= [c_1, c_3, c_4] = \begin{bmatrix}
1 & -1 & 0\\
0 & 1 & -1\\
0 & 0 & 1\\
-1 & 0 & -1\\
1 & 0 & 1\\
0 & 0 & 0
\end{bmatrix},\\
\mathbf{C}_p &= [c_2, c_5, c_6, c_7, c_8] = \begin{bmatrix}
-1 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0\\
0 & -1 & 0 & 0 & 0\\
-1 & 0 & 1 & -1 & -1\\
1 & 0 & -1 & 1 & 2\\
0 & 0 & 0 & 0 & -1
\end{bmatrix},
\end{aligned} \tag{7}$$

and $\mathbf{r}_r = [r_1, r_3, r_4]$ and $\mathbf{r}_p = [r_2, r_5, r_6, r_7, r_8]$. The rank of matrix $\mathbf{C}_r$ equals 3, which is the number of rational reaction rates.

Step 2. Calculate the left inverse matrix of $\mathbf{C}_r$. That is, calculate $\mathbf{C}_r^-$ such that

$$\mathbf{C}_r^-\mathbf{C}_r = I. \tag{8}$$

As matrix $\mathbf{C}_r$ has the column full rank, matrix $\mathbf{C}_r^-$ satisfying (8) exists although it is typically not unique. For a given matrix $\mathbf{C}_r$, $\mathbf{C}_r^-$ can be easily found by solving (8) which is a linear algebraic system. If it is not unique, any matrix satisfying (8) works for our proposed method.

For the running example, we can have

$$\mathbf{C}_r^- = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0\\
0 & 1 & 1 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0
\end{bmatrix}. \tag{9}$$

Step 3. Multiply (6) by matrix $\mathbf{C}_r^-$ from the left to obtain

$$\begin{aligned}
\mathbf{C}_r^-\frac{dX}{dt} &= \mathbf{C}_r^-\mathbf{C}_r\mathbf{r}_r(X,\boldsymbol{\beta}) + \mathbf{C}_r^-\mathbf{C}_p\mathbf{r}_p(X,\boldsymbol{\beta})\\
&= \mathbf{r}_r(X,\boldsymbol{\beta}) + \mathbf{C}_r^-\mathbf{C}_p\mathbf{r}_p(X,\boldsymbol{\beta})
\end{aligned} \tag{10}$$

or

$$\mathbf{r}_r(X,\boldsymbol{\beta}) + \mathbf{C}_r^- \mathbf{C}_p \mathbf{r}_p(X,\boldsymbol{\beta}) = \mathbf{C}_r^- \frac{dX}{dt}. \tag{11}$$

From its expression, each differential equation in the system (11) contains only one-rational reaction rates plus a linear combination of polynomial reaction rates.

For the running example, we have

$$\begin{aligned} r_1 - r_2 - r_5 &= \frac{d}{dt}\left(\text{Gluc6P} + \text{Fruc6P} + \text{Fruc1,6P}_2\right), \\ r_3 - r_5 &= \frac{d}{dt}\left(\text{Fruc6P} + \text{Fruc1,6P}_2\right), \\ r_4 - r_5 &= \frac{d}{dt}\text{Fruc1,6}P_2. \end{aligned} \tag{12}$$

Step 4. Calculate matrix $\mathbf{C}_r^{\perp}$ such that

$$\mathbf{C}_r^{\perp}\mathbf{C}_r = \mathbf{0}, \tag{13}$$

where $\mathbf{C}_r^{\perp}$ has the full row rank and $\text{rank}(\mathbf{C}_r^{\perp}) + \text{rank}(\mathbf{C}_r^-) =$ the number of rows in $\mathbf{C}_r$. Note that $\mathbf{C}_r^{\perp}$ can be easily found by solving (13), which is a homogenous linear algebraic system. Again if it is not unique, any matrix satisfying (13) works for our proposed method.

Then multiply (6) by matrix $\mathbf{C}_r^{\perp}$ from the left to obtain

$$\mathbf{C}_r^{\perp}\frac{dX}{dt} = \mathbf{C}_r^{\perp}\mathbf{C}_r\mathbf{r}_r(X,\boldsymbol{\beta}) + \mathbf{C}_r^{\perp}\mathbf{C}_p\mathbf{r}_p(X,\boldsymbol{\beta}) = \mathbf{C}_r^{\perp}\mathbf{C}_p\mathbf{r}_p(X,\boldsymbol{\beta}) \tag{14}$$

or

$$\mathbf{C}_r^{\perp}\mathbf{C}_p\mathbf{r}_p(X,\boldsymbol{\beta}) = \mathbf{C}_r^{\perp}\frac{dX}{dt}. \tag{15}$$

For the running example, we can have

$$\begin{aligned} \mathbf{C}_r^{\perp} &= \begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \\ \mathbf{C}_r^{\perp}\mathbf{C}_p &= \begin{bmatrix} -2 & -2 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}. \end{aligned} \tag{16}$$

Step 5. Let $D = \mathbf{C}_r^{\perp}\mathbf{C}_p$. If $\text{rank}(D) \geq$ the number of columns, then solving (15) yields

$$\mathbf{r}_p(X,\boldsymbol{\beta}) = \left(D^T D\right)^{-1} D^T \mathbf{C}_r^{\perp}\frac{dX}{dt}. \tag{17}$$

If $\text{rank}(D)$ < the number of columns, it means that some polynomial reaction rates in (15) are linearly dependent. Then combine the linearly dependent rates and construct a new reaction rate vector $\overline{\mathbf{r}}_p(X,\boldsymbol{\beta})$ and full column rank matrix $\overline{D}$ such that

$$\overline{D}\overline{\mathbf{r}}_p(X,\boldsymbol{\beta}) = D\mathbf{r}_p(X,\boldsymbol{\beta}) = \mathbf{C}_r^{\perp}\mathbf{C}_p\mathbf{r}_p(X,\boldsymbol{\beta}) = \mathbf{C}_r^{\perp}\frac{dX}{dt}, \tag{18}$$

and then solving (18) yields

$$\overline{\mathbf{r}}_p(X,\boldsymbol{\beta}) = \left(\overline{D}^T\overline{D}\right)\overline{D}^T\mathbf{C}_r^{\perp}\frac{dX}{dt}. \tag{19}$$

For the running example, we have $\text{rank}(D)$ < the number of columns. As the first four columns are linearly dependent, we can have a new reaction rates $-2r_2 - 2r_5 + r_6 - r_7$. Therefore, we have

$$\overline{D} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}, \qquad \overline{D}^T\mathbf{C}_r^{\perp} = \begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 \\ -1 & -1 & -2 & 0 & 1 & -1 \end{bmatrix}, \tag{20}$$

and furthermore, noting that $(d/dt)(\text{ATP} + \text{ADP} + \text{AMP}) = 0$, from (19) we have

$$\begin{aligned} r_6 - r_7 - 2r_2 - 2r_5 &= \frac{d}{dt}\left(\text{Gluc6P} + \text{Fruc6P} + 2\text{Fruc1,6P}_2 + \text{ATP} - \text{AMP}\right), \\ r_8 &= -\frac{d}{dt}\text{AMP}. \end{aligned} \tag{21}$$

After these five steps, dynamic metabolic system (2) is transferred into a system of differential equations, in which each differential equation contains one-rational function plus polynomial functions ((11) or (12)) or only polynomial function ((19) or (21)). Parameters in (19) can be analytically estimated by well-known least squares methods. In the next section, we describe an algorithm to estimate parameters in (11).

3. Parameter Estimation Algorithm

After its complexity analysis, estimating parameters in dynamic metabolic system is reduced to mainly estimating parameters in a rational function plus polynomial, which we call the improper rational function. These functions are nonlinear in both parameters and state variables. Therefore, estimation of parameters in these models is a nonlinear estimation problem. In general, all algorithms for nonlinear parameter estimation can be used to estimate parameters in the improper rational functions, for example, Gauss-Newton iteration method and its variants such as Box-Kanemasu interpolation method, Levenberg damped least squares methods, Marquardt's method [9–12, 15], and more sophisticated methods [16]. However, these iteration methods are initial sensitive. Another main shortcoming is that most of these methods may converge to the local minimum of the least squares cost function and thus cannot find the real values of parameters. In the following, we describe an iterative linear least squares method to estimate parameters in the improper rational functions. The basic idea is to transfer optimizing a nonlinear least squares objective function into iteratively solving a sequence of linear least squares problems.

Consider the general form of the following improper rational functions:

$$\eta(\mathbf{X},\boldsymbol{\beta}) = \frac{N_0(\mathbf{X}) + \sum_{i=1}^{p_N} N_i(\mathbf{X})\beta_{N_i}}{D_0(\mathbf{X}) + \sum_{j=1}^{p_D} D_j(\mathbf{X})\beta_{D_j}} + \sum_{k=1}^{p_P} P_k(\mathbf{X})\beta_{P_k}, \tag{22}$$

where the vector $\mathbf{X}$ consists of the state variables and the p-dimensional vector $\boldsymbol{\beta}$ consists of all parameters in the improper rational function (22), which can naturally be divided into three groups: those in the numerator of the rational functions β_{N_i} $(i = 1, \ldots, p_N)$, those in the denominator of the rational function β_{D_j} $(j = 1, \ldots, p_D)$, and those in the polynomial β_{P_k} $(k = 1, \ldots, p_P)$, where we have that $p_D + p_N + p_P = p$. $N_i(\mathbf{X})$ $(i = 0, 1, \ldots, p_N)$, $D_j(\mathbf{X})$ $(j = 0, 1, \ldots, p_D)$, and $P_k(\mathbf{X})$ $(k = 1, \ldots, p_P)$ are the known functions nonlinear in the state variable $\mathbf{X}$ and do not contain any unknown parameters. Either $N_0(\mathbf{X})$ or $D_0(\mathbf{X})$ must be nonzero, and otherwise from sensitivity analysis [9, 16] the parameters in model (22) cannot be uniquely identified.

If there is no polynomial part, model (22) is reduced to a rational function. Recently, several methods have been proposed for estimating parameters in rational functions [5, 6, 13, 14]. The authors in [5, 6] have employed general nonlinear parameter estimation methods to estimate parameters in rational functions. As shown in their results, the estimation error is fairly large. We have observed that in rational functions both the denominator and numerator are linear in the parameters. Based on this observation, we have developed iterative linear least squares methods in [13, 14] for estimating parameters in rational functions. Mathematically, improper rational function (22) can be rewritten as the following rational function:

$$\eta(\mathbf{X}, \boldsymbol{\beta}) = \left(N_0(\mathbf{X}) + \sum_{i=1}^{p_N} N_i(\mathbf{X}) \beta_{N_i} + \left(\sum_{k=1}^{p_P} P_k(\mathbf{X}) \beta_{P_k} \right) \times \left(D_0(\mathbf{X}) + \sum_{j=1}^{p_D} D_j(\mathbf{X}) \beta_{D_j} \right) \right) \times \left(D_0(\mathbf{X}) + \sum_{j=1}^{p_D} D_j(\mathbf{X}) \beta_{D_j} \right)^{-1}. \tag{23}$$

However, in the numerator of the model above, there are $p_D p_P + p_N + p_P$ coefficients while there are $p_D + p_N + p_P$ unknown parameters. When $p_P = 1$, the number of parameters is equal to the numbers of coefficients, and the methods developed in [13, 14] can be applied. However, when $p_P > 1$, those methods are not applicable as the number of parameters is less than the number of coefficients in the numerator.

In order to describe an algorithm to estimate parameters in the improper rational function (22) for n given groups of observation data y_t and $\mathbf{X}_t$ $(t = 1, 2, \ldots, n)$, we introduce the following notation:

$$\begin{aligned}
\boldsymbol{\beta}_N &= \left[\beta_{N1}, \beta_{N2}, \ldots, \beta_{Np_N}\right]^T \in R^{p_N}, \\
\boldsymbol{\beta}_D &= \left[\beta_{D1}, \beta_{D2}, \ldots, \beta_{Dp_D}\right]^T \in R^{p_D}, \\
\boldsymbol{\beta}_P &= \left[\beta_{P1}, \beta_{P2}, \ldots, \beta_{Pp_D}\right]^T \in R^{p_P}, \\
\boldsymbol{\beta} &= \left[\begin{array}{ccc} \boldsymbol{\beta}_P^T & \boldsymbol{\beta}_N^T & \boldsymbol{\beta}_D^T \end{array}\right]^T,
\end{aligned}$$

$$\begin{aligned}
\varphi_N(\mathbf{X}_t) &= \left[N_1(\mathbf{X}_t), N_2(\mathbf{X}_t), \ldots, N_{p_N}(\mathbf{X}_t)\right] \in R^{p_N}, \\
\varphi_D(\mathbf{X}_t) &= \left[D_1(\mathbf{X}_t), D_2(\mathbf{X}_t), \ldots, D_{p_D}(\mathbf{X}_t)\right] \in R^{p_D}, \\
\varphi_P(\mathbf{X}_t) &= \left[P_1(\mathbf{X}_t), P_2(\mathbf{X}_t), \ldots, P_{p_P}(\mathbf{X}_t)\right] \in R^{p_P}, \\
\mathbf{Y} &= \left[y(1), y(2), \ldots, y(n)\right]^T \in R^n, \\
\Phi_{N_0} &= \left[N_0(\mathbf{X}_1), N_0(\mathbf{X}_2), \ldots, N_0(\mathbf{X}_n)\right]^T \in R^n, \\
\Phi_{D_0} &= \left[D_0(\mathbf{X}_1), D_0(\mathbf{X}_2), \ldots, D_0(\mathbf{X}_n)\right]^T \in R^n, \\
\Phi_N &= \begin{bmatrix} \varphi_N(\mathbf{X}_1) \\ \varphi_N(\mathbf{X}_2) \\ \vdots \\ \varphi_N(\mathbf{X}_n) \end{bmatrix} \in R^{n \times p_N}, \\
\Phi_D &= \begin{bmatrix} \varphi_D(\mathbf{X}_1) \\ \varphi_D(\mathbf{X}_2) \\ \vdots \\ \varphi_D(\mathbf{X}_n) \end{bmatrix} \in R^{n \times p_D}, \\
\Phi_P &= \begin{bmatrix} \varphi_P(\mathbf{X}_1) \\ \varphi_P(\mathbf{X}_2) \\ \vdots \\ \varphi_P(\mathbf{X}_n) \end{bmatrix} \in R^{n \times p_P}, \\
\Psi(\boldsymbol{\beta}_D) &= \operatorname{diag} \begin{bmatrix} D_0(\mathbf{X}_1) + \varphi_D(\mathbf{X}_1) \boldsymbol{\beta}_D \\ D_0(\mathbf{X}_2) + \varphi_D(\mathbf{X}_2) \boldsymbol{\beta}_D \\ \vdots \\ D_0(\mathbf{X}_n) + \varphi_D(\mathbf{X}_n) \boldsymbol{\beta}_D \end{bmatrix} \in R^{n \times n}.
\end{aligned} \tag{24}$$

To estimate parameters in the improper rational function (22), as in [11], we form a sum of the weighted squared errors (the cost function) with the notions above as follows:

$$\begin{aligned}
J(\boldsymbol{\beta}) &= J(\boldsymbol{\beta}_P, \boldsymbol{\beta}_N, \boldsymbol{\beta}_D) \\
&= \sum \left(D_0(\mathbf{X}_t) + \varphi_D(\mathbf{X}_t) \boldsymbol{\beta}_D\right)^2 \\
&\quad \times \left(\frac{N_0(\mathbf{X}_t) + \varphi_N(\mathbf{X}_t) \boldsymbol{\beta}_N}{D_0(\mathbf{X}_t) + \varphi_D(\mathbf{X}_t) \boldsymbol{\beta}_D} + \Phi_P \boldsymbol{\beta}_P - y_t \right)^2.
\end{aligned} \tag{25}$$

Minimizing $J(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta} = [\boldsymbol{\beta}_P^T, \boldsymbol{\beta}_N^T, \boldsymbol{\beta}_D^T]^T$ can give the nonlinear least squares estimation of parameters $\boldsymbol{\beta}_P$, $\boldsymbol{\beta}_N$, and $\boldsymbol{\beta}_D$. We rewrite the objective function (22) as follows:

$$\begin{aligned}
J(\boldsymbol{\beta}) = \sum \big[&\left(D_0(\mathbf{X}_t) + \varphi_D(\mathbf{X}_t) \boldsymbol{\beta}_D\right) \Phi_P \boldsymbol{\beta}_P + \varphi_N(\mathbf{X}_t) \boldsymbol{\beta}_N \\
&- \varphi_D(\mathbf{X}_t) y_t \boldsymbol{\beta}_D - D_0(\mathbf{X}_t) y_t + N_0(\mathbf{X}_t) \big]^2.
\end{aligned} \tag{26}$$

Table 1: The true value (from [4]), estimated value, and relative estimation errors.

Parameter name	True value	Estimated value	REE (%)
$V_{max,2}$ (mM·min^{-1})	50.2747	50.2447	0.0001
$K_{ATP,1}$ (mM)	0.10	0.10000	0.0399
k_2 (mM^{-1}·min^{-1})	2.26	2.2599	0.0049
$V^f_{max,3}$ (mM·min^{-1})	140.282	139.4917	0.5633
$V^r_{max,3}$ (mM·min^{-1})	140.282	141.3623	0.7701
$K_{Gluc6P,3}$ (mM)	0.80	0.7999	1.3884
$K_{Fruc6P,3}$ (mM)	0.15	0.1499	0.0930
$V_{max,4}$ (mM·min^{-1})	44.7287	44.6664	0.1372
$K_{Fruc6P,4}$ (mM^2)	0.021	0.0206	1.8457
k	0.15	0.1526	1.7447
k_5 (min^{-1})	6.04662	6.0466	0.0007
k_6 (min^{-1})	68.48	68.4837	0.0054
k_7 (min^{-1})	3.21	3.20797	0.0078
k_{8f} (min^{-1})	432.9	432.8408	0.0137
k_{8r} (min^{-1})	133.33	133.314	0.0120

In the objective function (26), for a given parameters $\overline{\boldsymbol{\beta}}_D$ in the first term, we have

$$J(\boldsymbol{\beta}) = J\left(\boldsymbol{\beta}_P, \boldsymbol{\beta}_N, \boldsymbol{\beta}_D, \overline{\boldsymbol{\beta}}_D\right) = \left[\mathbf{A}\left(\overline{\boldsymbol{\beta}}_D\right)\boldsymbol{\beta} - \mathbf{b}\right]^T \left[\mathbf{A}\left(\overline{\boldsymbol{\beta}}_D\right)\boldsymbol{\beta} - \mathbf{b}\right], \tag{27}$$

where

$$A\left(\overline{\boldsymbol{\beta}}_D\right) = \begin{bmatrix} \Psi\left(\overline{\boldsymbol{\beta}}_D\right)\Phi_P^T \\ \Phi_N^T \\ -\operatorname{diag}(Y)\,\Phi_D^T \end{bmatrix} \in R^{n\times p}, \tag{28}$$

$$\mathbf{b} = \left(\Phi_{D_0}\operatorname{diag}(Y) - \Phi_{N_0}\right) \in R^n. \tag{29}$$

Then for given parameters $\overline{\boldsymbol{\beta}}_D$, we can estimate the parameters $\boldsymbol{\beta} = [\boldsymbol{\beta}_P^T, \boldsymbol{\beta}_N^T, \boldsymbol{\beta}_D^T]^T$ by linear least squares method as follows:

$$\boldsymbol{\beta} = \left[\mathbf{A}^T\left(\overline{\boldsymbol{\beta}}_D\right)\mathbf{A}\left(\overline{\boldsymbol{\beta}}_D\right)\right]^{-1}\mathbf{A}^T\left(\overline{\boldsymbol{\beta}}_D\right)\mathbf{b}. \tag{30}$$

Based on the above discussion, we propose the following iterative linear least squares method.

Step 1. Choose the initial guess for $\boldsymbol{\beta}_D^0$.

Step 2. Iteratively construct matrix $\mathbf{A}(\boldsymbol{\beta}_D^s)$ and vector $\mathbf{b}$ by (28) and (29), respectively, and then solve the linear least squares problem:

$$J\left(\boldsymbol{\beta}^{s+1}\right) = \left[\mathbf{A}\left(\boldsymbol{\beta}_D^s\right)\boldsymbol{\beta}^{s+1} - \mathbf{b}\right]^T \left[\mathbf{A}\left(\boldsymbol{\beta}_D^s\right)\boldsymbol{\beta}^{s+1} - \mathbf{b}\right], \tag{31}$$

which gives the solution

$$\boldsymbol{\beta}^{s+1} = \left[\mathbf{A}^T\left(\boldsymbol{\beta}_D^s\right)\mathbf{A}\left(\boldsymbol{\beta}_D^s\right)\right]^{-1}\mathbf{A}^T\left(\boldsymbol{\beta}_D^s\right)\mathbf{b} \tag{32}$$

until the stopping criterion is met, where $\boldsymbol{\beta}^s = [\boldsymbol{\beta}_P^{sT}, \boldsymbol{\beta}_N^{sT}, \boldsymbol{\beta}_D^{sT}]^T$ is the estimation of parameters $\boldsymbol{\beta}$ at step s.

From (31), if the estimation sequence $\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \ldots$ is converged to $\boldsymbol{\beta}^*$, the objective function (26) reaches its minimum value at $\boldsymbol{\beta}^*$. That is, $\boldsymbol{\beta}^*$ is the estimation of parameters in model (22).

There are several ways to set up a stopping criterion. In this paper the stopping criteria are chosen as

$$\frac{\left\|\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1}\right\|}{\left\|\boldsymbol{\beta}^{k-1}\right\| + 1} \le \varepsilon, \tag{33}$$

where $\|\cdot\|$ is the Euclidean norm of the vector and ε is a preset small positive number, for example, 10^{-5}.

4. Application

To investigate the method developed in previous sections, this study generates artificial data from the dynamic metabolic system in the running example with the biochemically plausible parameter values [4] listed in column 2 of Table 1 and initial values: Gluc6P(0) = 1 mM, Fruc6P(0) = 0 mM, Fruc1,6P$_2$(0) = 0 mM, ATP(0) = 2.1 mM, ADP(0) = 1.4 mM, and AMP (0) = 0.1 mM. The trajectory of this system is depicted in Figure 2. From Figure 2, the concentrations of all molecules except for Frucose-1,6-biphosphate reach their its steady states after about 0.1 minutes while Frucose-1,6-biphosphate after 0.5 minutes. Therefore, we do not use the data simulated after 0.5 minutes.

Although no noise is added to the artificial data in the simulation, noises are introduced in numerically calculating the derivatives by finite difference formulas. In general, the higher the sampling frequency and more data points are used, the more accurate the numerical derivatives are. On the other hand, we may not obtain data with the high frequency because of experimental limitations in practice. In this study, the sampling frequency is 100 data points per minute. In numerically calculating the concentration change rate at each

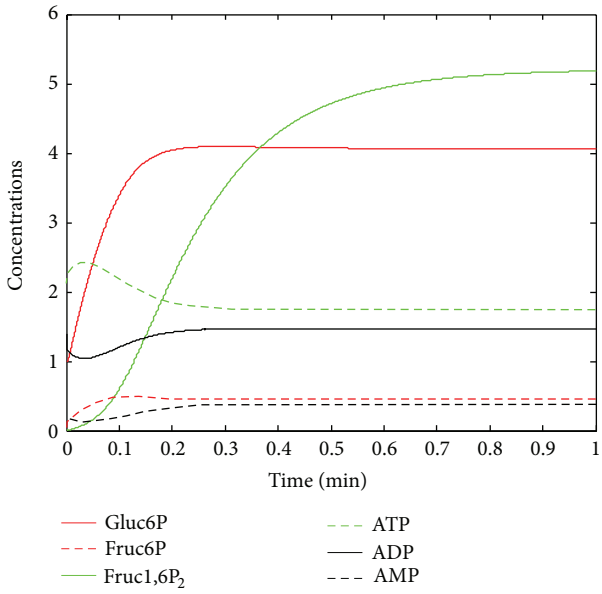


Figure 2: Trajectory of system (3).

time point from concentration x, we adopt the five-point central finite difference formula as follows:

$$\dot{x}(t_n) = \frac{1}{12\Delta t}\left[x(t_{n-2}) - 8x(t_{n-1}) + 8x(t_{n+1}) - x(t_{n+2})\right]. \tag{34}$$

The estimation accuracy of the proposed method is investigated in terms of relative estimation error which is defined as

$$\text{REE} = \frac{\|\text{estimate\_value} - \text{true\_value}\|}{\|\text{true\_value}\|}. \tag{35}$$

As all parameters to be estimated are nonnegative, initial values are chosen as 0 or 1 in this study. The experimental results are listed in columns 3 and 4 in Table 1. From column 3 in Table 1, the estimated parameter values are very close to the corresponding true values. Actually the relative estimation errors calculated from (29) for all estimated parameters except for two are less than 1%. This indicates that the proposed method can accurately estimate the parameters in this system.

5. Conclusions and Future Work

In this study, we have first described a method to analyze the complexity of metabolic systems for parameter estimation, based on the stoichiometric matrix of the metabolic systems. As a result, the estimation of parameters in the metabolic systems has been reduced to the estimation of parameters in the improper rational functions or polynomial functions. Then we have developed an iterative linear least squares method for estimating parameters in the improper rational models. The results from its application to a metabolism system have shown that the proposed method can accurately estimate the parameters in metabolic systems.

We do not consider the noises in the data except those introduced by numerical derivatives in this study. One direction of future work is to investigate the influence of noises in the data to the estimation accuracy. In addition, low sampling frequency is expected, particularly for molecular biological systems as in practice measurements from them may be very expensive or it is impossible to sample measurements with high frequencies. Another direction of future work is to improve the estimation accuracy of the proposed method with low sampling frequencies.

Acknowledgments

This work was supported by the Special Fund of Ministry of Education of Beijing for Distinguishing Professors and Science and Technology Funds of Beijing Ministry of Education (SQKM201210037001) to Li-Ping Tian, by National Natural Science Foundation of China (NSFC 61134004) to Zhong-Ke Shi, and by Natural Sciences and Engineering Research Council of Canada (NSERC) to Fang-Xiang Wu.

References

[1] M. Fussenegger, J. E. Bailey, and J. Varner, "A mathematical model of caspase function in apoptosis," *Nature Biotechnology*, vol. 18, no. 7, pp. 768–774, 2000.

[2] J. Nielsen, J. Villadsen, and G. Liden, *Bioreaction Engineering Principles*, Kluwer Academic Publishers, New York, NY, USA, 2nd edition, 2003.

[3] G. N. Stephanopoulos, A. A. Aritidou, and J. Nielsen, *Metabolic Engineering: Principles and Methodologies*, Academic Press, San Diego, Calif, USA, 1998.

[4] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach, *Systems Biology in Practice: Concepts, Implementation and Application*, Wiley-VCH and KGaA, Weinheim, Germany, 2005.

[5] K. G. Gadkar, J. Varner, and F. J. Doyle III, "Model identification of signal transduction networks from data using a state regulator problem," *Systems Biology*, vol. 2, no. 1, pp. 17–29, 2005.

[6] K. G. Gadkar, R. Gunawan, and F. J. Doyle III, "Iterative approach to model identification of biological networks," *BMC Bioinformatics*, vol. 6, article 155, 2005.

[7] I.-C. Chou and E. O. Voit, "Recent developments in parameter estimation and structure identification of biochemical and genomic systems," *Mathematical Biosciences*, vol. 219, no. 2, pp. 57–83, 2009.

[8] J. V. Beck and K. J. Arnold, *Parameter Estimation in Engineering and Science*, John Wiley & Sons, New York, NY, USA, 1977.

[9] A. van den Bos, *Parameter Estimation for Scientists and Engineers*, John Wiley & Sons, Hoboken, NJ, USA, 2007.

[10] P. Mendes and D. B. Kell, "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation," *Bioinformatics*, vol. 14, no. 10, pp. 869–883, 1998.

[11] C. G. Moles, P. Mendes, and J. R. Banga, "Parameter estimation in biochemical pathways: a comparison of global optimization methods," *Genome Research*, vol. 13, no. 11, pp. 2467–2474, 2003.

[12] E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, H. Lehracj, and R. Herwing, *Systems Biology: A Textbook*, Wiley-VCH and KGaA, Weiheim, Germany, 2009.

[13] F. X. Wu, L. Mu, and Z. K. Shi, "Estimation of parameters in rational reaction rates of molecular biological systems via weighted least squares," *International Journal of Systems Science*, vol. 41, no. 1, pp. 73–80, 2010.

[14] F. X. Wu, Z. K. Shi, and L. Mu, "Estimating parameters in the caspase activated apoptosis system," *Journal of Biomedical Engineering and Technology*, vol. 4, no. 4, pp. 338–354.

[15] L. Marucci, S. Santini, M. di Bernardo, and D. di Bernardo, "Derivation, identification and validation of a computational model of a novel synthetic regulatory network in yeast," *Journal of Mathematical Biology*, vol. 62, no. 5, pp. 685–706, 2011.

[16] L. Cheng, Z. G. Hou, Y. Lin, M. Tan, W. C. Zhang, and F.-X. Wu, "Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks," *IEEE Transactions on Neural Networks*, vol. 22, no. 5, pp. 714–726, 2011.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 167069, 13 pages
http://dx.doi.org/10.1155/2013/167069

Research Article

Wavelet-Based Artifact Identification and Separation Technique for EEG Signals during Galvanic Vestibular Stimulation

Mani Adib and Edmond Cretu

Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada V6T 1Z4

Correspondence should be addressed to Mani Adib; mani.adib@gmail.com

Received 22 March 2013; Accepted 5 June 2013

Academic Editor: Carlo Cattani

We present a new method for removing artifacts in electroencephalography (EEG) records during Galvanic Vestibular Stimulation (GVS). The main challenge in exploiting GVS is to understand how the stimulus acts as an input to brain. We used EEG to monitor the brain and elicit the GVS reflexes. However, GVS current distribution throughout the scalp generates an artifact on EEG signals. We need to eliminate this artifact to be able to analyze the EEG signals during GVS. We propose a novel method to estimate the contribution of the GVS current in the EEG signals at each electrode by combining time-series regression methods with wavelet decomposition methods. We use wavelet transform to project the recorded EEG signal into various frequency bands and then estimate the GVS current distribution in each frequency band. The proposed method was optimized using simulated signals, and its performance was compared to well-accepted artifact removal methods such as ICA-based methods and adaptive filters. The results show that the proposed method has better performance in removing GVS artifacts, compared to the others. Using the proposed method, a higher signal to artifact ratio of −1.625 dB was achieved, which outperformed other methods such as ICA-based methods, regression methods, and adaptive filters.

1. Introduction

Brain stimulation by means of electrical currents has been employed in neurological studies for therapy purposes for many years [1–5]. However, the ability to analyze the ongoing neural activities during the stimulation is limited due to the artifact generated by GVS. The leakage of the stimulation current through the scalp generates an additional electrical potential with a much higher amplitude than that of the neural activities. As a result, higher artifactual potentials are collected by the EEG electrodes, especially in the neighbourhood of stimulation areas. The stimulation artifacts which are superimposed on the EEG signals are the main obstacle in understanding the effects of the GVS interactions with neural circuitries in different brain regions. Analyzing the EEG signals during GVS stimulation is of high importance, as it provides information on how it affects the neural activities. For instance, in suppressing the symptoms of some neurological disorders using GVS, researchers are interested in eliciting GVS responses in different brain regions. Furthermore, to be able to perform GVS studies in closed-loop mode, where the delivered GVS stimuli are adjusted in response to ongoing neural activities, it is necessary to remove the stimulation artifacts from neural activities signals. An experimentally measured example of EEG signals contaminated with the GVS artifacts is illustrated in Figure 1.

Considering that the frequency spectra of the neural signals and GVS artifacts overlap, filtering the frequency components of GVS artifacts results in the loss of the original neural signals. The four major EEG frequency bands are Delta (the lowest frequency band up to 4 Hz), Theta (4 Hz to 8 Hz), Alpha (8 Hz to 12 Hz), and Beta (12 Hz to 30 Hz). In order to analyze and understand the effect of GVS on EEG patterns, it is essential to be able to remove the artifact signals from the frequency band of interest, before establishing any GVS-EEG interaction models.

There are various methods to remove different types of artifacts, such as myogenic artifacts [6–9], ocular artifacts [10–15], extrinsic artifacts such as MRI induced artifacts in simultaneous EEG/fMRI studies [16], stimulation artifacts

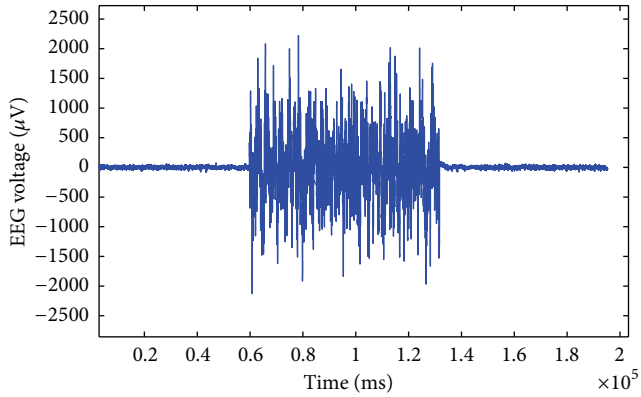


Figure 1: Measured EEG data during 72 seconds of GVS stimulation and 60 seconds before and after applying the GVS.

[17–20], and general artifacts and signals that have noncerebral origin [21, 22]. One of the most commonly used methods to remove artifacts from EEG signals is the Independent Component Analysis (ICA). Generally, in the component-based methods such as ICA, the EEG signals are decomposed into statistically independent and uncorrelated terms; the artifact components are then identified and filtered out, and the EEG signals can be reconstructed from the neural components without artifacts. However, applying ICA to remove the GVS stimulation artifacts is challenging, particularly when we increase the amplitude of the GVS over 1 mA with a signal to artifact ratio less than −35 dB. We will discuss this in more detail later in the section "Comparison of the performance of different artifact removal methods".

We propose a novel method for GVS artifacts removal by combining time-series regression methods and wavelet decomposition methods. To enhance the precision of the artifact estimation using regression models, the models should account for the complex behavior of the GVS interactions in the frequency domain. So we decomposed the recorded EEG and GVS signals into different frequency bands and then used regression models to estimate the GVS artifacts in each frequency band. We used multiresolution wavelet analysis to decompose nonstationary EEG signals in the time-frequency plane. Both the discrete wavelet transform (DWT) and the stationary wavelet transform (SWT) algorithms were employed, and the results were compared. To estimate the GVS current distribution through the scalp using time-series regression methods based on biophysical models, we used and compared the performance of different parametric regression models, such as discrete-time polynomials, nonlinear Hammerstein-Wiener, and state-space models.

In this study, we firstly used simulated data to assess and optimize the performance of the proposed method using various regression models and different wavelet algorithms. The resulting optimized method was then applied to real data. We compared the results of the proposed method and other methods, such as ICA, using both simulated and real data. This paper is organized as follows: Section 2 provides a detailed description of the equipment and set-up, the data simulation, the signal processing methods, and the comparison of their performances. Section 3 shows the results of the proposed artifact removal method, and in Section 4, we discuss the proposed method, its results, and suggested works for the future.

Table 1: EEG channels.

ch1	ch2	ch3	ch4	ch5	ch6	ch7	ch8	ch9	ch10
FP1	FP2	F7	F3	Fz	F4	F8	T7	C3	Cz
ch11	ch12	ch13	ch14	ch15	ch16	ch17	ch18	ch19	ch20
C4	T8	P7	P3	Pz	P4	P8	O1	O2	Ref

2. Materials and Methods

2.1. Equipment and Setup. The EEG recording was carried out with a NeuroScan SynAmps2 system, with 20 electrodes located according to the international 10–20 EEG system (Table 1) and with a sampling frequency set to 1 kHz.

The GVS signal was applied using a Digitimer DS5 isolated bipolar current stimulator. This stimulator can generate a stimulation current with a waveform proportional to the controlling voltage applied to its input. The waveform was generated using LabVIEW and sent to the stimulator through a National Instrument (NI) Data Acquisition (DAQ) board. In this study, we applied a zero-mean pink noise current, with a $1/f$-type power spectrum within a frequency range of 0.1 to 10 Hz and duration of 72 seconds. We kept the amplitude of the delivered stimuli lower than the feeling threshold, in the range of 100 μA to 800 μA, with the root mean square values between 60 μA and 450 μA. The stimulator is equipped with a data acquisition device to record the delivered stimulus, which allows us to make a continuous record of the delivered stimulation current and voltage. We recorded the EEG signals during the stimulation, 60 seconds before and 60 seconds after the stimulation. The EEG data for these experiments were acquired by our collaborator in the Pacific Parkinson's Research Centre. Nine healthy subjects (6 males, 3 females), between the ages of 21 and 53 yr, with no known history of neurological disease or injury, participated in this study. All subjects were asked to relax, remain still, and concentrate on a focal point on the screen in front of them so that less myogenic and ocular artifacts occur. Also, under resting conditions, there are less variations in the head impedance [23], which is important for data acquisition in this study.

2.2. Simulated Data. To quantitatively assess and optimize the performance of the proposed method and compare the accuracy of different methods in removing the GVS artifacts from the EEG recordings, we used simulated data. The simulation study was carried out by combining the clean (artifact free) EEG recordings with the simulated GVS contamination. To simulate the actual process of the GVS contamination, we paid attention to the physical structure of the electrode-skin interface and the electrical impedance of the head between the points that the EEG and the GVS electrodes are placed. As the skull impedance is much higher than scalp impedance [23], we can assume that the GVS current mainly distributes through the scalp. The skin and the electrode-skin interface

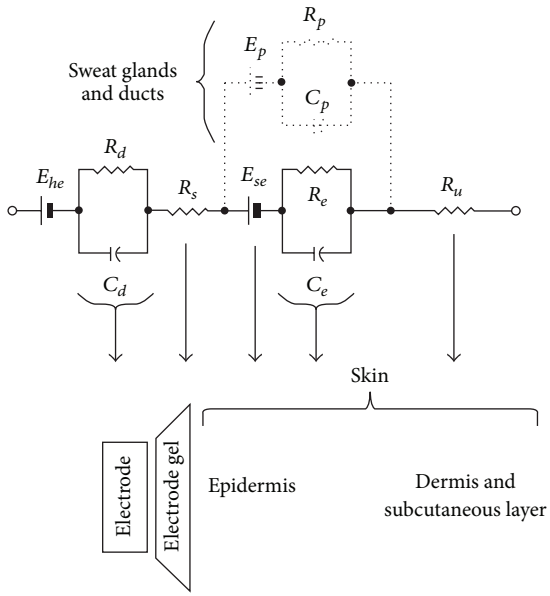


Figure 2: Electrical equivalent circuit for the electrode-skin interface and the underlying skin [24].

can be modeled using a resistive-capacitive circuit [24], as shown in Figure 2.

In this electrical equivalent circuit, E_{he} is the half cell potential of the electrode/gel interface, and the parallel combination of resistive R_d and capacitive C_d components represents the impedance associated with the electrode-gel interface. R_s is the series impedance associated with the resistance of the electrode gel. E_{se} is the potential difference across the epidermis, whose impedance is represented by the resistance R_e and capacitance C_e. In general, the dermis and the subcutaneous layer under it behave as an equivalent pure resistance R_u. The deeper layers of the skin, containing vascular, nervous components and hair follicles, contribute very less to the electrical skin impedance, but sweat glands and ducts add an equivalent parallel RC network (represented by broken lines in Figure 2) and a potential difference between sweat glands, ducts, dermis, and subcutaneous layers [24]. If we neglect the pure resistance of the deeper layers of skin and the resistance of the electrode gel, we can simplify the impedance structure as follows:

$$Z(s) \approx \left(\frac{R_d}{sR_dC_d + 1} + \frac{R_e}{sR_eC_e + 1} \parallel \frac{R_p}{sR_pC_p + 1} \right). \quad (1)$$

This equation can be rewritten as

$$Z(s) \approx \frac{sB_1 + B_0}{s^2A_2 + sA_1 + 1}, \quad (2)$$

where s is the complex frequency variable, A_2, A_1, B_2, and B_1 represent specific combinations of R_d, R_e, R_p, C_d, C_e, and C_p for each electrode. This model-based identification approach suggests the following relation between the injected GVS current and the collected EEG voltage at a given electrode:

$$E_m = X_{\text{in}} \frac{sB_1 + B_0}{s^2A_2 + sA_1 + 1} + E + W_{\text{noise}}, \quad (3)$$

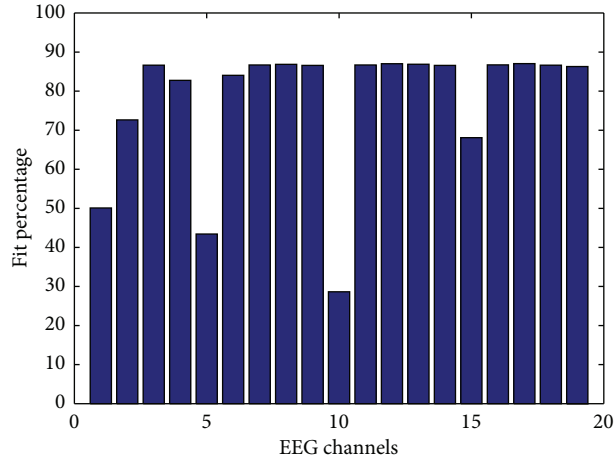


Figure 3: Fit percentage between the simulation output and the measured EEG at each channel.

where E_m is the measured EEG, X_{in} is the injected GVS current, E is the original neural signals or EEG without artifact, and W_{noise} is the measurement noise. We simulated this impedance structure to be able to compute the GVS contribution at each EEG channel output:

$$E_m^* = X_{\text{in}} \frac{sB_1 + B_0}{s^2A_2 + sA_1 + 1}, \quad (4)$$

where E_m^* represents the GVS artifacts in the measured EEG signals. The simulated impedance structure between GVS electrodes and all 19 EEG electrodes was used to calculate the output voltage due to the GVS current (the GVS artifact) at each EEG electrode (Figure 3).

The fit percentage is a measure of the relative energy fraction in the simulated GVS artifact calculated as given by:

$$\text{fit} = 100 \left(1 - \frac{\sum \left(E_m(t) - E_m^*(t) \right)^2}{\left(\sum \left(E_m(t) - \text{mean}(E_m(t)) \right)^2 \right)} \right). \quad (5)$$

The results show that the fitness of simulated GVS artifact is higher at the EEG electrodes which are closer to the GVS electrodes and it is lower at further channels like channel 15 (Pz), channel 10 (Cz), channel 5 (Fz), channel 1 (FP1), and channel 2 (FP2). According to (2), we can assume that the skin impedance model is a low-order, continuous-time transfer function with one zero and two poles. To simulate the skin impedance structure, we used an iterative nonlinear least-squares algorithm to minimize a selected cost function taken as the weighted sum of the squares of the errors. This algorithm has been applied to real measured data, and the parameters of the impedance model were identified for each EEG electrode. For instance, the simulated electrical equivalent impedance for channel 18 (O1, occipital) has been calculated as:

$$Z(s) = K_p \frac{1 + sT_z}{s^2T_w^2 + 2s\zeta \cdot T_w + 1} \quad (6)$$

with $K_p = -40921$, $T_w = 0.10848$, $\zeta = 4.7863$, and $T_z = -2.3726$. We used this modeled impedance to simulate

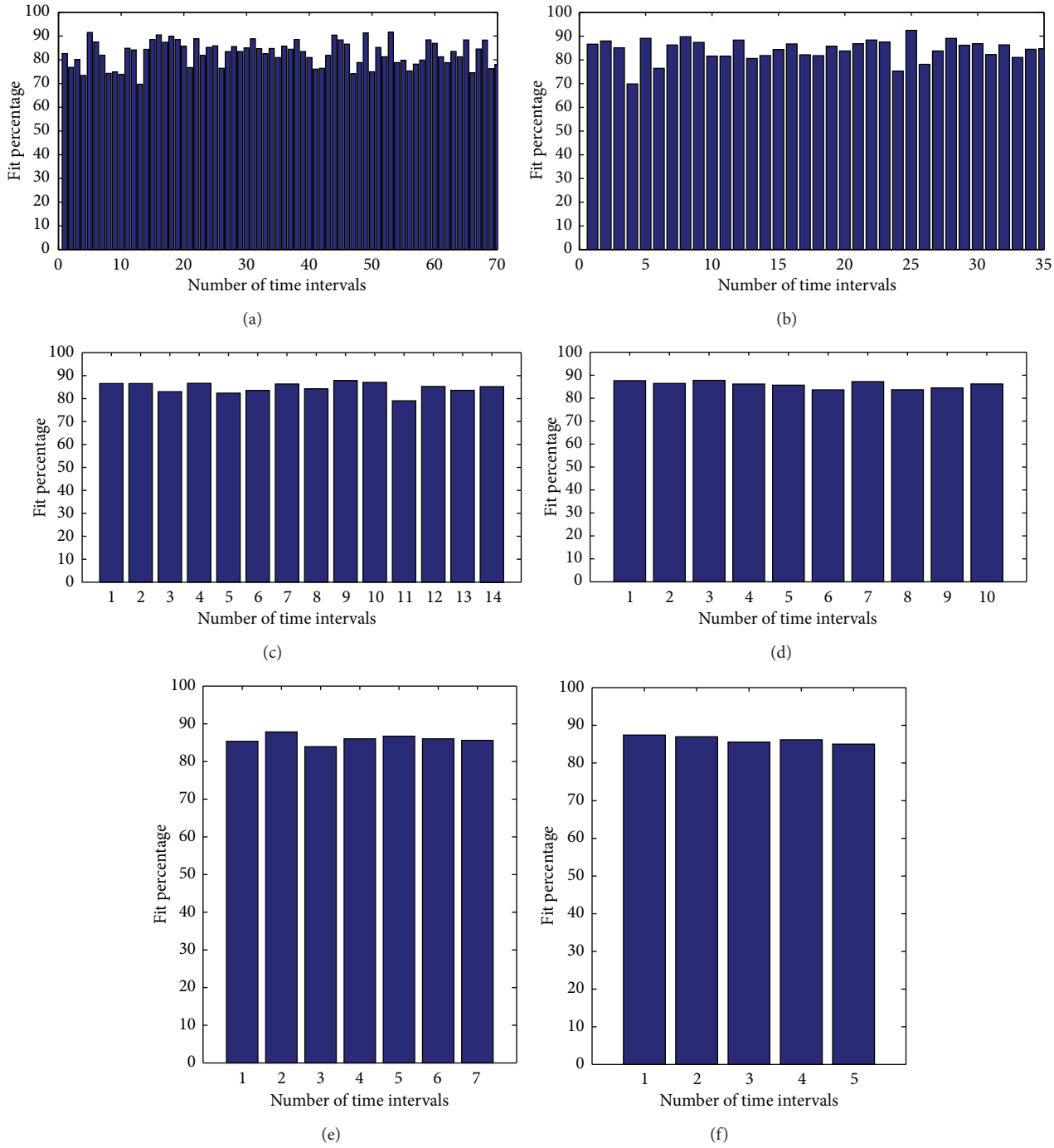


Figure 4: The fit percentage for the simulated GVS artifact at channel 18 for time intervals (a) 1 sec, (b) 2 sec, (c) 5 sec, (d) 7 sec, (e) 10 sec, and (f) 14 sec.

the output signal due to scalp propagation between channel 18 and the GVS electrodes (the simulated GVS artifact) which is the dominant term of the total measured EEG signals, with a high fit percentage of about 87%.

We calculated the impedance models using the entire EEG data collected in each trial (70 seconds). To address the concern about the time-variant properties of the scalp impedance, we computed the impedance models for shorter time intervals (e.g., 1s, 2s, 5s, 7s, 10s, and 14s) and analyzed the fitness of the simulated GVS artifact with the measured EEG data (Figure 4).

The results show that the fitness of the models does not vary for different lengths of time intervals, and for different time intervals it is very close to the fitness of the output model using the entire 70 seconds EEG data, which is around 87%. The above results indicate that the impedance of the scalp can

be represented by one transfer function for the entire trial. To simulate the measured EEG data during the GVS, we combined the simulated GVS artifacts with the clean EEG data collected right before the GVS is applied, in order to get a global data set with known EEG and GVS artifact components. This facilitates a quantitative comparison of the effectiveness of the method in removing the undesirable artifact signals.

2.3. Regression-Based Methods for Artifact Removal. The injected GVS current and the EEG signals are recorded concurrently by the measurement system, while the GVS current distribution through the scalp contaminates the recorded EEG signals. We can use the recorded GVS current as a reference to identify the GVS artifacts in the measured EEG signals. To identify the GVS artifacts in the contaminated EEG signals, we applied time-series regression methods using different model structures. One class of model structures is the *discrete-time polynomial* models, described by the following general equation:

$$A(q)\,y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t). \quad (7)$$

Here $u(t)$ is the recorded GVS current, $y(t)$ is the estimated GVS artifact, and $e(t)$ is a white noise (mean = 0, variance = σ^2) which represents the stochastic part of the model. $A(q)$, $B(q)$, $C(q)$, $D(q)$, and $F(q)$ are polynomials in terms of the time-shift operator q which describe the influence of the GVS current and measurement noise on the EEG data. Model structures such as ARMAX, Box-Jenkins, and Output-Error (OE) are the subsets of the above general polynomial equation. In ARMAX model $F(q)$ and $D(q)$ are equal to 1, in Box-Jenkins $A(q)$ is equal to 1, and in Output-Error model $A(q)$, $C(q)$, and $D(q)$ are equal to 1.

Another class of model structures is Hammerstein-Wiener model, which uses one or two static nonlinear blocks in series with a linear block. This model structure can be employed to capture some of the nonlinear behavior of the system. The linear block is a discrete transfer function, represents the dynamic component of the model, and will be parameterized using an Output-Error model similar to the previous model. The nonlinear block can be a nonlinear function such as dead-zone, saturation, or piecewise-linear functions. As we have not observed any dead-zone or saturation type of nonlinearity in our data, we chose the piecewise-linear function by which we can break down a nonlinear system into a number of linear systems between the breakpoints.

We also used *state-space* models in which the relation between the GVS signals, noise, and the GVS artifacts are described by a system of first-order differential equations relating functions of the state variables, noise, and the GVS signal to the first derivatives of the state variables and Output equations relating the state variables and the GVS signal to the GVS artifact.

2.4. Adaptive Filtering Methods for Artifact Removal. Adaptive filtering is another approach to remove artifacts. This method is specifically suitable for real time applications. The adaptive filter uses the received input data point to refine its properties (e.g., transfer function or filter coefficients) and match the changing parameters at every time instant. These filters have been employed to remove different EEG artifacts [25].

In our application, the primary input to the adaptive filter system is the measured contaminated EEG signal $E_m(n)$ as a mixture of a true EEG $E_t(n)$ and an artifact component $z(n)$. The adaptive filter block takes the GVS current $i_{\mathrm{GVS}}(n)$ as the reference input and estimates the artifact component. The filter coefficients h_m are adjusted recursively in an optimization algorithm driven by an error signal:

$$e(n) = E_m(n) - \widehat{E}_{\mathrm{GVS}}(n) = E_t(n) - \left[z(n) - \widehat{E}_{\mathrm{GVS}}(n)\right], \quad (8)$$

where

$$\widehat{E}_{\mathrm{GVS}}(n) = \sum_{m=1}^{M} h_m \cdot i_{\mathrm{GVS}}(n+1-m). \quad (9)$$

Because of the function of vestibular system which modulates the stimulation signals [26], there is no direct linear correlation between the true EEG $E(n)$ and the GVS current $i_{\mathrm{GVS}}(n)$. On the other hand, there is a strong correlation between the GVS artifact $z(n)$ and $i_{\mathrm{GVS}}(n)$, so we can calculate the expected value of e^2 as follows:

$$E\left[e^2(n)\right] = E\left[\left(E_m(n) - \widehat{E}_{\mathrm{GVS}}(n)\right)^2\right] \quad (10)$$

or

$$E\left[e^2(n)\right] = E\left[E_t^2(n)\right] - E\left[\left(z(n) - \widehat{E}_{\mathrm{GVS}}(n)\right)^2\right]. \quad (11)$$

And as the adjustment of the filter coefficients does not affect the $E[E_t^2(n)]$, therefore minimizing the term $E[(z(n) - \widehat{E}_{\mathrm{GVS}}(n))^2]$ is equivalent to minimizing $E[e^2(n)]$.

Among the various optimization techniques, we chose the Recursive Least-Squares (RLS) and the Least Mean Squares (LMS) for our application. In the section "Comparison of the performance of different artifact removal methods", we compared the results of adaptive filters with those of the other methods.

2.5. Wavelet Decomposition Methods. In this section, we explain how we employ the wavelet methods to enhance the performance of our artifact removal method. The applied GVS current in this study is a pink noise with frequency band of 0.1–10 Hz. Both the GVS current and the EEG data are acquired at the sampling rate of 1000 Hz. After antialiasing filtering, the signals are in a frequency range of 0–500 Hz. The following is the power spectrum of the GVS current using *Welch's method* (Figure 5).

As shown above, the main GVS frequency components are in the range of 0.1 to 10 Hz. To achieve the best fit between the estimated GVS contribution and the measured EEG at each EEG channel, we broke down the recorded GVS current and the contaminated EEG data into various frequency bands by means of wavelet analysis and estimated the GVS artifacts in each frequency band. Wavelet transform is able to

Table 2: Frequency bands for approximation and details components.

	L1	L2	L3	L4	L5	L6
Approximation	0–250	0–125	0–62.5	0–31.25	0–15.75	0–7.87
Details	250–500	125–250	62.5–125	31.25–62.5	15.75–31.25	7.87–15.75
	L7	L8	L9	L10	L11	L12
Approximation	0–3.93	0–1.96	0–0.98	0–0.49	0–0.24	0–0.12
Details	3.93–7.87	1.96–3.93	0.98–1.96	0.49–0.98	0.24–0.49	0.12–0.24

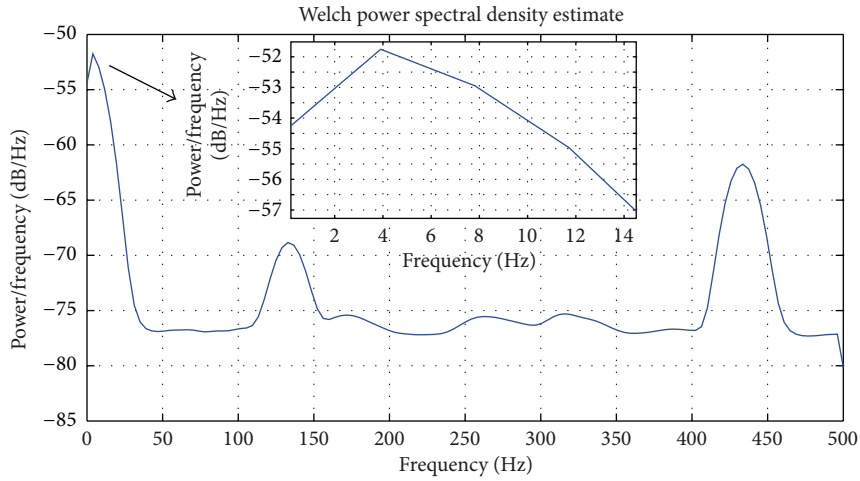


Figure 5: The GVS current power spectrum.

construct a high resolution time-frequency representation of nonstationary signals like EEG signals. In wavelet transform, the signal is decomposed into a set of basis functions, obtained by dilations and shifts of a unique function ψ called the *mother* or the *prototype* wavelet, as opposed to a sine wave which is used as the basis function in the Fourier Transform. When the signals are discrete, the *discrete wavelet transform* (DWT) algorithm can be applied, and the set of basis functions are defined on a *"dyadic"* grid in the time-scale plane as

$$\psi_{j,k}(t) = 2^{-(j/2)}\psi\left(2^{-j}t - k\right), \tag{12}$$

where 2^j governs the amount of scaling and $k2^j$ governs the amount of translation or time shifting. The wavelet transform is the inner product of the basis wavelet functions and the signal in the time domain. In the DWT algorithm, the discrete time-domain signal is decomposed into high frequency or details components and low frequency or approximation components through successive low pass and high pass filters. For multi resolution analysis, the original signal is decomposed into an approximation and details parts. The approximation part is decomposed again by iterating this process; thus one signal can be decomposed into many components. The basic DWT algorithm does not preserve translation invariance. Consequently a translation of wavelet coefficients does not necessarily correspond to the same translation of the original signal. This nonstationary property originates from the downsampling operations in the pyramidal algorithm. The algorithm can be modified by inserting $2^j - 1$ zeros between filters coefficients of the layer j, instead of down-sampling. This modified version of the DWT algorithm is called *stationary wavelet transform* (SWT), and it can preserve the translation invariance property. In this study, we applied both DWT and SWT, to decompose the EEG signals using different mother wavelets such as *Symlet* and *Daubechies* of different orders. Both the GVS current and the simulated EEG signals were decomposed into 12 levels, and thus we have the frequency bands for approximation and detail components, shown in Table 2.

2.6. ICA-Based Methods for Artifact Removal. Independent Component Analysis (ICA) is a statistical method used to extract independent components from a set of measured signals. This method is a special case of the *Blind Source Separation* methods, where the K channels of the recorded EEG signals ($E(t) = e_1(t), \ldots, e_K(t)$) are assumed to be a linear combination of $N(N \leq K)$ unknown independent sources ($S(t) = s_1(t), \ldots, s_N(t)$):

$$E(t) = MS(t), \tag{13}$$

where M is the unknown mixing matrix defining weights for each source contributions to the EEG signals recorded at each channel. In ICA, the measured K channel EEG signals are taken into an N dimensional space and projected onto a coordinate frame where the data projections are minimally overlapped and maximally independent of each other. There are various algorithms with different approaches to find the independent components, such as minimizing the mutual information or maximizing the joint entropy among the data

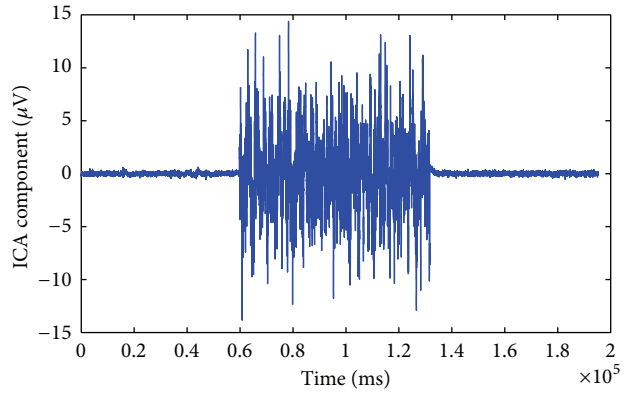


Figure 6: The ICA component attributed to the stimulus artifact, 72 seconds in the middle.

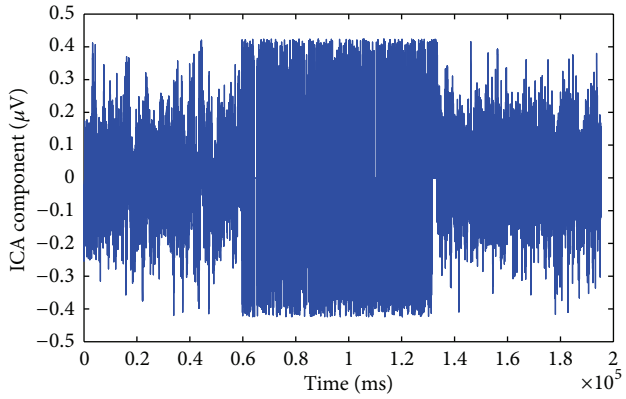


Figure 7: The ICA component attributed to the stimulus artifact after applying the threshold.

projections. The ICA algorithm we used in this study is the *extended Infomax* algorithm [27] which is a modified version of the Infomax algorithm proposed by Bell and Sejnowski [28]. It uses a learning rule that switches between different types of distributions such as Sub-gaussian and Super-gaussian sources. The extended Infomax algorithm is implemented in EEGLAB MATLAB toolbox [29] and widely used to analyze EEG studies. The ICA was applied to the measured EEG set to find the GVS artifacts components. To remove the GVS artifact we need to find all components that are attributed to the GVS applied to the subject. These components can be identified by calculating the correlation coefficient between the ICA components and the GVS signal. The temporal structure of the GVS artifact components is also different from the other components as, during the time that the GVS is applied, a large amplitude artifact appears (Figure 6).

We tried two approaches to remove the artifact. The first approach is to zero out the artifact signals from the components that account for the GVS parasitic influence and obtain a new cleaned-up source matrix $\widehat{S}(t)$. The second approach is to apply a threshold on the artifact components, in order to extract the artifact spikes and set them to zero. The threshold was set at three standard deviations above the mean of the EEG signal without the artifact (e.g., the signal before applying the GVS), and all data points with amplitude over the threshold were set to zero. Thus we obtained a new source matrix, $\widehat{S}(t)$, with the modified components. The threshold at 3 standard deviations of the original neural signals enables us to keep a major part of the original neural activities untouched as much as possible (Figure 7).

Eventually, we reconstruct ICA-corrected EEG signals as:

$$\widehat{E}(t) = M\widehat{S}(t), \tag{14}$$

where $\widehat{E}(t)$ is the new data set which represents the estimated artifact-free data.

2.7. The Proposed Artifact Removal Method. In the proposed method, we decomposed the EEG and GVS current signals in 12 frequency bands (Table 2), and then using the regression methods, we estimated the GVS artifact components in each frequency band. Figure 8 shows the process for detecting GVS artifacts. As shown in this flowchart, in each frequency band, the GVS artifacts are detected through a regression analysis, where the GVS signals are taken as the reference signals.

The estimated GVS artifact frequency components are subtracted from the contaminated EEG frequency components. The wavelet decomposition enables us to focus on the frequency bands of interest and calculate the estimated GVS artifacts in each frequency band independently; thus the regression method can deal better with some nonlinear behaviors of the skin in the frequency domain. This wavelet-based time-frequency analysis approach enhances the performance of the artifact removal method. The cleaned-up signal is reconstructed from the proper frequency components of the estimated GVS signal components in the frequency range of interest (e.g., 1 Hz to 32 Hz). We calculated the correlation coefficients between the GVS signals and the estimated GVS artifacts reconstructed from different frequency bands, and we observed that the regression results improve when we reconstruct the estimated GVS artifact components from corresponding frequency bands separately.

The result of the correlation analysis is tabulated in Table 3. In this analysis, the real data from channel O1, occipital EEG, was decomposed into 12 frequency bands, using the SWT algorithm with the mother wavelet db3, and the GVS current was estimated using OE regression model of order 2. We calculated *Pearson's correlation* for the correlation analysis as

$$\mathrm{Corr}\left(u, \widehat{y}\right) = \frac{\mathrm{Cov}\left(u, \widehat{y}\right)}{\sigma_u \cdot \sigma_{\widehat{y}}}, \tag{15}$$

where $u(t)$ is the recorded GVS current and $\widehat{y}_i(t)$ is the estimated GVS artifact reconstructed from different frequency components.

The result shows that the correlation between the GVS signal and the estimated GVS artifact significantly increases by using wavelet decomposition method. We applied the wavelet transform to remove frequency components lower

Table 3: Correlation between the GVS signal and the estimated GVS artifact reconstructed from different frequency components.

	Estimated GVS artifact without wavelet decomposition	Estimated GVS artifact from 0.12 Hz to 250 Hz	Estimated GVS artifact from 0.24 Hz to 125 Hz	Estimated GVS artifact from 0.49 Hz to 62.5 Hz
Correlation	0.6960	0.8463	0.9168	0.9725
	Estimated GVS artifact from 0.49 Hz to 31.25 Hz	Estimated GVS artifact from 0.49 Hz to 15.75 Hz	Estimated GVS artifact from 0.98 Hz to 31.25 Hz	Estimated GVS artifact from 0.98 Hz to 15.75 Hz
Correlation	0.9776	0.9769	0.9899	0.9899

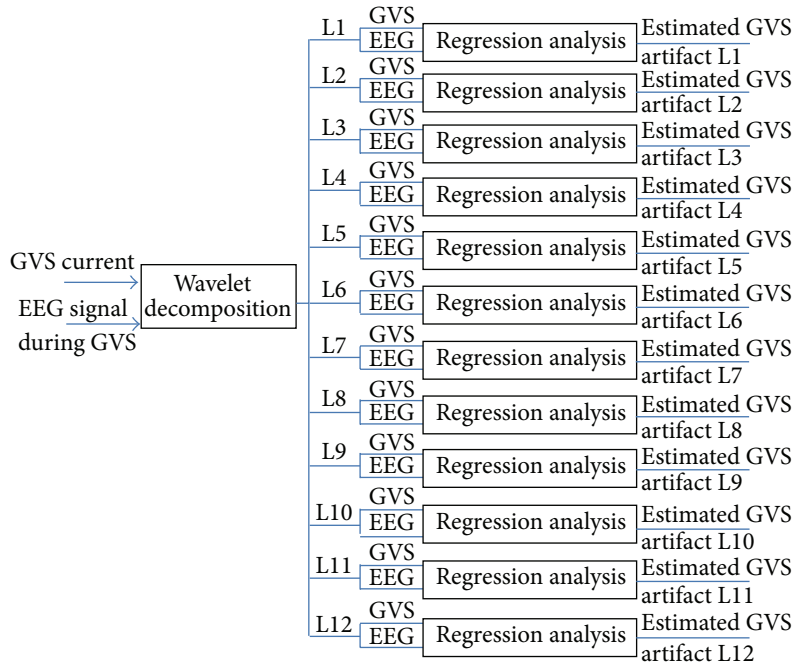


Figure 8: Flowchart of the process for detecting GVS artifacts in the proposed method.

than 0.98 Hz and higher than 31.25 Hz, which are not of the main interest, and the correlation between the GVS signal and estimated GVS artifact was increased up to 0.9899.

We employed both SWT and DWT algorithms in the proposed artifact removal method. The difference between SWT and DWT algorithms was briefly explained in the wavelet analysis section. We also used various regression models to estimate the GVS artifact. To assess the performance of the proposed method using different algorithms and models, we applied our method to the simulated data and examined the cleaned-up EEG signals in comparison with the original artifact-free EEG signals. For this assessment, not only did we calculate the correlation between the artifact-removed EEG signals and the original artifact-free EEG signals, but also we measured the fitness of the artifact-removed signals based on the *normalized residual sum of squares* which is sometime introduced as the *normalized quadratic error* defined by

$$\mathrm{RSS}_N = \frac{\sum \left(E_o(t) - \widehat{E}_o(t)\right)^2}{\sum \left(E_o(t) - \mathrm{mean}\left(E_o(t)\right)\right)^2}, \tag{16}$$

where $E_o(t)$ represents the original artifact-free signal and $\widehat{E}_o(t)$ is the artifact-removed signal.

We measured the performance of the proposed method based on the correlation (15) and the normalized residual sum of squares (16). The choice for the wavelet algorithm and mother wavelet was made such that the performance of the artifact removal method is maximized. To compare different wavelet algorithms and mother wavelets, we employed a number of mother wavelets from two different wavelet families which have been commonly used in EEG signal processing, *Daubechies* (*db*3, *db*4, and *db*5) and *Symlets* (*sym*3, *sym*4, and *sym*5). Both SWT and DWT were used with these mother wavelets in the proposed artifact removal method and applied to the simulated data. We tabulated the normalized residual sum of squares and the correlation between the artifact-removed signals and the original artifact-free signals in the frequency range lower than 31.25 Hz (Table 4).

The results show that SWT algorithm has a superior performance compared to DWT algorithm, and between different mother wavelets both Daubechies and Symlet wavelets with order of 4 performed better than the others.

Another step to improve the performance of the method, is finding an optimum regression method to calculate the estimated GVS artifacts as accurate as possible. We used three different classes of model structure, Output-Error (OE)

Table 4: Correlation and normalized residual sum of squares between the artifact-removed signals and the original artifact-free EEG signals for simulated data using different wavelet decomposition algorithms.

	DWT db3	DWT db4	DWT db5	DWT db6	DWT sym3	DWT sym4	DWT sym5	DWT sym6
Corr.	0.8781	0.9023	0.9155	0.9242	0.8781	0.9023	0.9156	0.9242
RSS_N	0.5517	0.4870	0.4503	0.4255	0.5517	0.4870	0.4503	0.4255
	SWT db3	SWT db4	SWT db5	SWT db6	SWT sym3	SWT sym4	SWT sym5	SWT sym6
Corr.	0.9932	0.9933	0.9933	0.9932	0.9932	0.9933	0.9933	0.9932
RSS_N	0.1710	0.1700	0.1705	0.1714	0.1710	0.1700	0.1705	0.1714

Table 5: Correlation and normalized residual sum of squares between the artifact-removed signals and the original artifact-free EEG signals for simulated data using different models for estimating the GVS artifacts.

	OE2	OE3	OE4	OE5	NLHW2
Corr.	0.9933	0.9933	0.9933	0.9822	0.9934
RSS_N	0.1700	0.1701	0.1704	0.2267	0.1711
	SS2	SS3	SS4	NLHW3	NLHW4
Corr.	0.9933	0.8105	0.7466	0.9926	0.9851
RSS_N	0.1704	0.7628	0.9174	0.1230	0.1725

as a simple special case of the general polynomial model, Hammerstein-Wiener with the piecewise-linear function, and Space-State models, which were all introduced in the "Regression-based approach" section. We employed these models with different orders in the proposed artifact removal method and applied the proposed method using each of these models to the simulated data. In order to compare the performance, we used SWT with Daubechies 4 to decompose the contaminated signals, estimated the GVS artifact using different models, and then assessed the performance in terms of the correlation and the normalized residual sum of squares between the original artifact-free signal and the artifact-removed signal reconstructed in the frequency range lower than 31.25 Hz. The results are tabulated in Table 5.

For nonlinear Hammerstein-Wiener models we used the piecewise-linear function and broke down the EEG signal into a number of intervals. We tried a various number of intervals and observed that, with 4 intervals (or less), we could get the highest correlation and the least residual.

The results show that, between all those models, both Output-Error and nonlinear Hammerstein-Wiener have better performance. We employed these regression models to maximize the performance of the proposed method, then we applied the proposed method to the real data.

We also used two ICA-based methods for removing the artifact: filtering out the artifact components and applying a threshold on the artifact components amplitude to remove the artifact spikes beyond the threshold.

To assess the performances of the ICA methods on the simulated data, we calculated both the correlation and the normalized residual sum of squares between the artifact-removed EEG signals and the original artifact-free EEG signals.

We compared the ICA-based methods with the proposed methods using the Output-Error and nonlinear Hammerstein-Wiener models order 2, along with 12-level STW decomposition with DB4 mother wavelet (Tables 6 and 7).

Table 6: Correlation and normalized residual sum of squares between the artifact-removed signals and the original artifact-free EEG signals for simulated data using the proposed method and ICA-based methods.

	Removing the ICA artifact component	Applying threshold to the ICA artifact component	SWT decomposition with DB4 modeled with OE2	SWT decomposition with DB4 modeled with NLHW2
Corr.	0.6445	0.6171	0.9933	0.9934
RSS_N	0.9567	1.0241	0.1700	0.1711

Table 7: Correlation between the GVS signals and the estimated GVS artifact extracted from EEG signals for real data using the proposed method and ICA-based methods.

	Removing the ICA artifact component	Applying threshold to the ICA artifact component	SWT decomposition with DB4 modeled with OE2	SWT decomposition with DB4 modeled with NLHW2
Corr.	0.6859	0.6858	0.8743	0.8743

2.8. Comparison of Different Artifact Removal Methods. We applied different artifact removal methods on real EEG data acquired during application of GVS. We used the data from channel O1 (occipital EEG) of different subjects in EEG/GVS studies. We applied stimulation signals of different amplitudes in our experiments and observed consistent results from these experiments. By calculating the correlation coefficients between the GVS signals and the estimated GVS artifacts, we compared the performance of these methods. First we compare ICA-based, regression-based, and adaptive filters without using the wavelet analysis. Then we use the proposed method where the wavelet analysis was employed to improve the performance of our artifact removal method.

The best algorithms for ICA-based methods, best models for regression-based methods and best filters for adaptive filtering methods were selected. Between different ICA algorithms (as mentioned in the section "ICA-based artifact removal methods"), the extended Infomax showed better results. Between regression-based methods (as previously introduced in the section "Regression-based artifact removal

Table 8: Correlation between the GVS signals and the estimated GVS artifact extracted from EEG signals for real data using different methods.

Method	Correlation
ICA-Infomax method (remove the artifact component)	0.6859
ICA-Infomax method (threshold the artifact component)	0.6858
Regression method with OE2	0.7673
RLS Adaptive filter (forgetting factor: 0.99997, length: 2)	0.7615
LMS Adaptive filter (adaptation gain: 0.5, length: 3)	0.7010

Table 9: Correlation between the GVS signal and the estimated GVS artifact reconstructed from different frequency components for real data.

Frequency band	Correlation
Estimated GVS artifact without wavelet decomposition	0.7673
Estimated GVS artifact from 0.12 Hz to 250 Hz	0.8463
Estimated GVS artifact from 0.24 Hz to 125 Hz	0.9168
Estimated GVS artifact from 0.49 Hz to 62.5 Hz	0.9725
Estimated GVS artifact from 0.49 Hz to 31.25 Hz	0.9776
Estimated GVS artifact from 0.49 Hz to 15.75 Hz	0.9769
Estimated GVS artifact from 0.98 Hz to 31.25 Hz	0.9899
Estimated GVS artifact from 0.98 Hz to 15.75 Hz	0.9899

methods"), OE order 2 showed better performance, and between adaptive filters (as previously introduced in the section "Adaptive filtering methods for artifact removal"), RLS filter with the forgetting factor of 0.99997, the filter length of 2, LMS filter with the adaptation gain of 0.5, and the filter length of 3 had better performance. We tabulated (Table 8) the correlation between the GVS signals and the estimated GVS artifacts.

The results show that, between all the above methods, the regression-based methods are able to estimate the GVS artifacts with higher correlation with the original GVS signals. Thus, we employed the regression-based method along with the wavelet analysis in our proposed method to achieve the best performance in removing GVS artifact. The wavelet decomposition method improves the estimation of the GVS artifacts in both correlation performance and robustness. This is due to the separate transfer function estimations for each frequency band, aspect that makes it less prone to nonlinear skin behavior or to other noise sources. Furthermore, with wavelet decomposition, we can filter out the frequency components that are not of interest. Removing those frequency components can improve the results of the regression analysis as well. The cleaned EEG data is reconstructed from the frequency range of interest (e.g., 1 Hz to 32 Hz).

Using a correlation analysis, we show how the wavelet-based time-frequency analysis approach enhances the performance of the artifact removal method. We calculated the correlation coefficients between the GVS signals and the estimated GVS artifacts reconstructed from different frequency bands (tabulated in Table 9). We observed that by focusing on the frequency components of interest, for example, between 1 Hz to 32 Hz, we could achieve much higher correlation between the estimated and original GVS signals.

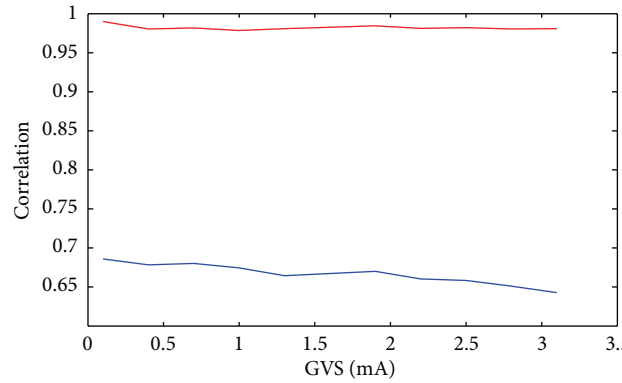


Figure 9: Correlation between the GVS signal and the estimated GVS artifact using the proposed method (red) and the ICA method (blue) for different GVS amplitudes.

As shown in Table 9, after removing the frequency bands lower than 0.98 Hz and larger than 31.25 Hz, which were outside our interest at the present time, the correlation between the GVS signal and the estimated GVS artifact significantly increases from 0.7673 to 0.9899 by using wavelet decomposition method.

So far, we showed the proposed method has superior performance than the other methods when it is applied to low-amplitude stochastic GVS signals up to 1 mA. We also applied our artifact removal method to EEG/GVS data sets collected by our other collaborator in the Sensorimotor Physiology Laboratory, where higher amplitude pink noise GVS up to 3100 μA was applied in the EEG/GVS studies. In one data sets, pink noise GVS in a wide range of amplitudes from 100 μA to 3100 μA (each 300 μA) was applied, and the EEG/GVS data were collected. We compared the performance of the proposed method and the extended Infomax ICA method. The results show that while the performance of the ICA method deteriorates as the GVS amplitude is increased, the proposed method provides a robust performance (Figure 9).

3. Results

In the section "The proposed artifact removal method", we optimized the proposed method using the simulated data. To find the optimum algorithms for signal decomposition, we compared the SWT and DWT decomposition algorithms using different mother wavelets (the results shown in Table 4), and to achieve better estimation of the GVS artifacts, we employed different model structures (results shown in Table 5).

In the optimized algorithm, we employed the SWT decomposition algorithm using DB4 mother wavelet and decomposed the signals into 12 frequency bands. This enabled us to separate the GVS artifact into different frequency bands

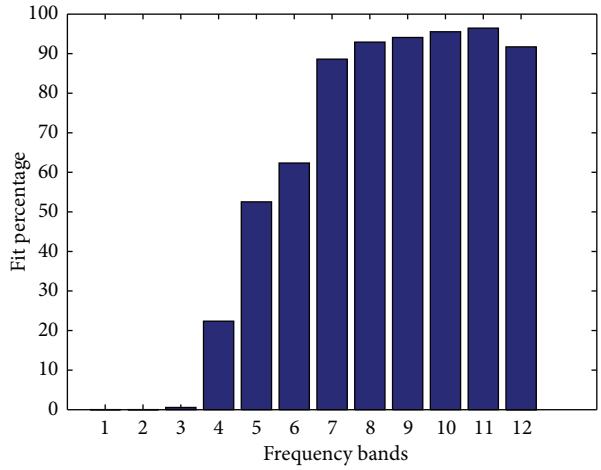


Figure 10: The fit percentage of the detail components of the estimated GVS artifacts using the OE model order 2 in each frequency band.

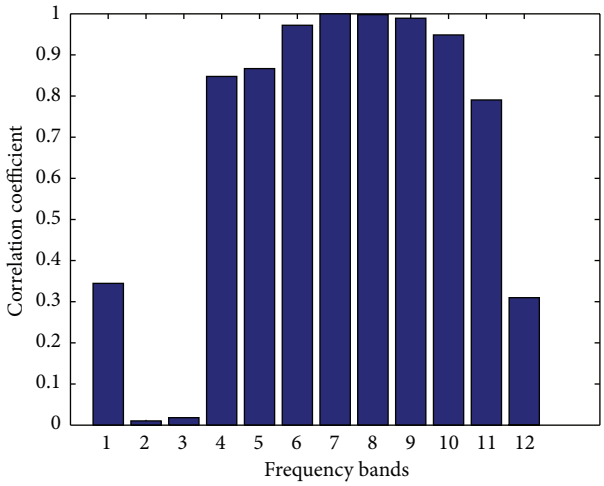


Figure 11: The correlation between the detail components of the estimated GVS signals and the GVS signals for the simulated data using the OE model order 2 in each frequency bands.

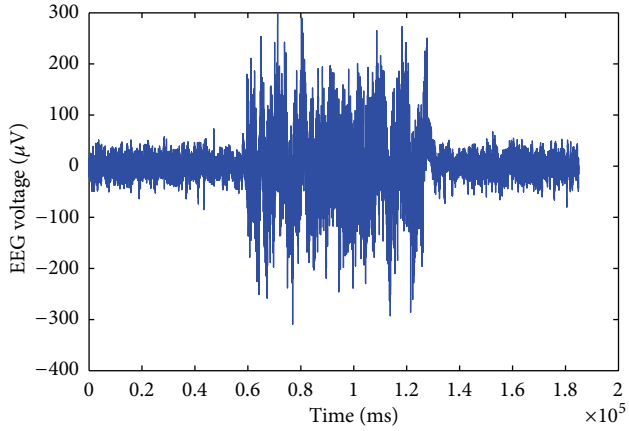


Figure 12: The occipital EEG channel data after applying the proposed artifact removal method using the frequency components lower than 64 Hz.

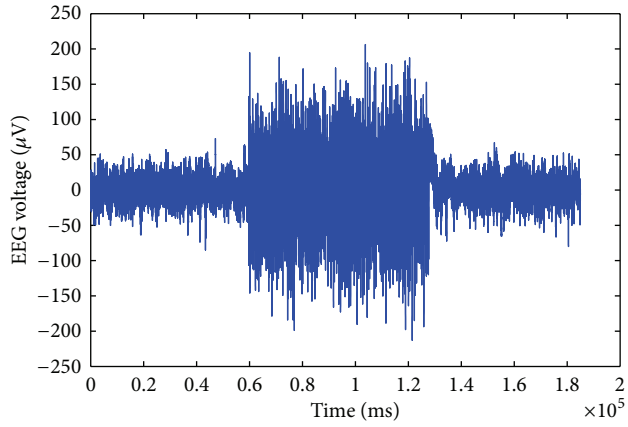


Figure 13: The occipital EEG channel data after applying the proposed artifact removal method using the frequency components between 1 Hz to 32 Hz.

and estimate the artifact using a time-domain regression model. The comparison of the different model structures shows that the Output-Error (OE) and the nonlinear Hammerstein-Wiener order 2 have similar performances, better than the other models.

In the previous section, we compared the performance of different methods and observed that how the combining of wavelet decomposition and regression analysis (Table 9) can improve the performance of the artifact removal method for GVS/EEG studies.

Using the proposed method, we can focus on specific frequency bands and remove the GVS artifact with better performance in each frequency band, separately. Figures 10 and 11 show the fit percentage (5) and the correlation (15) between the detail components of the estimated GVS signals and the GVS signals for the simulated data in the frequency bands introduced in Table 2.

The results show that for frequency components L6 to L10, which correspond approximately to 8–16 Hz, 4–8 Hz, 2–4 Hz, 1-2 Hz, and 0.5–1 Hz bands, we can achieve higher performance in rejecting the GVS artifacts separately. One of the reasons of the robustness of the method is building separate equivalent transfer functions for the GVS signals for each frequency band which helps in maintaining the performance of the algorithms for a large range of GVS intensity levels and frequency ranges. To illustrate the importance of the wavelet analysis, we depicted the artifact-removed signals using different frequency components (Figures 12, 13, and 14).

Figure 14 shows that when we use specific frequency components to estimate the GVS artifacts, we can significantly suppress the GVS artifact and achieve high *signal to artifact ratio* (SAR). SAR is defined as the ratio of the signal amplitude to the artifact amplitude in decibels (dB). We can achieve an SAR of −1.625 dB in the frequency range of 1 Hz–16 Hz, while, using the frequency components in the range of 1 Hz–32 Hz

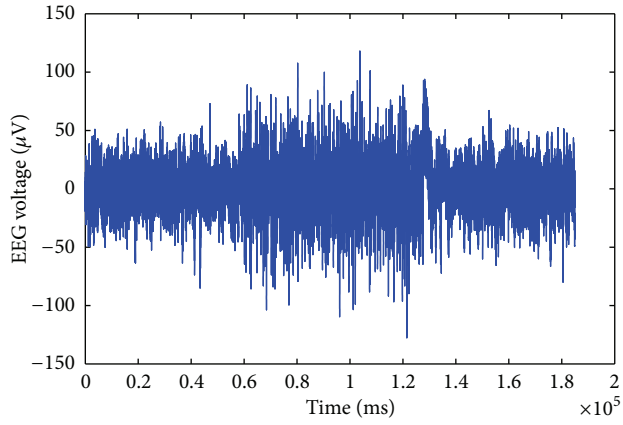


Figure 14: The occipital EEG channel data after applying the proposed artifact removal method using the frequency components between 1 Hz to 16 Hz.

(Figure 13), we can obtain a SAR of −10.498 dB; using the frequency components in the range of 1 Hz–64 Hz (Figure 12), we have an SAR of −13.863 dB. In the original contaminated EEG signals, without removing the GVS artifact (Figure 1), the SAR is −32.189 dB.

4. Discussion

In the section "Simulated data", we showed that by simulating the skin impedance and estimating the transfer function of the skin (one function for the whole frequency range), we could reconstruct a major portion of the GVS artifact. As an example, for channel 18, around 87% of the GVS artifact was reconstructed (Figure 3), thus we could simulate the contaminated EEG signals to assess the performance of the proposed method.

Using the wavelet decomposition, we were able to reconstruct up to 96% of the GVS artifact components in some frequency bands, especially in the frequency range of the GVS signals (Figure 10).

We showed that the use of the wavelet decomposition can improve the time domain regression approach to estimate the GVS artifacts. By means of the combination of the regression and wavelet analysis in the proposed artifact removal method, we were able to focus on different frequency bands and significantly improve the SAR of the contaminated EEG data in specific frequency bands.

The proposed method and the ICA-based methods behave differently in rejecting the GVS artifact. We observed a high correlation between the estimated GVS artifacts and the original GVS signals using the proposed method, but we could not obtain a good correlation using the ICA-based methods.

As illustrated earlier, we cannot completely remove the GVS contamination in all frequency ranges (e.g., over 16 Hz). Removing the whole GVS artifacts remains a problem for the future approaches.

In this study we also observed that nonlinear Hammerstein-Wiener model of the second order, using piecewise-linear blocks with 4 breakpoints (or less), provided the same performance as the Output-Error model of the second order. This implies that the relationships between the GVS artifacts at the EEG electrodes and the injected GVS current are linear and remain constant over the entire epoch. Our simulation study results also showed that the impedance models between the EEG electrodes and the GVS electrodes remain constant over the entire epoch (Figure 4) and using short epochs would not improve the fitness of the impedance models and the estimation of the GVS artifacts. As a matter of fact, it may even worsen the estimation of time-domain characteristics.

We also showed that, when we apply the proposed method to remove the GVS artifacts, less distortion is introduced in the cleaned EEG signals, compared to the distortion that the other methods (e.g., ICA-based methods) introduce. Furthermore, using the proposed method, we do not need to collect and process all EEG channels as in the ICA-based analysis; therefore it is much faster than the ICA-based methods. This allows us to have a simple experimental setup for collecting EEG signals with less EEG channels for the GVS studies which makes the preparation for the data acquisition session take less time before the subject gets tired, and more myogenic and ocular artifacts are introduced. Compared to the ICA methods, the proposed method is easier to be implemented in a real time system for future applications.

Acknowledgments

The authors would like to thank the research team of Professor Martin J. McKeown, from Pacific Parkinson's Research Centre, and also the research team of Professor Jean-Sébastien Blouin, from the Sensorimotor Physiology Laboratory, University of British Columbia, for the collection of the experimental data and for the useful dialogs during our work.

References

[1] Y. Yamamoto, Z. R. Struzik, R. Soma, K. Ohashi, and S. Kwak, "Noisy vestibular stimulation improves autonomic and motor responsiveness in central neurodegenerative disorders," *Annals of Neurology*, vol. 58, no. 2, pp. 175–181, 2005.

[2] W. Pan, R. Soma, S. Kwak, and Y. Yamamoto, "Improvement of motor functions by noisy vestibular stimulation in central neurodegenerative disorders," *Journal of Neurology*, vol. 255, pp. 1657–1661, 2008.

[3] S. Pal, S. M. Rosengren, and J. G. Colebatch, "Stochastic galvanic vestibular stimulation produces a small reduction in sway in parkinson's disease," *Journal of Vestibular Research*, vol. 19, pp. 137–142, 2009.

[4] Y. Yamamoto, R. Soma, Z. R. Struzik, and S. Kwak, "Can electrical vestibular noise be used for the treatment of brain diseases?" in *Proceedings of the 4th International Conference on Unsolved Problems of Noise and Fluctuations in Physics, Biology and High Technology (UPoN '05)*, pp. 279–286, Gallipoli, Italy, June 2005.

[5] K. S. Utz, V. Dimova, K. Oppenlander, and G. Kerkhoff, "Electrified minds: transcranial direct current stimulation (tdcs) and galvanic vestibular stimulation (gvs) as methods of non-invasive brain stimulation in neuropsychology—a review of current data and future implications," *Neuropsychologia*, vol. 48, no. 10, pp. 2789–2810, 2010.

[6] A. J. Shackman, B. W. McMenamin, H. A. Slagter, J. S. Maxwell, L. L. Greischar, and R. J. Davidson, "Electromyogenic artifacts and electroencephalographic inferences," *Brain Topography*, vol. 22, no. 1, pp. 7–12, 2009.

[7] B. W. McMenamin, A. J. Shackman, J. S. Maxwell et al., "Validation of ica-based myogenic artifact correction for scalp and source-localized EEG," *NeuroImage*, vol. 49, no. 3, pp. 2416–2432, 2010.

[8] M. Crespo-Garcia, M. Atienza, and J. L. Cantero, "Muscle artifact removal from human sleep EEG by using independent component analysis," *Annals of Biomedical Engineering*, vol. 36, no. 3, pp. 467–475, 2008.

[9] B. W. McMenamin, A. J. Shackman, J. S. Maxwell, L. L. Greischar, and R. J. Davidson, "Validation of regression-based myogenic correction techniques for scalp and source-localized EEG," *Psychophysiology*, vol. 46, no. 3, pp. 578–592, 2009.

[10] J. Gao, Y. Yang, P. Lin, P. Wang, and C. Zheng, "Automatic removal of eye-movement and blink artifacts from EEG signals," *Brain Topography*, vol. 23, no. 3, pp. 105–114, 2010.

[11] A. Schlogl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of eog artifacts in eeg recordings," *Clinical Neurophysiology*, vol. 118, no. 1, pp. 98–104, 2007.

[12] R. Magjarevic, M. A. Klados, C. Papadelis, C. D. Lithari, and P. D. Bamidis, "The removal of ocular artifacts from eeg signals: a comparison of performances for different methods," in *Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering (IFMBE '09)*, J. Sloten, P. Verdonck, M. Nyssen, and J. Haueisen, Eds., vol. 22, pp. 1259–1263, Springer, Berlin, Germany, 2009.

[13] P. He, G. Wilson, C. Russell, and M. Gerschutz, "Removal of ocular artifacts from the EEG: a comparison between time-domain regression method and adaptive filtering method using simulated data," *Medical and Biological Engineering and Computing*, vol. 45, no. 5, pp. 495–503, 2007.

[14] A. Schloegl, A. Ziehe, and K. R. Müller, "Automated ocular artifact removal: comparing regression and component-based methods," Nature Precedings, 2009.

[15] G. L. Wallstrom, R. E. Kass, A. Miller, J. F. Cohn, and N. A. Fox, "Automatic correction of ocular artifacts in the eeg: a comparison of regression-based and component-based methods," *International Journal of Psychophysiology*, vol. 53, no. 2, pp. 105–119, 2004.

[16] F. Grouiller, L. Vercueil, A. Krainik, C. Segebarth, P. Kahane, and O. David, "A comparative study of different artefact removal algorithms for eeg signals acquired during functional MRI," *NeuroImage*, vol. 38, no. 1, pp. 124–137, 2007.

[17] Y. Erez, H. Tischler, A. Moran, and I. Bar-Gad, "Generalized framework for stimulus artifact removal," *Journal of Neuroscience Methods*, vol. 191, no. 1, pp. 45–59, 2010.

[18] F. Morbidi, A. Garulli, D. Prattichizzo, C. Rizzo, and S. Rossi, "Application of Kalman filter to remove TMS-induced artifacts from EEG recordings," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 6, pp. 1360–1366, 2008.

[19] T. I. Aksenova, D. V. Nowicki, and A.-L. Benabid, "Filtering out deep brain stimulation artifacts using a nonlinear oscillatory model," *Neural Computation*, vol. 21, no. 9, pp. 2648–2666, 2009.

[20] T. Hashimoto, C. M. Elder, and J. L. Vitek, "A template subtraction method for stimulus artifact removal in highfrequency deep brain stimulation," *Journal of Neuroscience Methods*, vol. 113, no. 2, pp. 181–186, 2002.

[21] G. Inuso, F. La Foresta, N. Mammone, and F. C. Morabito, "Brain activity investigation by EEG processing: wavelet analysis, kurtosis and Renyi's entropy for artifact detection," in *Proceedings of the International Conference on Information Acquisition (ICIA '07)*, pp. 195–200, Seogwipo-si, South Korea, July 2007.

[22] G. Inuso, F. La Foresta, N. Mammone, and F. C. Morabito, "Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '07)*, pp. 1524–1529, Orlando, Fla, USA, August 2007.

[23] A. T. Tidswell, A. Gibson, R. H. Bayford, and D. S. Holder, "Electrical impedance tomography of human brain activity with a two-dimensional ring of scalp electrodes," *Physiological Measurement*, vol. 22, no. 1, pp. 167–175, 2001.

[24] J. G. Webster, *Medical Instrumentation-Application and Design*, Wiley, New York, NY, USA, 4th edition, 2009.

[25] A. Garcés Correa, E. Laciar, H. D. Patîo, and M. E. Valentinuzzi, "Artifact removal from EEG signals using adaptive filters in cascade," *Journal of Physics*, vol. 90, Article ID 012081, 2007.

[26] R. C. Fitzpatrick and B. L. Day, "Probing the human vestibular system with galvanic stimulation," *Journal of Applied Physiology*, vol. 96, no. 6, pp. 2301–2316, 2004.

[27] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation*, vol. 11, no. 2, pp. 417–441, 1999.

[28] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[29] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 324325, 7 pages
http://dx.doi.org/10.1155/2013/324325

Research Article

Multiscale Cross-Approximate Entropy Analysis as a Measure of Complexity among the Aged and Diabetic

Hsien-Tsai Wu,[1] Cyuan-Cin Liu,[1] Men-Tzung Lo,[2] Po-Chun Hsu,[1] An-Bang Liu,[3] Kai-Yu Chang,[1] and Chieh-Ju Tang[4]

[1] *Department of Electrical Engineering, National Dong Hwa University, No. 1, Section 2, Da Hsueh Road, Shoufeng, Hualien 97401, Taiwan*
[2] *Research Center for Adaptive Data Analysis & Center for Dynamical Biomarkers and Translational Medicine, National Central University, Chungli 32001, Taiwan*
[3] *Department of Neurology, Buddhist Tzu Chi General Hospital and Buddhist Tzu Chi University, Hualien 97002, Taiwan*
[4] *Department of Internal Medicine, Hualien Hospital, Health Executive Yuan, Hualien 97061, Taiwan*

Correspondence should be addressed to Hsien-Tsai Wu; dsphans@mail.ndhu.edu.tw

Received 22 March 2013; Revised 27 May 2013; Accepted 1 June 2013

Academic Editor: Shengyong Chen

Complex fluctuations within physiological signals can be used to evaluate the health of the human body. This study recruited four groups of subjects: young healthy subjects (Group 1, $n = 32$), healthy upper middle-aged subjects (Group 2, $n = 36$), subjects with well-controlled type 2 diabetes (Group 3, $n = 31$), and subjects with poorly controlled type 2 diabetes (Group 4, $n = 24$). Data acquisition for each participant lasted 30 minutes. We obtained data related to consecutive time series with R-R interval (RRI) and pulse transit time (PTT). Using multiscale cross-approximate entropy (MCE), we quantified the complexity between the two series and thereby differentiated the influence of age and diabetes on the complexity of physiological signals. This study used MCE in the quantification of complexity between RRI and PTT time series. We observed changes in the influences of age and disease on the coupling effects between the heart and blood vessels in the cardiovascular system, which reduced the complexity between RRI and PTT series.

1. Introduction

Multiple temporal and spatial scales produce complex fluctuations within the output signals of physiological systems [1]. In recent studies on translational medicine [1–5], researchers have found that implicit information within the complex fluctuations of physiological signals can be used to evaluate health conditions.

Many recent studies [2, 3] have employed nonlinear dynamical analysis to quantify the complexity of physiological signals in the cardiovascular system. Costa et al. [2] were the first to propose multiscale entropy (MSE) as an approach to analyze the R-R interval (RRI) series of healthy individuals and discovered that the RRI series of young individuals were more complex than that of elderly people. Wu et al. [3] adopted the same method in an examination of pulse wave velocity (PWV) and found that the complexity of these series decreased with aging and/or the progression of diabetes. In addition to time and space, "coupling behavior" in the physiological system also affects the complexity of individual physiological signals, such as RRI or PWV [6]. Drinnan et al. [7] indicated that pulse transit time (PTT) is influenced by RRI and other cardiovascular variables and used cross-correlation functions to quantify the phase relationship between the two time series signals in the cardiovascular system. They established that there was a strong correlation between PTT and RRI variations in healthy subjects. However, Pincus [8] claimed that cross-approximate entropy (Co_ApEn) is more effective than cross-correlation functions in the evaluation of complexity between the two series.

Despite the fact that Co_ApEn has been widely applied to evaluate the complexity between two time series [9–12], single-scale entropy values are not necessarily able to

identify the dynamic complexity of physiological signals. Therefore, this study was an attempt to use a multiscale Co_ApEn (MCE) [13] to quantify the complexity between the synchronous time series of cardiac functions and the degree of atherosclerosis. We assumed that complexity would exist in RRI and PTT series of the cardiovascular system due to the mutual interaction between the heart and blood vessels. Moreover, we assumed that complexity reduces with aging and the influence of disease. We used MCE to develop an index for the quantification of complexity between the two time series capable of distinguishing between healthy individuals and those with diabetes.

2. Methods

2.1. Study Design. This study evaluated the influences of age and diabetes on RRI and PTT. Considering that RRI and PTT are nonlinear, cardiovascular variables, we tested the applicability of MCE in the study subjects and investigated whether this dynamic parameter could provide further information related to the clinical control of diabetes.

2.2. Subject Populations and Experiment Procedure. Between July 2009 and March 2012, four groups of subjects were recruited for this study: young healthy subjects (Group 1, age range: 18–40, $n = 32$), healthy upper middle-aged subjects (Group 2, age range: 41–80, $n = 36$), subjects with well-controlled type 2 diabetes (Group 3, age range: 41–80, $n = 31$, 6.5% $\leqq$ glycosylated hemoglobin (HbA1c) $< 8\%$), and subjects with poorly controlled type 2 diabetes (Group 4, age range: 41–80, $n = 24$, HbA1c $\geqq 8\%$) [3]. The other 22 subjects were excluded due to incomplete or unstable waveform data acquisition. All diabetic subjects were recruited from the Hualien Hospital Diabetic Outpatient Clinic; healthy controls were recruited from a health examination program at the same hospital. None of the healthy subjects had personal or family history of cardiovascular disease. Type 2 diabetes was diagnosed as either fasting sugar higher than 126 mg/dL or HbA1c $\geqq$ 6.5%. All diabetic subjects had been receiving regular treatment and follow-up care in the clinic for more than two years. Regarding the use of medications, there was no significant difference in the type (i.e., antihypertensive, lipid-lowering, and hypoglycemic medications), dosage, and frequency among the well-controlled and poorly controlled diabetic subjects. This study was approved by the Institutional Review Board (IRB) of Hualien Hospital and National Dong Hwa University. All subjects refrained from caffeinated beverages and theophylline-containing medications for 8 hours prior to each hospital visit. Each subject gave informed consent, completed questionnaires on demographic data and medical history, and underwent blood sampling prior to data acquisition. Blood pressure was obtained once from the left arm of supine subjects using an automated oscillometric device (BP3AG1, Microlife, Taiwan) with a cuff of appropriate size, followed by the acquisition of waveform data from the second toe using a six-channel ECG-PWV [14, 15] as previously described.

2.3. Data Collection and Calculation of RRI and PTT Series. All subjects were permitted to rest in a supine position in a quiet, temperature-controlled room at $25 \pm 1^{\circ}$C for 5 minutes prior to subsequent 30-minute measurements. Again, a good reproducibility of six-channel ECG-PWV system [14, 15] was used for waveform measurement from the second toe. Infrared sensors were simultaneously applied to points of reference for the acquisition of data. Electrocardiogram (ECG) measurements were obtained using the conventional method. After being processed through an analog-to-digital converter (USB-6009 DAQ, National Instruments, Austin, TX USA) at a sampling frequency of 500 Hz, the digitized signals were stored on a computer. Because of its conspicuousness, the R wave in Lead II was selected as a reference point: the time interval between the R-wave peak of the jth cardiac cycle to the footpoint of the toe pulse from the left foot was defined as PTT(j); the time difference between the two continues peak of ECG R wave was defined as RRI(i), as shown as Figure 1.

Using ECG and photoplethysmography (PPG), we obtained the RRI series $\{\mathrm{RRI}(i)\} = \{\mathrm{RRI}(1), \mathrm{RRI}(2), \ldots, \mathrm{RRI}(1000)\}$ and PTT series $\{\mathrm{PTT}(j)\} = \{\mathrm{PTT}(1), \mathrm{PTT}(2), \ldots, \mathrm{PTT}(1000)\}$ from each subject. All series were retrieved from 1000 consecutive, stable ECG tracings and PPG toe pulse signals synchronous with the cardiac cycle [14].

Due to a trend within physiological signals [6, 16], nonzero means may be included; therefore, we used empirical mode decomposition (EMD) [17] to deconstruct the $\{\mathrm{RRI}(i)\}$ and $\{\mathrm{PTT}(j)\}$ series, thereby eliminating the trend from the original series. We then normalized the $\{\mathrm{RRI}(i)\}$ and $\{\mathrm{PTT}(j)\}$ series, as shown in (1). In these equations, SD_x and SD_y represent the standard deviations of series $\{\mathrm{RRI}(i)\}$ and $\{\mathrm{PTT}(j)\}$, respectively. Complexity analysis was performed on the normalized results, $\{\mathrm{RRI}'(i)\}$ and $\{\mathrm{PTT}'(j)\}$. Consider

$$\begin{aligned} \left\{\mathrm{RRI}'(i)\right\} &= \frac{\{\mathrm{RRI}(i)\}}{\mathrm{SD}_x}, \\ \left\{\mathrm{PTT}'(j)\right\} &= \frac{\{\mathrm{PTT}(j)\}}{\mathrm{SD}_y}. \end{aligned} \tag{1}$$

2.4. Multiscale Cross-Approximate Entropy (MCE) Using Normalized RRI and PTT Series Together. Previous studies [1–3, 18] have employed MSE to overcome comparison difficulties at a scale factor of 1, when physiological complexity is reduced due to age or disease. However, other research [7] has indicated a strong relationship between variations in PTT series and RRI series; therefore, we used MCE to investigate the interactions between PTT and RRI.

2.4.1. Coarse-Grained Process and Cross-Approximate Entropy (Co_ApEn). MSE involves the use of a scale factor τ ($\tau = 1, 2, 3, \ldots, n$), which is selected according to a 1D series of consecutive cycles. This factor enables the application of a coarse-graining process capable of deriving a new series prior to the calculation of entropy in each new individual series [1–3, 18]. Using this approach, we performed coarse-graining on the normalized 1D consecutive cycles of the $\{\mathrm{RRI}'(i)\}$ and $\{\mathrm{PTT}'(j)\}$ series based on scale factor τ,

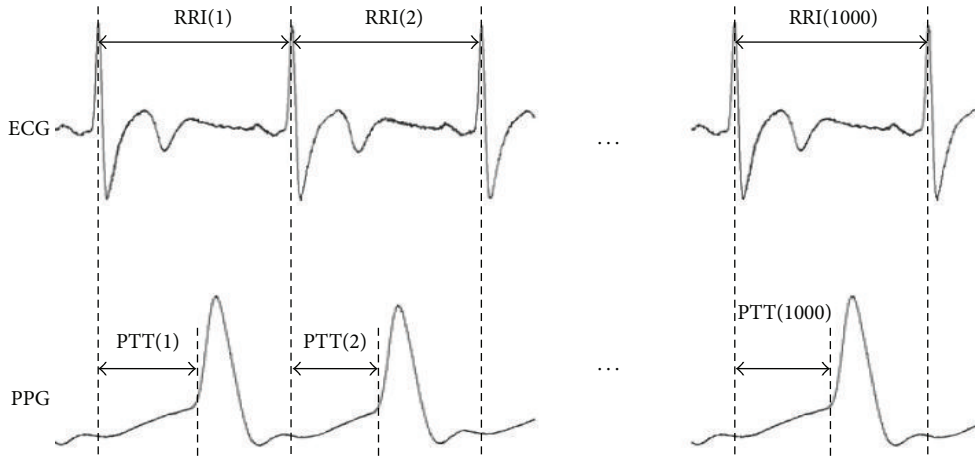


Figure 1: 1000 consecutive data points from ECG signals and PPG signals: PTT(j) refers to the time interval between the R-wave peak of the jth cardiac cycle to the footpoint of the toe pulse from the left foot.

thereby obtaining the series $\{\mathrm{RRI}'^{(\tau)}\}$ and $\{\mathrm{PTT}'^{(\tau)}\}$ as shown in (2). We then calculated entropy as follows:

$$\mathrm{RRI}'(u)^{(\tau)} = \frac{1}{\tau}\sum_{i=(u-1)\tau+1}^{u\tau}\mathrm{RRI}'(i), \quad 1 \le u \le \frac{1000}{\tau},$$
$$\mathrm{PTT}'(u)^{(\tau)} = \frac{1}{\tau}\sum_{j=(u-1)\tau+1}^{u\tau}\mathrm{PTT}'(j), \quad 1 \le u \le \frac{1000}{\tau}. \tag{2}$$

Previous studies [19, 20] have used Co_ApEn, an improved analysis method of approximate entropy, to analyze two synchronous physiological time series, define their relationship, and calculate the complexity within that relationship [8, 21]. This method utilizes the dynamic changes between the two series to evaluate the physiological system. Similarities between changes in the two series can be used to observe the regulatory mechanisms in the physiological system. However, many studies [8, 19–21] presented their results at a scale factor of 1. To obtain a deeper understanding of the complexity of the physiological system, we utilized coarse-grained $\{\mathrm{RRI}'^{(\tau)}\}$ and $\{\mathrm{PTT}'^{(\tau)}\}$ series to calculate the Co_ApEn at each scale, using (7). We refer to this approach as multiscale cross-approximate entropy (MCE). The details of the algorithm are as follows [22].

(1) For given m, for two sets of m-vectors,

$$\mathbf{x}(i) \equiv \left[\mathrm{RRI}'^{(\tau)}(i)\ \ \mathrm{RRI}'^{(\tau)}(i+1)\ \cdots\ \mathrm{RRI}'^{(\tau)}(i+m-1)\right], \quad i = 1,\ N-m+1,$$
$$\mathbf{y}(j) \equiv \left[\mathrm{PTT}'^{(\tau)}(j)\ \ \mathrm{PTT}'^{(\tau)}(j+1)\ \cdots\ \mathrm{PTT}'^{(\tau)}(j+m-1)\right], \quad j = 1,\ N-m+1. \tag{3}$$

(2) Define the distance between the vectors $\mathbf{x}(i)$, $\mathbf{y}(j)$ as the maximum absolute difference between their corresponding elements, as follows:

$$d[\mathbf{x}(i), \mathbf{y}(j)] = \max_{k=1}^{m}\left[\left|\mathrm{RRI}'^{(\tau)}(i+k-1) - \mathrm{PTT}'^{(\tau)}(j+k-1)\right|\right]. \tag{4}$$

(3) With the given $\mathbf{x}(i)$, find the value of $d[\mathbf{x}(i),\mathbf{y}(j)]$ (where $j = 1$ to $N-m+1$) that is smaller than or equal to r and the ratio of this number to the total number of m-vectors ($N-m+1$). That is,

let $N^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(i)$ = the number of $\mathbf{y}(j)$ satisfying the requirement $d[\mathbf{x}(i),\mathbf{y}(j)] \leqq r$, then

$$C^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(i) = \frac{N^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(i)}{N-m+1}. \tag{5}$$

$C^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(i)$ measures the frequency of the m-point $\mathrm{PTT}'^{(\tau)}$ pattern being similar (within a tolerance of $\pm r$) to the m-point $\mathrm{RRI}'^{(\tau)}$ pattern formed by $\mathbf{x}(i)$.

(4) Average the logarithm of $C^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(i)$ over i to obtain $\phi^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(r)$, as follows:

$$\phi^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\ln C^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(i). \tag{6}$$

(5) Increase m by 1, and repeat steps 1 ~ 4 to obtain $C^{m+1}_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(i)$, $\phi^{m+1}_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(r)$.

(6) Finally, take $\mathrm{Co\_ApEn}_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(m,r) = \lim_{N\to\infty}[\phi^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(r) - \phi^{m+1}_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(r)]$ and for N-point data, the estimate is

$$\mathrm{Co\_ApEn}_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(m,r,N) = \phi^m_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(r) - \phi^{m+1}_{\mathrm{RRI}'^{(\tau)}\mathrm{PTT}'^{(\tau)}}(r), \tag{7}$$

where m represents the chosen vector dimension, r represents a tolerance range, and N is the data length. To ensure efficiency and accuracy of calculation, the parameters of this study were set at $m = 3$, $r = 0.15$, and $N = 1000$.

2.4.2. RRI and PTT-Based Multiscale Cross-Approximate Entropy Index (MCEI) for Small and Large Scales. The values of $\text{Co\_ApEn}_{\text{RRI}'^{(\tau)}\text{PTT}'^{(\tau)}}(\tau)$ were obtained from a range of scale factors between 1 and 20 using the MCE data analysis method. The values of $\text{Co\_ApEn}_{\text{RRI}'^{(\tau)}\text{PTT}'^{(\tau)}}(\tau)$ between scale factors 1 and 5 were defined as small scale; those between scale factors 6 and 20 were defined as large scale [23]. The sum of MCE between scale factors 1 and 5 was MCEI_{SS} in (8), while the sum of MCE between scale factors 6 and 20 was MCEI_{LS} in (9). Defining and calculating these two indices of multiscale cross-approximate entropy enables the assessment and quantification of complexity in RRI and PTT between different scale factors. Consider

$$\text{MCEI}_{\text{SS}} = \sum_{\tau=1}^{5} \text{Co\_ApEn}_{\text{RRI}'^{(\tau)}\text{PTT}'^{(\tau)}}(\tau), \tag{8}$$

$$\text{MCEI}_{\text{LS}} = \sum_{\tau=6}^{20} \text{Co\_ApEn}_{\text{RRI}'^{(\tau)}\text{PTT}'^{(\tau)}}(\tau). \tag{9}$$

2.5. Multiscale Entropy Index (MEI) Using RRI or PTT Only. Sample entropy (S_E) was used to quantify the complexity of RRI or PTT series in twenty scales. The values of S_E between scale factors 1 and 5 were defined as small scale, whereas those between scale factors 6 and 20 were defined as large scale. The sum of MSE in small scale was defined as MEI_{SS}, while the sum of MSE in large scale was MEI_{LS} [3].

2.6. Statistical Analysis. Average values were expressed as mean ± SD. Significant differences in anthropometric, hemodynamic, and computational parameters (i.e., RRI, PTT, MCEI_{SS}, and MCEI_{LS}) between different groups were determined using an independent sample t-test. Statistical Package for the Social Science (SPSS, version 14.0 for Windows) was used for all statistical analysis. A P value less than 0.05 was considered statistically significant.

3. Results

3.1. Comparison of Basic Demographic and Cardiovascular Parameters in Different Groups. Table 1 presents the basic demographic parameters of Group 1 and Group 2, showing no significant difference in major demographic parameters except for age, HbA1c levels, and body height. Significant differences were observed in body mass index (BMI), waist circumference, systolic blood pressure (SBP), pulse pressure (PP), HbA1c levels, and fasting blood sugar level between Group 2 and Group 3 (Group 3 > Group 2). In addition, significant differences were also observed in HbA1c levels, triglycerides, and fasting blood sugar level between Group 3 and Group 4.

3.2. MCEI_{LS} as Parameters Indicative of Age and Diabetic Control. There were no significant differences in the values of S_E(RRI) and S_E(PTT) at any scale (Figure 2), or in MEI_{SS}(RRI), MEI_{LS}(RRI), MEI_{SS}(PTT), and MEI_{LS}(PTT) among the 4 groups (Table 1).

Figure 3 summarizes the results of the MCE analysis for the values of RRI and PTT time series over 1000 identical cardiac cycles obtained from the four groups of participants. At a scale factor of 1 ($\tau = 1$), the magnitudes of $\text{Co\_ApEn}_{\text{RRI}'^{(1)}\text{PTT}'^{(1)}}(1)$ ranked as follows: Group 1/Group 3/Group 4/Group 2. The value of $\text{Co\_ApEn}_{\text{RRI}'^{(\tau)}\text{PTT}'^{(\tau)}}(\tau)$ began dropping in all groups at a scale factor of 2 ($\tau = 2$).

Beginning at a scale factor of 3 ($\tau = 3$), the reduction in $\text{Co\_ApEn}_{\text{RRI}'^{(\tau)}\text{PTT}'^{(\tau)}}(\tau)$ in Group 1 slowed. However, in the other groups, the values continued decreasing rapidly. Beginning at a scale factor of 5 ($\tau = 5$), the $\text{Co\_ApEn}_{\text{RRI}'^{(\tau)}\text{PTT}'^{(\tau)}}(\tau)$ of Group 2 achieved stability with only minor fluctuations. The decline in $\text{Co\_ApEn}_{\text{RRI}'^{(\tau)}\text{PTT}'^{(\tau)}}(\tau)$ in Group 4 remained greater than that in Group 3. When plotted against large scale factors (i.e., 6–20), the magnitudes of $\text{Co\_ApEn}_{\text{RRI}'^{(\tau)}\text{PTT}'^{(\tau)}}(\tau)$ ranked as follows: Group 1, Group 2, Group 3, and Group 4.

MCEI_{SS} only presented a significant difference between Groups 1 and 2 (10.18 ± 0.52 versus 9.42 ± 0.70, $P < 0.01$). The differences among Groups 2, 3, and 4 did not reach statistical significance. In comparison, MCEI_{LS} presented significant differences among all four of the groups (Group 1 versus Group 2: 28.30 ± 1.26 versus 25.96 ± 1.99, $P < 0.01$; Group 2 versus Group 3: 25.96 ± 1.99 versus 23.14 ± 1.85, $P < 0.01$; Group 3 versus Group 4: 23.14 ± 1.85 versus 20.13 ± 1.73, $P < 0.01$) (Table 1).

4. Discussion

Since Pincus and Singer's study [19], Co_ApEn has generally been used to reveal similarities between two synchronous, consecutive variables within a single network. This approach has also been used to research the complexity of physiological signals [12, 19]; however, the influence of multiple temporal and spatial scales creates complexity. Thus, this study employed multiscale Co_ApEn (MCE) to evaluate the complexity between the cardiac function-related parameter, RRI, and the atherosclerosis-related parameter, PTT, in the cardiovascular systems of various subject groups.

Previous studies [1, 2, 18] have also indicated that physiological signals are generally nonlinear and exist in nonstationary states. The use of MSE to quantify complexity within the times series of a single type of physiological signal (i.e., RRI or PWV) demonstrated that the complexity of physiological signals decreases with aging [2] or with the influence of diabetes [3]. In this study, although we used MSE to quantify complexity of RRI or PTT series, there were no significant differences in MEI_{SS}(RRI), MEI_{LS}(RRI), MEI_{SS}(PTT), and MEI_{LS}(PTT) between well-controlled and poor-controlled diabetic subjects. Therefore, the influence of the degree of glycemic control on complexity of physiological signals might not be evaluated efficiently according to the use of MSE when analyzing single time series (i.e., RRI or PTT).

Drinnan et al.'s study [7] stated that cardiovascular variables such as RRI and PTT are regulated by complex

Table 1: Comparisons of demographic, anthropometric, and serum biochemical parameters, $MCEI_{SS}$, and $MCEI_{LS}$ among different subject populations.

Parameters	Group 1	Group 2	Group 3	Group 4
Age, year	26.56 ± 9.60	58.19 ± 8.29**	62.74 ± 0.55	60.58 ± 7.68
Body height, cm	169.38 ± 7.92	162.83 ± 6.85**	161.56 ± 8.97	161.17 ± 7.28
Body weight, kg	66.38 ± 12.21	65.22 ± 11.55	69.40 ± 11.37	73.75 ± 14.86
BMI, kg/m^2	23.02 ± 3.27	24.55 ± 3.90	26.52 ± 3.21†	28.42 ± 5.47
Waist circumference, cm	81.20 ± 11.09	82.94 ± 11.00	93.33 ± 9.37††	97.46 ± 3.77
SBP, mmHg	116.50 ± 12.89	115.67 ± 14.12	128.32 ± 16.08††	128.46 ± 16.36
DBP, mmHg	71.44 ± 6.70	74.75 ± 9.93	75.58 ± 9.63	78.21 ± 9.89
PP, mmHg	42.97 ± 0.96	40.92 ± 9.29	52.74 ± 14.34††	50.25 ± 13.12
HbA1c, %	5.43 ± 0.32	5.84 ± 0.34**	6.74 ± 0.62††	9.36 ± 1.59‡‡
Triglyceride, mg/dL	88.88 ± 62.54	114.06 ± 88.15	120.87 ± 47.74	168.04 ± 98.43‡
Fasting blood sugar, mg/dL	93.13 ± 6.96	97.78 ± 14.69	127.27 ± 24.75††	183.96 ± 58.66‡‡
MEI_{SS}(RRI)	9.31 ± 0.54	8.54 ± 0.78	8.00 ± 1.08†	7.64 ± 0.81
MEI_{LS}(RRI)	27.11 ± 2.16	26.38 ± 2.07	25.59 ± 2.89	25.45 ± 3.25
MEI_{SS}(PTT)	9.97 ± 0.38	9.90 ± 0.40	9.85 ± 0.56	9.50 ± 1.41
MEI_{LS}(PTT)	26.73 ± 2.40	23.86 ± 3.71**	21.65 ± 2.55†	21.06 ± 4.92
$MCEI_{SS}$	10.18 ± 0.52	9.42 ± 0.70**	9.41 ± 0.62	9.25 ± 0.39
$MCEI_{LS}$	28.30 ± 1.26	25.96 ± 1.99**	23.14 ± 1.85††	20.13 ± 1.73‡‡

Group 1: healthy young subjects, Group 2: healthy upper middle-aged subjects, Group 3: type 2 diabetic well-controlled patients, Group 4: type 2 diabetic poorly controlled patients. Values are expressed as mean ± SD. BMI: body mass index; SBP: systolic blood pressure; DBP: diastolic blood pressure; PP: pulse pressure; HbA1c: glycosylated hemoglobin; MEI_{SS}(RRI): R-R interval-based multiscale entropy index with small scale; MEI_{LS}(RRI): R-R interval-based multiscale entropy index with large scale; MEI_{SS}(PTT): pulse transit time-based multiscale entropy index with small scale; MEI_{LS}(PTT): pulse transit time-based multiscale entropy index with large scale; $MCEI_{SS}$: multiscale $\text{Co\_ApEn}_{\text{RRI}^{\prime(\tau)}\text{PTT}^{\prime(\tau)}}(\tau)$ index with small scale; $MCEI_{LS}$: multiscale $\text{Co\_ApEn}_{\text{RRI}^{\prime(\tau)}\text{PTT}^{\prime(\tau)}}(\tau)$ index with large scale.

$^{\dagger}P < 0.05$ Group 2 versus Group 3, $^{\ddagger}P < 0.05$ Group 3 versus Group 4. $^{**}P < 0.01$ Group 1 versus Group 2, $^{\dagger\dagger}P < 0.01$ Group 2 versus Group 3, and $^{\ddagger\ddagger}P < 0.01$ Group 3 versus Group 4.

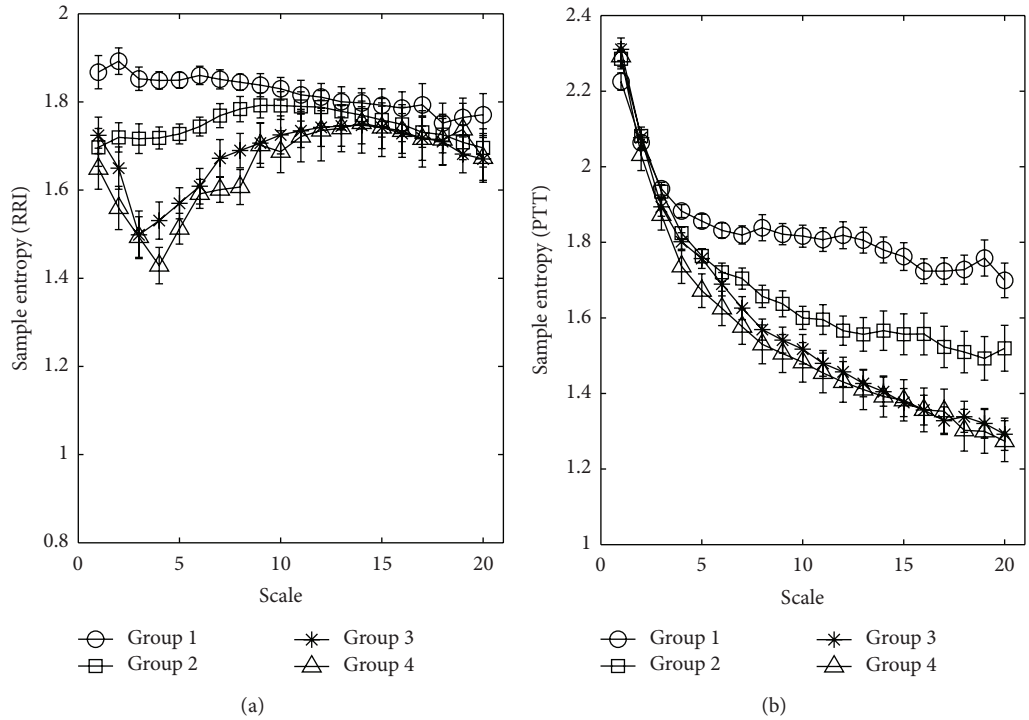


Figure 2: Multiscale entropy (MSE) analysis of (a) RRI and (b) PTT time series showing changes in sample entropy, S_E, among the four groups of study subjects for different scale factors. Symbols represent the mean values of entropy for each group, and bars represent the standard error (given by SE = SD/$\sqrt{n}$, where n is the number of subjects).

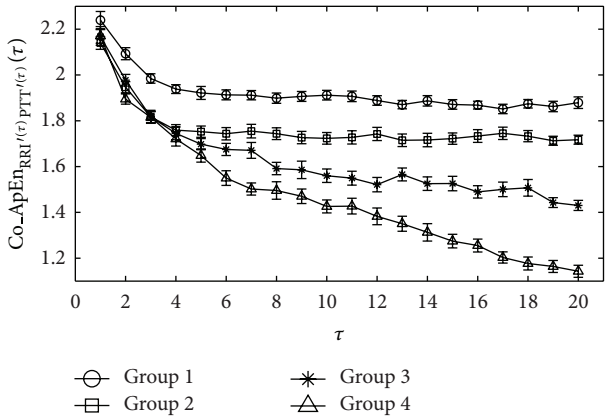


FIGURE 3: Co_ApEn$_{RRI'^{(\tau)}PTT'^{(\tau)}}(\tau)$ curve of the four groups was calculated using the MCE calculation (τ =1~20) on 1000 consecutive RRI and PTT times series. Symbols represent the mean values of entropy for each group, and bars represent the standard error (given by SE = SD/$\sqrt{n}$, where n is the number of subjects).

physiological systems and that a strong relationship exists between variations in PTT and those in RRI. We therefore employed the Co_ApEn integrated with preprocessing coarse-graining to calculate MCEI values as well as the complexity between the synchronous time series RRI and PTT. Figure 3 shows that at small-scale factors (from 1 to 5), it is difficult to determine the influence of age, diabetes, or glycemic control based on the complexity between the time series RRI and PTT using Co_ApEn$_{RRI'^{(\tau)}PTT'^{(\tau)}}(\tau)$. Similarly, MCEI$_{SS}$ indicates only that aging reduces the complexity between the two time series. This finding is similar to that of previous studies [3]. As the scale factor increased (from 6 to 20), Co_ApEn$_{RRI'^{(\tau)}PTT'^{(\tau)}}(\tau)$ began revealing significant differences between the four study groups (Figure 3). Table 1 shows that the MCEI$_{LS}$ values of the young healthy subjects were the highest, whereas subjects with poorly controlled type 2 diabetes were the lowest. This may be due to the fact that the coupling effect between the heart and the blood vessels in the cardiovascular system varies according to age and the influence of disease [24, 25]. In other words, the complexity between the time series RRI and PTT decreases due to age and disease.

Although the MCEI$_{LS}$ can be used to quantify the complexity of RRI and PTT and have been shown to effectively identify significant difference among study groups, limitations still exist. First, a lengthy process of data acquisition and considerable calculation and off-line processing is needed. MCE analysis involves a 30-minute measurement, as opposed to the relatively shorter duration measurement of only RRI and PTT, making the process tiring for participants. The nature of analysis postmeasurement further prevented subjects from receiving their MCEI test results immediately. Second, the medications that the diabetic patients used such as hypoglycemic, antihyperlipidemic, and antihypertensive drugs may also affect autonomic nervous activity. These effects, however, were difficult to assess. The potential effect of medications, therefore, was not considered in the statistical analysis of this study.

5. Conclusions

This study integrates cross-approximate entropy with multiple scales to analyze the complexity between two synchronous physiological signals (RRI and PTT) in the cardiovascular system. According to our results, MCEI$_{LS}$ clearly reveals a reduction in the complexity of two physiological signals caused by aging and diabetes.

Authors' Contribution

M.-T. Lo and A.-B. Liu equally contributed in this study compared with the corresponding author.

Conflict of Interests

The authors declare no conflict of interests.

Acknowledgments

The authors would like to thank the volunteers involved in this study for allowing them to collect and analyze their data. The authors are grateful for the support of Texas Instruments, Taiwan, in sponsoring the low power instrumentation amplifiers and ADC tools. The authors would also like to thank Miss Crystal J. McRae who is a native English speaker, to go over the whole paper. This research was partly supported by National Science Council under Grants NSC 100-2221-E-259-030-MY2 and NSC 101-2221-E-259-012 and National Dong Hwa University on campus interdisciplinary integration Projects no. 101T924-3 and 102T931-3. M.-T. Lo was supported by NSC (Taiwan, ROC), Grant no. 100-2221-E-008-008-MY2, joint foundation of CGH and NCU, Grant no. CNJRF-101CGH-NCU-A4, VGHUST102-G1-2-3 and NSC support for the Center for Dynamical Biomarkers and Translational Medicine, National Central University, Taiwan (NSC 101-2911-I-008-001).

References

[1] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, no. 2, part 1, 2005.

[2] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy to distinguish physiologic and synthetic RR time series," *Computing in Cardiology*, vol. 29, pp. 137–140, 2002.

[3] H.-T. Wu, P.-C. Hsu, C.-F. Lin et al., "Multiscale entropy analysis of pulse wave velocity for assessing atherosclerosis in the aged and diabetic," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2978–2981, 2011.

[4] R. T. Vieira, N. Brunet, S. C. Costa, S. Correia, B. G. A. Neto, and J. M. Fechine, "Combining entropy measures and cepstral analysis for pathological voices assessment," *Journal of Medical and Biological Engineering*, vol. 32, no. 6, pp. 429–435, 2012.

[5] J. Y. Lan, M. F. Abbod, R. G. Yeh, S. Z. Fan, and J. S. Shieh, "Review: intelligent modeling and control in anesthesia," *Journal of Medical and Biological Engineering*, vol. 32, no. 5, pp. 293–307, 2012.

[6] C.-K. Peng, M. Costa, and A. L. Goldberger, "Adaptive data analysis of complex fluctuations in physiologic time series," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 61–70, 2009.

[7] M. J. Drinnan, J. Allen, and A. Murray, "Relation between heart rate and pulse transit time during paced respiration," *Physiological Measurement*, vol. 22, no. 3, pp. 425–432, 2001.

[8] S. M. Pincus, "Approximate entropy in cardiology," *Herzschrittmachertherapie und Elektrophysiologie*, vol. 11, no. 3, pp. 139–150, 2000.

[9] S. M. Pincus, T. Mulligan, A. Iranmanesh, S. Gheorghiu, M. Godschalk, and J. D. Veldhuis, "Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 24, pp. 14100–14105, 1996.

[10] F. Roelfsema, S. M. Pincus, and J. D. Veldhuis, "Patients with Cushing's disease secrete adrenocorticotropin and cortisol jointly more asynchronously than healthy subjects," *Journal of Clinical Endocrinology and Metabolism*, vol. 83, no. 2, pp. 688–692, 1998.

[11] J. D. Veldhuis, A. Iranmanesh, T. Mulligan, and S. M. Pincus, "Disruption of the young-adult synchrony between luteinizing hormone release and oscillations in follicle-stimulating hormone, prolactin, and nocturnal penile tumescence (NPT) in healthy older men," *Journal of Clinical Endocrinology and Metabolism*, vol. 84, no. 10, pp. 3498–3505, 1999.

[12] D.-Y. Wu, G. Cai, Y. Yuan et al., "Application of nonlinear dynamics analysis in assessing unconsciousness: a preliminary study," *Clinical Neurophysiology*, vol. 122, no. 3, pp. 490–498, 2011.

[13] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: a tool for complexity analysis of multichannel data," *Physical Review E*, vol. 84, no. 6, Article ID 061918, 2011.

[14] A.-B. Liu, P.-C. Hsu, Z.-L. Chen, and H.-T. Wu, "Measuring pulse wave velocity using ECG and photoplethysmography," *Journal of Medical Systems*, vol. 35, no. 5, pp. 771–777, 2011.

[15] H. T. Wu, P. C. Hsu, A. B. Liu, Z. L. Chen, R. M. Huang, C. P. Chen et al., "Six-channel ECG-based pulse wave velocity for assessing whole-body arterial stiffness," *Blood Press*, vol. 21, no. 3, pp. 167–176, 2012.

[16] Z. Wu, N. E. Huang, S. R. Long, and C.-K. Peng, "On the trend, detrending, and variability of nonlinear and nonstationary time series," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 38, pp. 14889–14894, 2007.

[17] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, no. 1971, pp. 903–995, 1998.

[18] M. D. Costa, C.-K. Peng, and A. L. Goldberger, "Multiscale analysis of heart rate dynamics: entropy and time irreversibility measures," *Cardiovascular Engineering*, vol. 8, no. 2, pp. 88–93, 2008.

[19] S. Pincus and B. H. Singer, "Randomness and degrees of irregularity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 5, pp. 2083–2088, 1996.

[20] M. Kreuzer, H. Hentschke, B. Antkowiak, C. Schwarz, E. F. Kochs, and G. Schneider, "Cross-approximate entropy of cortical local field potentials quantifies effects of anesthesia—a pilot study in rats," *BMC Neuroscience*, vol. 11, article 122, 2010.

[21] S. M. Pincus, "Irregularity and asynchrony in biologic network signals," *Methods in Enzymology*, vol. 321, pp. 149–182, 2000.

[22] F. Yang, B. Hong, and Q. Tang, "Approximate entropy and its application to biosignal analysis," in *Nonlinear Biomedical Signal Processing: Dynamic Analysis and Modeling*, M. Akay, Ed., vol. 2, John Wiley & Sons, Hoboken, NJ, USA, 2000.

[23] D. Cheng, S.-J. Tsai, C.-J. Hong, and A. C. Yang, "Reduced physiological complexity in robust elderly adults with the APOE $\varepsilon 4$ allele," *PLoS ONE*, vol. 4, no. 11, Article ID e7733, 2009.

[24] D. E. Vaillancourt and K. M. Newell, "Changing complexity in human behavior and physiology through aging and disease," *Neurobiology of Aging*, vol. 23, no. 1, pp. 1–11, 2002.

[25] D. T. Kaplan, M. I. Furman, S. M. Pincus, S. M. Ryan, L. A. Lipsitz, and A. L. Goldberger, "Aging and the complexity of cardiovascular dynamics," *Biophysical Journal*, vol. 59, no. 4, pp. 945–949, 1991.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 368514, 15 pages
http://dx.doi.org/10.1155/2013/368514

Research Article

Constructing Benchmark Databases and Protocols for Medical Image Analysis: Diabetic Retinopathy

Tomi Kauppi,[1] Joni-Kristian Kämäräinen,[2] Lasse Lensu,[1] Valentina Kalesnykiene,[3] Iiris Sorri,[3] Hannu Uusitalo,[4] and Heikki Kälviäinen[1]

[1] *Machine Vision and Pattern Recognition Laboratory, Department of Mathematics and Physics, Lappeenranta University of Technology (LUT), Skinnarilankatu 34, FI-53850 Lappeenranta, Finland*
[2] *Department of Signal Processing, Tampere University of Technology, Korkeakoulunkatu 10, FI-33720 Tampere, Finland*
[3] *Department of Ophthalmology, University of Eastern Finland, Yliopistonranta 1, FI-70211 Kuopio, Finland*
[4] *Department of Ophthalmology, University of Tampere, Biokatu 14, FI-33520 Tampere, Finland*

Correspondence should be addressed to Lasse Lensu; lasse.lensu@lut.fi

Received 25 January 2013; Accepted 26 May 2013

Academic Editor: Carlo Cattani

We address the performance evaluation practices for developing medical image analysis methods, in particular, how to establish and share databases of medical images with verified ground truth and solid evaluation protocols. Such databases support the development of better algorithms, execution of profound method comparisons, and, consequently, technology transfer from research laboratories to clinical practice. For this purpose, we propose a framework consisting of reusable methods and tools for the laborious task of constructing a benchmark database. We provide a software tool for medical image annotation helping to collect class label, spatial span, and expert's confidence on lesions and a method to appropriately combine the manual segmentations from multiple experts. The tool and all necessary functionality for method evaluation are provided as public software packages. As a case study, we utilized the framework and tools to establish the DiaRetDB1 V2.1 database for benchmarking diabetic retinopathy detection algorithms. The database contains a set of retinal images, ground truth based on information from multiple experts, and a baseline algorithm for the detection of retinopathy lesions.

1. Introduction

Image databases and expert ground truth are regularly used in medical image processing. However, it is relatively common that the data is not public, and, therefore, reliable comparisons and state-of-the-art surveys are difficult to conduct. In contrast to, for example, biometrics including face, iris, and fingerprint recognition, the research has been driven by public databases and solid evaluation protocols. These databases have been extended and revised resulting in continuous pressure for the development of better methods. For every medical application, it should be an acknowledged scientific contribution to provide a set of images, collect accurate and reliable ground truth for the images, and devise a meaningful evaluation protocol. Once this pioneering work has been done, it sets an evaluation standard for a selected problem.

We have set our primary goal to the automatic detection of diabetic retinopathy [1] which is very well motivated since diabetes has become one of the most rapidly increasing health threats worldwide [2, 3]. Since the retina is vulnerable to microvascular changes of diabetes and diabetic retinopathy is the most common complication of diabetes, retinal imaging is considered a noninvasive and painless mean to screen and monitor the progress of the disease [4]. Since these diagnostic procedures as well as regular monitoring of state of diabetes require the attention of medical personnel, for example, GP and ophthalmologists, the workload and shortage of personnel will eventually exceed the current resources for screening. To cope with

these challenges, digital imaging of the eye fundus, and automatic or semiautomatic image analysis algorithms based on image processing and computer vision techniques provide a great potential. For this, suitable retinal image databases containing well-defined and annotated ground truth are needed.

In this work, our main contributions are (1) an image annotation tool for medical experts, (2) a public retinal image database with expert annotations, (3) a solid evaluation framework for the image analysis system development and comparison (Figure 1), and (4) image-based and pixel-based evaluation methods. We particularly focus on constructing benchmark databases and protocols. We have experienced that developing databases from scratch is demanding, laborious, and time consuming. However, certain tasks occur repeatedly and are reusable as such. Here, we discuss the related practical issues, point out and solve repeated occurring subtasks, and provide the solutions as open-source tools on our website. In the experimental part, we utilize the proposed framework and construct a revised version of the diabetic retinopathy database DiaRetDB1 originally published in [5, 6], and later discussed in [7].

The paper is organized as follows: in Section 2, we discuss medical benchmarking in general, provide relevant guidelines, and briefly survey the related works. In Section 3, we discuss collecting patient images and the spatial ground truth. We propose a portable data format for the ground truth, and represent and solve the problem of fusing multiple expert annotations. In Section 4, we discuss evaluation practices in general, and provide an evaluation approach based on the standard ROC analysis. We evaluate our color-cue-based detection method (baseline) by using the constructed database. In Section 5, we utilize the given results and tools to establish the diabetic retinopathy evaluation and benchmarking database DiaRetDB1 V2.1, and we draw the conclusions in Section 6.

2. Benchmarking in General and Previous Work

Public image databases for benchmarking purposes are essential resources in the development of image analysis algorithms and help medical imaging researchers evaluate and compare state-of-the-art methods. Eventually, this leads to the development of better algorithms and, consequently, will support technology transfer from research laboratories to clinical practice. However, the public availability of image databases is limited because of the amount of work needed to make internal data publicly available, including the ground truth annotation and the privacy protection of the patient information. Therefore, reliable comparisons and state-of-the-art surveys are difficult to perform. In this section, a benchmarking framework is described that provides guidelines on how to construct benchmarking image databases with a particular emphasis on retinal image analysis. The benchmarking framework comprises three important requirements: (1) patient images, (2) the ground truth, and (3) an evaluation protocol.

2.1. Key Questions in Constructing Benchmarks. Thacker et al. [10] studied the performance characterization of computer vision methods. They provide good examples which are easily transferable to applications of medical image processing. The results in [10] can be utilized in every step of the method development, but we set special attention to the final diagnosis, that is, the subject-wise decision making directly serving the clinical work. In other words, the framework omits the development and research phase evaluations and constructs the good practices to evaluate the performance of retinal image analysis algorithms. For that purpose, the eight general considerations adopted from [10] are addressed and referred to as the key questions.

C1: "How is testing currently performed?" If a commonly used database and protocol are available, their validity for the development and evaluation needs to be examined. In the worst case, a new database needs to be constructed for which the proposed framework can be useful.

C2: "Is there a data set for which the correct answers are known?" Such a data set can be used to report the results in accordance to other studies. This enables method comparison.

C3: "Are there data sets in common use?" See C1 and C2. Common data sets facilitate fair method comparison.

C4: "Are there experiments which show that algorithms are stable and work as expected?" These experiments can be realized if representative data and expert ground truth are available.

C5: "Are there any strawman algorithms?" If a strawman algorithm is included in the database, it defines the baseline performance for other methods. In this paper, we call these kinds of baseline methods as strawman algorithms.

C6: "What code and data are available?" By publishing the method's code or at least executable version of it, other research groups can avoid laborious reimplementation.

C7: "Is there a quantitative methodology for the design of algorithms?" This depends on the medical problem, but the methodology can be typically devised by following corresponding clinical work and practices. Understanding of the medical practitioners' task which should be assisted or automated provides a conceptual guideline. If the database is correctly built to reflect the real-world conditions, then the database implicitly reflects the applicability of the algorithm's design to the problem.

C8: "What should we be measuring to quantify performance? which metrics are used?" At least in the image-wise (subject-wise) experiments, the receiver operating characteristic (ROC) curve is in accordance with the medical practice, where the sensitivity and specificity values are in common use. The ROC curve, also known as ROC analysis, is a widely used tool in medical community for visualizing and comparing

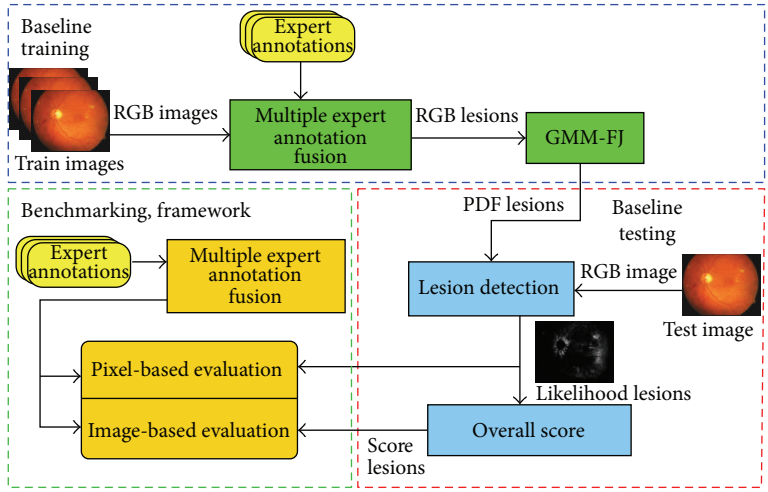


Figure 1: A framework for constructing benchmark databases and protocols [1].

methods based on their performance [11]. It is a graphical representation that describes the trade-off between the sensitivity and specificity (e.g., correctly classified normal images versus correctly classified abnormal images). In the curve, the x-axis is defined as 1 − specificity, and the y-axis is directly the sensitivity [12].

In general, C1 ∈ C2 ∈ C3, which means that if there is a commonly used data set in the form of, for example, a benchmark database, the answers to C1 and C2 are known. Similarly, C4 ∈ C5 ∈ C6 defines the maturity of the existing solutions. In the case where the data and code are both available and have been shown to work by achieving the required sensitivity and specificity rates, the solution is at a mature level and true clinical experiments can be started. C7 is a general guideline for the design to find an acceptable work flow for a specific problem, and C8 sets the quantitative and meaningful performance measures.

2.2. Requirements for Benchmarking. Benchmarking image databases in retinal imaging require three mandatory components: (1) patient images, (2) ground truth by domain experts, and (3) an evaluation protocol. Additional components, such as a baseline algorithm, provide notable additional value, but in the following, the three mandatory components are discussed.

2.2.1. True Patient Images. True patient images carry information which is meaningful for solving a given problem; that is, algorithms which work with these images are expected to perform well also in practice. The images can be recorded using alternative subjects, such as animals that are physiologically close to humans, and disease-related lesions can be produced artificially by using various substances. These are standard practices in medical research, but before drawing any general conclusions, their relevance and accuracy to the real world must be carefully verified. With true patient images, the results are biased by the distribution of database images with respect to the specific real population. The collection and selection of images are further discussed in Section 3. The true patient image requirement concerns the key questions C2, C3, C4, and C6.

2.2.2. Ground Truth Given by Experts. Ground truth must be accurate and reliable in the sense that it is statistically representative over experts. In the field of retinal image processing, it is advisable that the tools for ground truth annotation are provided by computer vision scientists, but the images are selected and annotated by medical experts specialized in the field. It is also clear that the ground truth must be independently collected from multiple experts. This can be laborious and expensive, but it enables statistical studies of reliability. In the case of multiple experts, disambiguation of the data is often necessary prior to the application of machine learning methods. Collecting the ground truth from experts concerns the key questions C2, C3, C4, and C6.

2.2.3. Evaluation Protocol. A valid evaluation protocol providing quantitative and comparable information is essential for reliable performance evaluations. Most articles related to retinal image analysis report the sensitivity and specificity separately, but they are meaningless metrics unless a method can produce superior values for both. The golden standard in similar problems is the ROC analysis. The approach is essentially the same as reporting the sensitivity and specificity but provides the evaluation result over all possible combinations of these values. It turns out that in benchmarking, the comparison of ROC curves is problematic, and, therefore, specific well-justified operation points or the area under curve (AUC) can be used as a single measure. This issue is further discussed in Section 4. In addition to the evaluation protocol, a baseline method (C5) or at least the results with the baseline method are helpful since they set the performance level which new methods should clearly outperform. From another viewpoint, the best reported results by using a commonly accepted database set the state of the art.

Table 1: Summary of the current state of the reference image databases in terms of the key questions addressed in Section 2.1.

Key questions	STARE (vessel)	STARE (disc)	DRIVE	MESSIDOR	CMIF	ROC	REVIEW
C2: "Is there a data set for which the correct answers are known?"	x		x	x		x	x
C3: "Are there data sets in common use?"	x	x	x	x	x	x	x
C4: "Are there experiments which show algorithms are stable and work as expected?"	x		x			x	
C5: "Are there any strawman algorithms?"	x	x	x				
C6.1: "What code is available?"						x	
C6.2: "What data is available?"	x	x	x	x	x	x	x
C7: "Is there a quantitative methodology for the design of algorithms?"							
C8.1: "What should we be measuring to quantify performance?"	x	x	x			x	x
C8.2: "What metrics are used?"		x	x			x	x
$\sum$	6	5	7	3	2	7	5

The evaluation protocol requirement concerns the key questions C1, C4, C7, and C8.

2.3. Eye Disease Databases. This section describes the most important public benchmarking databases in retinal image analysis. The database review provides a short description for each database, where the key questions C1–C8 addressed in Section 2.1 are used to highlight the main properties. Since each database is publicly available, they are expected to be in common use (C3). See Table 1 for a short summary.

STARE (structured analysis of the retina) [17] is one of the most used reference image database in the literature (C3, C4) for comparing blood vessel detection and optic disc localization algorithms. The STARE website [17] provides 20 images with pixel-wise hand-labeled ground truth for blood vessel detection (C2) and 81 images for optic disc localization without ground truth. The performance of blood vessel detection is measured using the ROC curve analysis, where the sensitivity is the proportion of correctly classified blood vessel pixels and the specificity is the proportion of correctly classified normal pixels (C8.1) [18]. In the evaluation of optic disc localization, the proportion of correctly localized optic discs indicates that the performance and the localization are successful if the center of optic disc generated by the algorithm is within 60 pixels from the ground truth (C8) [19]. The evaluation procedures for both data sets are published with vessel detection algorithm and baseline results (C5) [18, 19].

DRIVE (digital retinal images for vessel extraction) [20, 21] is another well-known reference database for blood vessel detection (C3), which contains 40 retinal images (C6.2) with manually segmented pixel-wise ground truth (C2, C6.2). The manual segmentation task was divided between three medical experts, and the database was published along with vessel detection algorithm (C5) [21]. The detection performance is measured similarly as in the STARE database, that is, comparing the sensitivity to the specificity (C8.1) from which the area under curve (AUC) is computed to produce the final measure for the algorithm comparison (C8.2) [20, 21]. In addition, the authors implemented and internally evaluated a number of blood vessel detection algorithms from various research groups and the results were published in [22] and on the DRIVE database website (C4) [20].

MESSIDOR (methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology) [23] is a reference image database collected to facilitate computer-assisted image analysis of diabetic retinopathy. Its primary objectives are to enable evaluation and comparison of algorithms for analyzing the severity of diabetic retinopathy, prediction of the risk of macular oedema, and indexing and managing image databases, that is, support image retrieval. For the evaluation, the MESSIDOR database website [23] provides 1200 images (C6.2) with image-wise severity grading (C2, C6.2) from three ophthalmologic departments including descriptions for the severity grading. It is noteworthy to mention that the severity grading is based on the existence and number of diabetic lesions and their distance from the macula.

CMIF (collection of multispectral images of the fundus) [24, 25] is a public multispectral retinal image database. The spectral images were obtained by implementing a "filter wheel" into a fundus camera containing a set of narrow-band filters corresponding to the set of desired wavelengths [25]. The database itself consists of normal and abnormal images (C6.2) spanning a variety of ethnic backgrounds covering 35 subjects in total [25]. As such, the database is not ready for benchmarking, but it provides a new insight into retinal pathologies.

ROC (retinopathy online challenge) [26, 27] follows the idea of asynchronous online algorithm comparison proposed by Scharstein and Szeliski [28] for stereo correspondence algorithms (Middlebury Stereo Vision Page), where a web evaluation interface with public evaluation data sets ensures that the submitted results are comparable. The research groups download the data set, they submit their results in the required format, and the results are evaluated by the web evaluation system. Since the evaluation is fully automatic, the research groups can submit and update their results continuously. In the current state, the ROC database website [26]

Table 2: Summary of the DiaRetDB1 V2.1 database in terms of the key questions addressed in Section 2.1.

Key questions	DiaRetDB1 V2.1
C2: "Is there a data set for which the correct answers are known?"	Yes
C3: "Are there data sets in common use?"	Yes (publicly available at [13])
C4: "Are there experiments which show algorithms are stable and work as expected?"	Experimental results reported in Section 4.4
C5: "Are there any strawman algorithms?"	No, but the baseline algorithm sets the baseline results for the DiaRetDB1 database
C6.1: "What code is available?"	Functionality for reading/writing images and ground truth, strawman algorithm, and annotation software (publicly available at [13, 14])
C6.2: "What data is available?"	Images and ground truth (XML) (publicly available at [13])
C7: "Is there a quantitative methodology for the design of algorithms?"	No, but medical practice is used as a guideline at each development step
C8.1: "What should we be measuring to quantify performance?"	Image- and pixel-based ROC analysis (description in Section 4)
C8.2: "What metrics are used?"	Equal error rate (EER) defined in Section 4

provides 100 retinal images (C6.2), a ground truth (C2, C6.2) and an online evaluation system for microaneurysms, and the evaluation results for a number of detection algorithms (C4). The algorithm performance is measured by comparing the sensitivity (the proportion of correctly classified lesions) against the average number of false positives in the image, that is, free-response receiver operating characteristic curve (FROC) (C8.1) [27]. The sensitivities of predefined false positive points are averaged to generate the final measure for algorithm comparison (C8.2) [27]. The annotations were gathered from 4 medical experts by marking the location, approximate size, and confidence of the annotation. Consensus of two medical experts was required for a lesion to be selected to the ground truth.

REVIEW (retinal vessel image set for estimation of widths) [29, 30] is a new reference image database to assess the performance of blood vessel width measurement algorithms. To characterize the different vessel properties encountered in the retinal images, the database consists of four image sets: (1) high-resolution image set (4 images); (2) vascular disease image set (8 images); (3) central light reflex image set (2 images), and (4) kick point image set (2 images) (C6.2). The REVIEW database concentrates on high-precision annotations, and, therefore, it provides only segments of blood vessels and not the whole network. To achieve high precision, the human observers used a semiautomatic tool to annotate a series of image locations from which the vessel widths were automatically determined [30]. The annotations were gathered from three medical experts, and the mean vessel width was defined as the ground truth (C2, C6.2). In the evaluation, the performance is measured using an unbiased standard deviation of the width difference between the algorithm-estimated vessel widths and the ground truth (C8) [30].

In general, most of the reference databases reach the minimal requirements for benchmarking image analysis algorithms; that is, they provide true patient images, ground truth from experts, and an evaluation protocol (Table 1). In some cases, the usability is already at a mature level, for example, in the case of the web evaluation system in the ROC database. The primary shortcomings appear to be related to the availability of software (C6.1) and how the algorithm's design for the medical problem is observed (C7). By publishing source codes or an executable, other researchers can avoid laborious reimplementation and if the database is correctly built to reflect real-world conditions, then the database implicitly reflects the applicability of the algorithm's design to the problem. The database properties in terms of the key questions are summarized in Table 1 and for comparison the proposed DiaRetDB1 database properties are summarized in Table 2. The framework for constructing benchmark databases and protocols has been summarized in Figure 1. The details of the framework are discussed in the next sections.

3. Patient Images and Ground Truth

3.1. Collecting Patient Images. The task of capturing and selecting patient images should be conducted by medical doctors or others specifically trained for photographing the eye fundus. With the images, there are two issues which should be justified: (1) distribution correspondence with the desired population and (2) privacy protection of patient data.

In DiaRetDB1, the ophthalmologists wanted to investigate the accuracy of automatic methods analyzing retinal images of patients who are diagnosed with having diabetes. Consequently, the images do not correspond to the actual severity or prevalence of diabetic retinopathy in the Finnish population but provide clear findings for automated detection methods. The data is, however, clinically relevant since the studied subpopulation is routinely screened by Finnish primary health care.

The privacy protection of patient data is a task related to the ethics of clinical practice, medical research, and also data security. A permission for collecting and publishing the data must be acquired from a corresponding national organization

(e.g., national or institutional ethical committee) and from the patients themselves. Moreover, all data must be securely stored; that is, all patient information, such as identifying metadata, must be explicitly removed from images which are to be used in a public database. In DiaRetDB1, the retinal images were acquired using a standard fundus camera and its accompanying software. The acquired images were converted to raw bitmaps and then saved to portable network graphics (PNG) format using lossless compression. The raw bitmaps contained nothing but the pixel data which guaranteed the removal of hidden metadata.

3.2. Image Annotations as the Ground Truth. In general, the image annotations are essential for training supervised algorithms, as well as for their evaluation and comparison. Such information is typically collected by manually annotating a set of images. In face recognition, for example, a ground truth contains identifiers of persons in the images and often also the locations of facial landmarks, such as eye centers, which can be very useful in training the methods. Commonly, simple tailored tools are used to collect the data, but also generic applications are available for problems which require an exhaustive amount of image data, for example, LabelMe [31] Web tool for annotating visual object categories. Annotating medical images is not an exception, but two essential considerations apply: (1) annotations must be performed by clinically qualified persons (specialized or specializing medical doctors, or other trained professionals for specific tasks), denoted as "experts" and (2) the ground truth should include annotations from multiple experts.

A more technical problem is to develop a reusable tool for the annotation task. To avoid biasing the results, the experts should be given minimal guidance for their actual annotation work. Basic image manipulation, such as zoom and brightness control, for viewing the images is needed, and a set of geometric primitives are provided for making the spatial annotations. In LabelMe [31], the only primitive is polygon region defined by an ordered set of points. A polygon can represent an arbitrarily complex spatial structure, but ophthalmologists found also the following primitives useful: small circle, which can be quickly put on a small lesion, and circle area and ellipse area which are described by their centroid, radius/radii, and orientation (ellipse). The system also requires at least one representative point for each lesion. This point should represent the most salient cue, such as color or texture, that describes the specific lesion. Furthermore, a confidence selection from the set of three discrete values, low, moderate, or high, is required for every annotation. The experts are allowed to freely define the types of annotations, that is, the class labels for the lesion types, but typically it is preferable to agree with the labels beforehand (e.g., in DiaRetDB1: hard exudates, soft exudates, microaneurysms, and haemorrhages). An important design choice is related to the usability of the tool with respect to its graphical user interface (GUI). For example, the GUI should not use colors which distract the annotators from image content.

The development of an annotation tool may take undesirable amount of research time and resources. To help other researchers in this task the tool is available upon request as Matlab M-files and as a Windows executable. Users have full access to the source code which enables tailoring of the tool for their specific needs. The default graphical user interface (GUI) is shown in Figure 2.

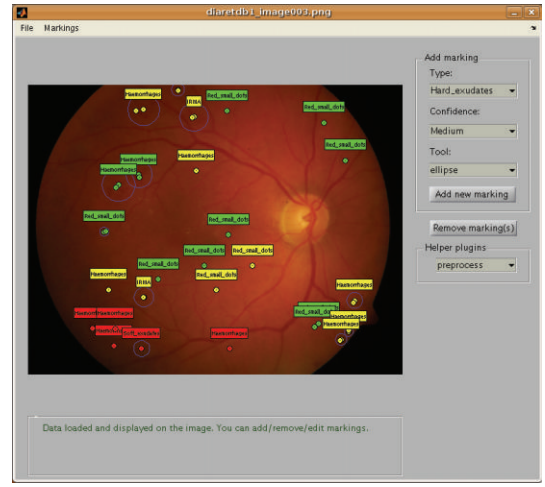


Figure 2: Graphical user interface of the image annotation tool [1].

3.3. Data Format for Medical Annotations. To store the annotations and to be able to restore their graphical layout, the data format must be defined. The data is naturally structured, and, therefore, structural data description languages are preferred. Several protocols for describing medical data exist, such as HL7 based on the extensible markup language (XML) [32], but these are complex protocols designed for patient information exchange between organizations and information systems. Since the requirements for benchmarking databases in general are considerably less comprehensive, a light-weight data format based on the XML data description language is adopted. Instead of the XML Schema document description, a more compact and, consequently, more interpretable Document Type Definition (DTD) description is applied. The used format is given in Listing 1.

3.4. Fusion of Manual Segmentations from Multiple Experts. A desired characteristic of collecting the ground truth for medical images is that one or several experts provide information on the image contents such as the disease-related lesions. Since there can exist inconsistencies in the case of a single expert (e.g., due to changing criteria while performing the annotation work) and nobody can be considered as the unparalleled expert, the use of several experts is preferred. Only in clear cases, however, the experts fully agree on the interpretation of the visible information. Since the early signs of retinopathy are very subtle changes in the images, it is necessary to develop a method to appropriately combine the expert information which is only partially coherent. To design such a method, the important questions relevant to training, evaluating, and benchmarking by using the database are as follows: (1) how to resolve inconsistencies

```
<!ELEMENT imgannotooldata (header, markinglist)>
<!ELEMENT header (creator, software?,
                  affiliation?, copyrightnotice)>
<!ELEMENT creator (#PCDATA)>
<!ELEMENT software (#PCDATA)>
<!ATTLIST   software version CDATA #REQUIRED>
<!ELEMENT affiliation (#PCDATA)>
<!ELEMENT copyrightnotice (#PCDATA)>
<!ELEMENT imagename (#PCDATA)>
<!ELEMENT imagesize (width, height)>
<!ELEMENT width (#PCDATA)>
<!ELEMENT height (#PCDATA)>
<!ELEMENT markinglist (marking*)>
<!ELEMENT marking ((polygonregion |
                    circleregion | ellipseregion),
                   representativepoint+, confidencelevel, markingtype)>
<!ELEMENT centroid (coords2d)>
<!ELEMENT polygonregion (centroid, coords2d,
                         coords2d, coords2d+)>
<!ELEMENT circleregion (centroid, radius)>
<!ELEMENT ellipseregion (centroid, radius, radius, rotangle)>
<!ELEMENT representativepoint (coords2d)>
<!ELEMENT coords2d (#PCDATA)>
<!ELEMENT radius (#PCDATA)>
<!ATTLIST   radius direction CDATA #REQUIRED>
<!ELEMENT rotangle (#PCDATA)>
<!ELEMENT markingtype (#PCDATA)>
<!ELEMENT confidencelevel (#PCDATA)>]>
```

Listing 1: DTD definition.

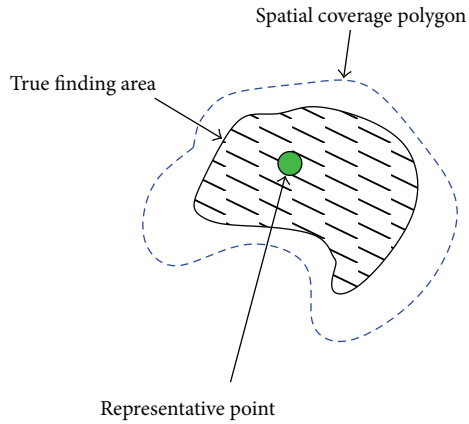


Figure 3: The available expert information in the DiaRetDB1 database. The expert's subjective confidence for the annotation is defined as follows: 100%, >50%, and <50% [1].

in the annotations from a single expert and (2) how to fuse equally trustworthy (no prior information on the superiority of the experts related to the task) information from multiple experts?

In our data format, the available expert information is the following (Figure 3): (1) spatial coverage (polygon area), (2) representative point(s) (small circle areas), and (3) the subjective confidence level. The representative points are distinctive "cue locations" that attracted the expert's attention to the specific lesion. The confidence level with a three-value scale describes the expert's subjective confidence for the lesion to represent a specific class (lesion type) as shown in Figure 4.

Combining the manual segmentations from multiple experts was originally studied in [9]. In the study, the area intersection provided the best fusion results in all experimental setups and is computed in a straightforward manner as the sum of expert-annotated confidence images divided by the number of experts. For DiaRetDB1, the fused confidence with the threshold 0.75 yielded the best results [1], resolving the inconsistencies of annotations either from a single expert or multiple expert cofusion problems.

The area intersection is intuitive and the result is based on processing the whole image ensemble. However, the threshold was selected with the baseline method, which undesirably tied the training and evaluation together. Therefore, the combination problem was revised in [8].

The most straightforward combination procedure is averaging where the expert segmentations are spatially averaged for each image and lesion type. In this procedure, the given confidence levels are used, and the only requirement for the confidence scale is that it is monotonically increasing. The average confidence image corresponds to the mean expert opinion, but it has two disadvantages: (1) it does not take

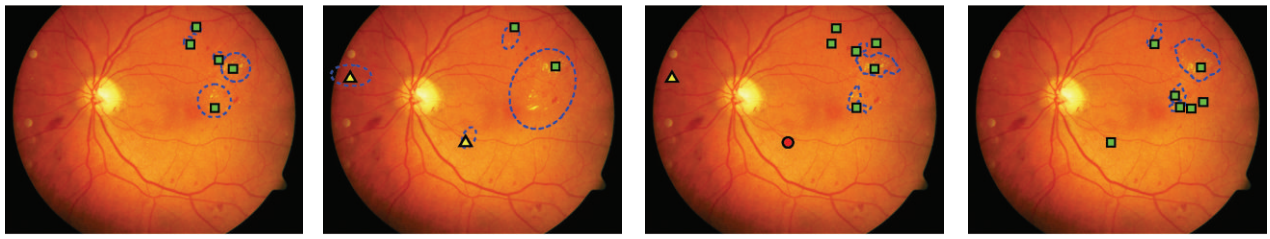
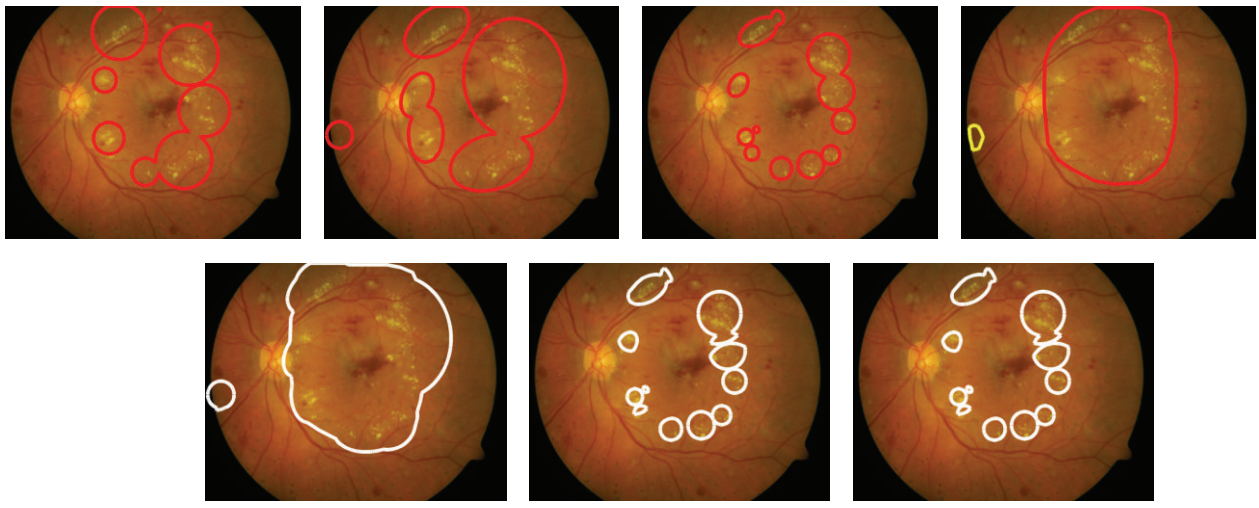


Figure 4: Four independent sets of spatial annotations (contours and representative points) for the same lesion type (hard exudates). The representative point markers denote the confidence level (*square* = 100%, *triangle* > 50%, and *circle* < 50%) [1].



(a) (b) (c)

Figure 5: 1st row: DiaRetDB1 expert spatial annotations for the lesion Hard exudate (red: high confidence, yellow: moderate, green: low). 2nd row: the ground truth (white) produced by the original method and (a) minimal and (b) maximal confidence. The disambiguated ground truth by (c) the revised method [8].

into account the possible differences of the experts in their use of the scale and (2) it does not produce binary values for the foreground (lesion of specific type) and background. As a solution, a binary mask can be generated by thresholding the average expert segmentation image. The threshold parameter $\tau \in [0, 1]$ adjusts experts' joint agreement: for $\tau \to 0$, the binary mask approaches *set union* and for $\tau \to 1$ approaches *set intersection* (see Figure 5).

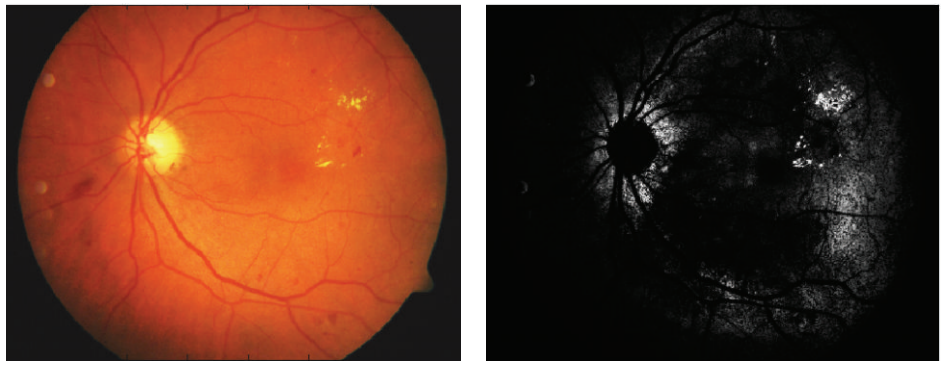
The revised combining method is based on the following principle: *The ground truth should optimally represent the mutual agreement of all experts*. To evaluate the degree of mutual agreement, a performance measure is needed. The performance depends only on two factors: *experts' markings* and the *ground truth*, and, without loss of generality, the measure is expected to output a real number

$$\text{perf}: \left\{I_{\text{exp}_{i,j,n}}, g_{t_{i,j}}\right\} \longrightarrow \mathbb{R}, \tag{1}$$

where expert segmentation masks $I_{\text{exp}_{i,j,n}}$ represents the expert segmentation mask for the input image i, lesion type j, and expert n, g_t is the ground truth, and $\{\cdot\}$ is used to denote that the performance is computed for a set of rated images. Generation of the image-wise ground truth is straightforward: if any of the pixels in the produced $I_{\text{mask}_{i,j}}$ for the lesion j is nonzero, the image is labeled to contain that lesion. A detection ROC curve can be automatically computed from the image-wise ground truth and image scores computed from the expert images. For the image-wise expert scores, we adopted the summax rule described in Section 4: pixel confidences of $I_{\text{exp}_{i,j,n}}$ are sorted, and 1% of the highest values are summed. The average equal error rate (EER point on the ROC curve) was chosen as the performance measure in (1), which can be given in an explicit form:

$$\begin{aligned}&\text{perf}\left(\left\{I_{\text{exp}_{i,j,n}}\right\}, \left\{g_{t_{i,j}}\right\}\right)\\ &\quad = \frac{1}{N}\sum_{n}\text{EER}\left(\left\{\text{summax}_{1\%}\left(I_{\text{exp}_{i,j,n}}\right)\right\}, \left\{I_{\text{mask}_{i,j}}(x, y; \tau)\right\}\right).\end{aligned} \tag{2}$$

A single EER value is computed for each expert n and over all images (i), and then the expert-specific EER values are summed for the lesion type j.



(a) (b)

Figure 6: Pixel-wise likelihoods for Hard exudates produced by the strawman algorithm: (a) original image (hard exudates are the small yellow spots in the right part of the image); (b) "likelihood map" for hard exudates [9].

The utilization of the summax rule is justified as a robust maximum rule by the multiple classifier theory [33]. Also the EER measure can be replaced with any other measure if, for example, prior information on the decision-related costs is available. The only factor affecting the performance in (2) is the threshold τ which is used to produce the ground truth. To maximize the mutual agreement, it is necessary to seek the most appropriate threshold $\hat{\tau}$ providing the highest average performance (EER) over all experts. Instead of a single threshold, lesion-specific thresholds $\hat{\tau}_j$ are determined since different lesions may significantly differ by their visual detectability. The optimal ground truth is equivalent to searching the optimal threshold:

$$\hat{\tau}_j \longleftarrow \underset{\tau_j}{\operatorname{argmin}} \frac{1}{N} \sum_{n} \mathrm{EER}(\cdot, \cdot). \tag{3}$$

A straightforward approach to implement the optimization is to iteratively test all possible values of τ from 0 to 1. Equation (3) maximizes the performance for each lesion type over all experts (N). The optimal thresholds $\hat{\tau}_j$ are guaranteed to produce the maximal mutual expert agreement according to the performance measure perf.

The revised combining method was shown to produce better results when compared to the original method and even to simultaneous truth and performance level estimation (STAPLE) [34]. The full description of the method and comparisons is presented in [8].

4. Algorithm Evaluation

4.1. Evaluation Methodology. The ROC-based analysis perfectly suits to medical decision making, being the acknowledged methodology in medical research [35]. An evaluation protocol based on the ROC analysis was proposed in [6] for image-based (patient-wise) evaluation and benchmarking, and the protocol was further studied in [9]. In clinical medicine, the terms *sensitivity* and *specificity* defined in the range [0%, 100%] or [0, 1] are used to compare methods and laboratory assessments. The sensitivity

$$\mathrm{SN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

depends on the diseased population whereas the specificity

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{5}$$

on the healthy population, defined by true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The x-axis of an ROC curve is 1 − specificity, whereas the y-axis represents directly the sensitivity [12].

It is useful to form an ROC-based quality measure. the quality measures preferred are as follows: The equal error rate (EER) [36] defined as when (SN = SP)

$$\mathrm{SN} = \mathrm{SP} = 1 - \mathrm{EER}, \tag{6}$$

or weighted error rate (WER) [37]

$$\mathrm{WER}\left(\hat{R}\right) = \frac{\mathrm{FPR} + \hat{R} \cdot \mathrm{FNR}}{1 + \hat{R}} = \frac{(1 - \mathrm{SP}) + \hat{R} \cdot (1 - \mathrm{SN})}{1 + \hat{R}}, \tag{7}$$

where $\hat{R} = C_{\mathrm{FNR}}/C_{\mathrm{FPR}}$ is the cost ratio between the false negative rate $\mathrm{FNR} = 1 - \mathrm{SN} = \mathrm{FN}/(\mathrm{TP} + \mathrm{FN})$ and false positive rate $\mathrm{FPR} = 1 - \mathrm{SP} = \mathrm{FP}/(\mathrm{FP} + \mathrm{TN})$. The main difference between the two measures is that EER assumes equal penalties for both false positives and negatives, whereas in the WER, the penalties are adjustable.

In the image-based evaluation, a single likelihood value for each lesion should be produced for all test images. Using the likelihood values, an ROC curve can be automatically computed [9]. If a method provides multiple values for a single image, such as the full-image likelihood map in Figure 6(b), the values must be fused to produce a single score.

4.2. Image-Based Evaluation. The automatic image-based evaluation follows the medical practice where the decisions

(1) **for each** test image **do**
(2) TN ← 0, TP ← 0, FN ← 0, FP ← 0
(3) curr_score ← image score
(4) **for each** test image **do**
(5) **if** curr_score ≥ image score **then**
(6) **if** ground truth assignment = "normal" **then**
(7) TN = TN + 1
(8) **else**
(9) FN = FN + 1
(10) **end if**
(11) **else**
(12) **if** ground truth assignment = "abnormal" **then**
(13) TP = TP + 1
(14) **else**
(15) FP = FP + 1
(16) **end if**
(17) **end if**
(18) **end for**
(19) $\mathrm{SN} = \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$ (Sensitivity)
(20) $\mathrm{SP} = \dfrac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$ (Specificity)
(21) Add new ROC point $(x; y) = (1 - \mathrm{SP}, \mathrm{SN})$
(22) **end for**
(23) Return the final ROC curve (all points)

Algorithm 1: Image-wise evaluation based on image scores.

are "subject-wise." An image analysis system is treated as a black-box which takes an image as the input. If the images are assumed to be either normal or abnormal, the system produces a score that corresponds to the probability of the image being abnormal, and a high score corresponds with high probability. The objective of the image-based evaluation protocol is to generate an ROC curve by manipulating the score values of the test images. The practices were adopted from [38].

Let the image analysis algorithm produced score values for n test images be $\boldsymbol{\zeta}^{\mathrm{im}} = \{\zeta_1^{\mathrm{im}}, \ldots, \zeta_n^{\mathrm{im}}\}$ and let the corresponding image-wise ground truths be $\boldsymbol{\omega}^{\mathrm{im}} = \{\omega_1^{\mathrm{im}}, \ldots, \omega_n^{\mathrm{im}}\}$, where each ω_i^{im} is either "normal" or "abnormal." Then, by selecting a threshold for the score values ($\boldsymbol{\zeta}^{\mathrm{im}}$), the test images can be classified as either normal or abnormal, and the performance expressed in the form of sensitivity and specificity can be determined by comparing the outcome with the corresponding image-wise ground truth ($\boldsymbol{\omega}^{\mathrm{im}}$). If the same procedure is repeated using each test image score as the threshold, the ROC curve can be automatically determined since each threshold generates a (sensitivity, specificity) pair that is a point on the ROC curve. Consequently, the procedure requires that the test images include samples from both populations, normal and abnormal. The image score-based evaluation method is presented in Algorithm 1.

4.3. Pixel-Based Evaluation. To validate a design choice in method development, it can be useful to measure also the spatial accuracy, that is, whether the detected lesions are found in correct locations. Therefore, a pixel-based evaluation protocol which is analogous to the image-based evaluation is proposed. In this case, the image analysis system takes an image as the input and outputs a similar score for each pixel. The objective of the pixel-based evaluation is to generate an ROC curve which describes the pixel-level success.

Let the image analysis algorithm-produced pixel score values for all n pixels in test set be $\boldsymbol{\zeta}^{\mathrm{pix}} = \{\zeta_1^{\mathrm{pix}}, \ldots, \zeta_n^{\mathrm{pix}}\}$ and let the corresponding pixel-wise ground truth be $\boldsymbol{\omega}^{\mathrm{pix}} = \{\omega_1^{\mathrm{pix}}, \ldots, \omega_n^{\mathrm{pix}}\}$, where the $\boldsymbol{\omega}^{\mathrm{pix}}$ is either "normal" or "abnormal." Then, by selecting a global pixel-wise threshold for the pixel score values ($\boldsymbol{\zeta}^{\mathrm{pix}}$), the pixels in all images can be classified to either normal or abnormal. Now, the sensitivity and specificity can be computed by comparing the outcome to the pixel-wise ground truth ($\boldsymbol{\omega}^{\mathrm{pix}}$). If the procedure is repeated using each unique pixel score as the threshold, the ROC curve can be automatically determined. The pixel-wise evaluation procedure is given in Algorithm 2. Note that the abnormal test image pixels contribute to both sensitivity and specificity, whereas the normal images only contribute to the specificity.

The evaluation forms a list of global pixel-wise scores from the test image pixel scores which determines the score thresholds. The use of all unique pixel scores in the test images is time consuming if the number of images in the test set is large or high-resolution images are used. The problem can be overcome by sampling the test image pixel scores.

(1) Form a list of tested pixel scores
(2) **for each** tested pixel score (curr_pix_score) **do**
(3) TN ← 0, TP ← 0, FN ← 0, FP ← 0
(4) **for each** test image **do**
(5) **for each** test image pixel score **do**
(6) **if** curr_pix_score ≥ pixel score **then**
(7) **if** ground truth pixel assignment = "normal" **then**
(8) TN = TN + 1
(9) **else**
(10) FN = FN + 1
(11) **end if**
(12) **else**
(13) **if** ground truth pixel assignment = "abnormal" **then**
(14) TP = TP + 1
(15) **else**
(16) FP = FP + 1
(17) **end if**
(18) **end if**
(19) **end for**
(20) **end for**
(21) $\mathrm{SN} = \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$ (Sensitivity)
(22) $\mathrm{SP} = \dfrac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$ (Specificity)
(23) Add new ROC point $(x; y) = (1 - \mathrm{SP}, \mathrm{SN})$
(24) **end for**
(25) Return the final ROC curve (all points)

Algorithm 2: Pixel-wise evaluation based on pixel scores.

(1) Extract colour information (r, g, b) of the lesion from the train set images (Section 3.4).
(2) Estimate $p(r, g, b \mid \text{lesion})$ from the extracted color information using a Gaussian mixture model determined by using the Figueiredo-Jain method [15, 16].
(3) Compute $p(r, g, b \mid \text{lesion})$ for every pixel in the test image (repeat step for every test image in the test set).
(4) Evaluate the performance (Section 4).

Algorithm 3: Strawman algorithm.

To preserve the test set's pixel score distribution, the global threshold scores can be devised as follows: (1) sort all the unique pixel scores in an ascending order to form an ordered sequence L and (2) compose the new reduced sequence of pixel scores L_{sampled} by selecting every jth likelihood in L.

4.4. The Strawman Algorithm. We provide a baseline method in the form of a strawman algorithm. The algorithm is based on the use of photometric cue as described in Algorithm 3 [9].

The score fusion in the strawman algorithm is based on the following reasoning: if we consider M medical evidence (features) extracted from the image, $\mathbf{x}_1, \ldots, \mathbf{x}_M$, where each evidence is a vector, then we can denote the score value of the image as $p(\mathbf{x}_1, \ldots, \mathbf{x}_M \mid \text{abnormal})$. The joint probability is approximated from the classification results (likelihoods) in terms of decision rules using the combined classifier theory (classifier ensembles) [33]. The decision rules for deriving the score were compared in the study [9] where the rules were devised based on Kittler et al. [33] and an intuitive rank-order-based rule "summax." The rule defines the image score $p(\mathbf{x}_1, \ldots, \mathbf{x}_M \mid \text{abnormal})$ using the compared decision rules when the prior values of the population characteristics are equal ($P(\text{normal}) = P(\text{abnormal})$) as follows:

$$\mathrm{SCORE}_{\text{summax}} = \sum_{m \in N_{Y\%}} p\left(\mathbf{x}_m \mid \text{abnormal}\right), \tag{8}$$

where $N_{Y\%}$ are the indices of $Y\%$ top-scoring pixel scores. Experimenting also with the max, mean, and product rules, strong empirical evidence supports the rank-order-based sum of maxima (summax; proportion fixed to 1%) [9].

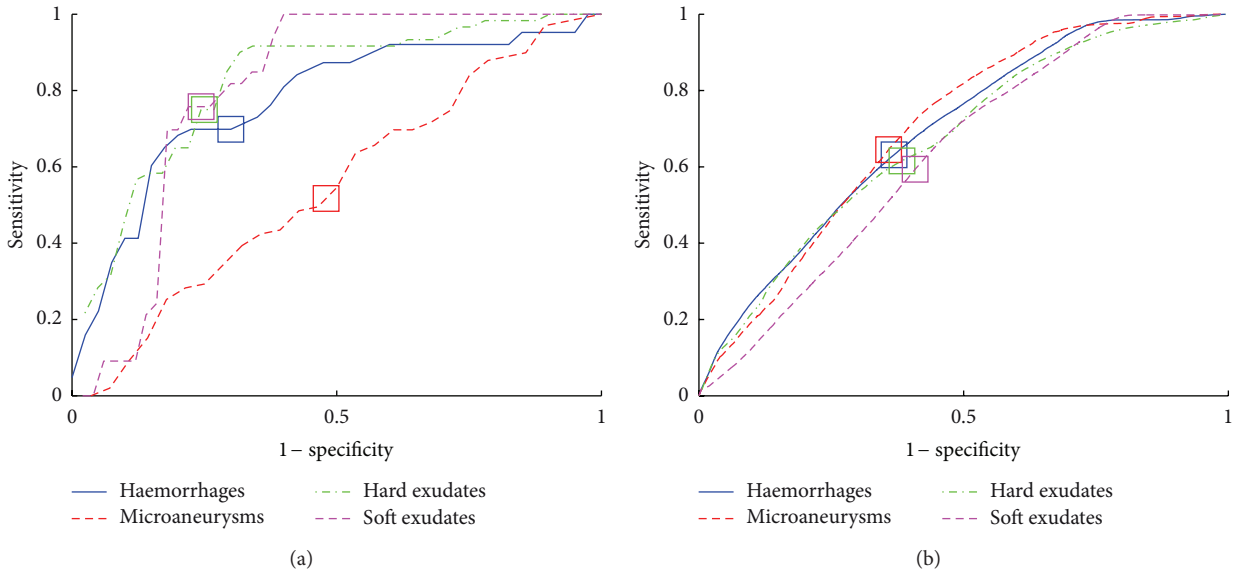


Figure 7: The ROC curves for the DiaRetDB1 strawman algorithm using the original ground truth (*squares* denote the EER points): (a) image based; (b) pixel based. Note the clear difference with microaneurysms as compared to the revised ground truth in Figure 8.

Table 3: The minimum, maximum, and average EER (5 random iterations) for the baseline method and evaluation protocol when using DiaRetDB1. The results include the original and the revised ground truth [8].

	Haemorrhage (HA)			Hard exud. (HE)			Microaneurysm (MA)			Soft exud. (SE)			Overall
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	
In [9]	0.233	0.333	0.273	0.200	0.220	0.216	0.476	0.625	0.593	0.250	0.333	0.317	0.349
In [8] (min)	0.263	0.476	0.322	0.250	0.250	0.250	0.286	0.574	0.338	0.333	0.333	0.333	0.311
In [8] (max)	0.263	0.476	0.322	0.250	0.250	0.250	0.386	0.574	0.338	0.200	0.268	0.241	**0.288**

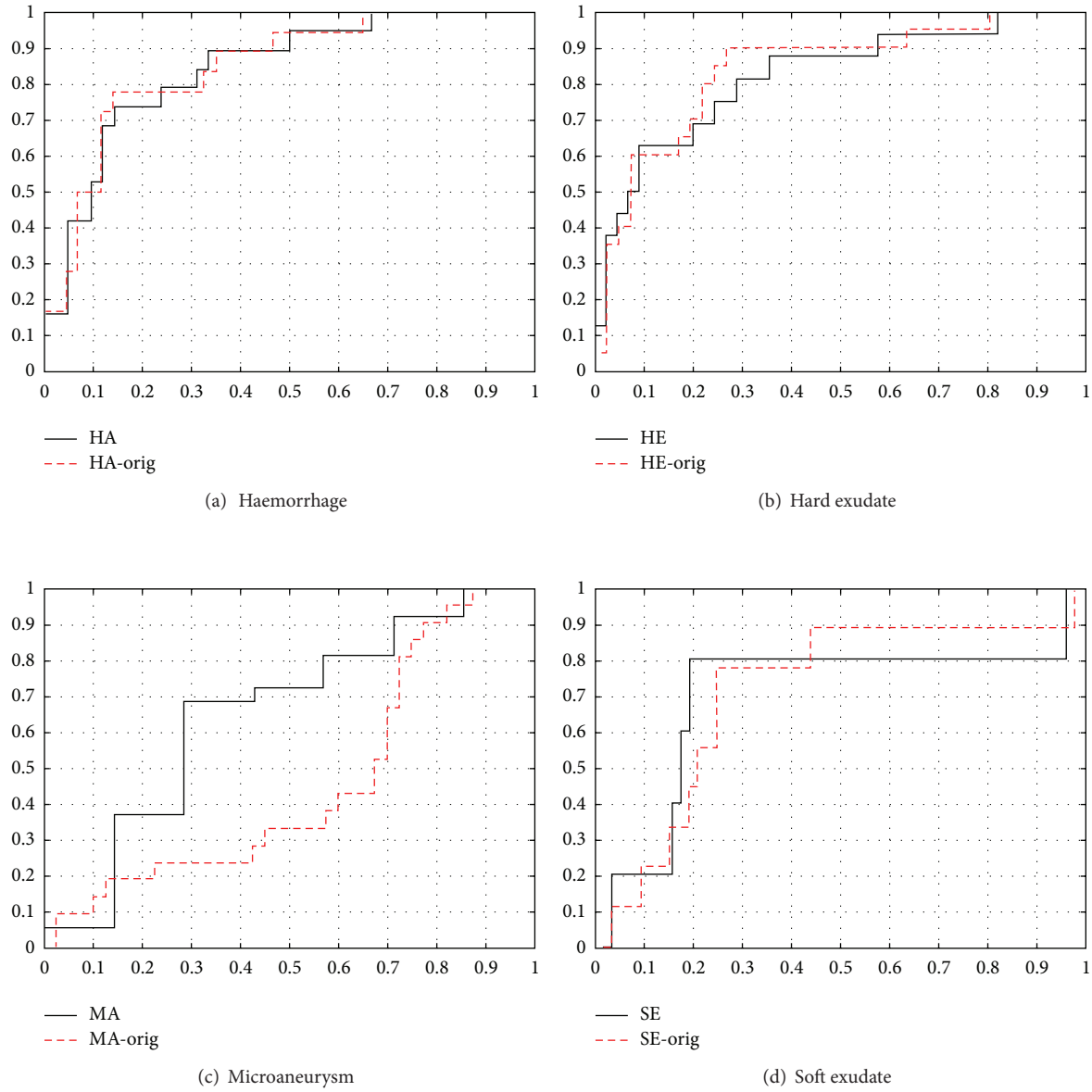
The achieved results for DiaRetDB1 are shown in Figure 7 (ROC curves) and in Table 3 (EER values). The performance is reported by using the EER which is justified since EER represents a "balanced error point" on the ROC curve and allows comparison to the previous works.

To quantify the effect of the revised method for combining the expert information, results from a comparison are shown in Table 3. It should be noted that the experiment is independent of the one presented above. The original confidence threshold (0.75) in [9] was not optimal for any of the lesion types and was clearly incorrect for haemorrhages (HA, 0.60) and microaneurysms (MA, 0.10). The underlined values in the table are the best achieved performances. The average performance for all lesion types significantly varies depending on the threshold.

The minimum and maximum thresholds for the revised combining method produce equal results except in the case of soft exudates, for which the maximum in the equally performing interval (1.0) is clearly better. The main difference from the original DiaRetDB1 method occurs with microaneurysms, since the optimal threshold (0.1) significantly differs from the original (0.75). For haemorrhages, the original result was too optimistic since the optimal confidence yields worse minimum and average EER. On average, the revised method provided 11–17% better performance. The related ROC curves are shown in Figure 8.

5. Case Study: DiaRetDB1 Diabetic Retinopathy Database and Protocol V2.1

The authors have published two medical image databases with the accompanied ground truth: DiaRetDB0 and DiaRetDB1. The work on DiaRetDB0 provided us with essential information on how diabetic retinopathy data should be collected, stored, annotated, and distributed. DiaRetDB1 was a continuation to establish a better database for algorithm evaluation. DiaRetDB1 contains retinal images selected by experienced ophthalmologists. The lesion types of interest were selected by the medical doctors (see Figure 9): microaneurysms (distensions in the capillary), haemorrhages (caused by ruptured or permeable capillaries), hard exudates (leaking lipid formations), soft exudates (microinfarcts), and neovascularisation (new fragile blood vessels). These lesions



(a) Haemorrhage (b) Hard exudate

(c) Microaneurysm (d) Soft exudate

Figure 8: ROC curves for the DiaRetDB1 baseline method using the original and revised (max) method to generate the training and testing data [8].

are signs of mild, moderate, and severe diabetic retinopathy, and they provide evidence also for the early diagnosis. The images were annotated by four independent and experienced medical doctors inspecting similar images in their regular work.

The images and ground truth are publicly available on the Internet [13]. The images are in PNG format, and the ground truth annotations follow the XML format. Moreover, we provide a DiaRetDB1 kit containing full Matlab functionality (M-files) for reading and writing the images and ground truth, fusing expert annotations, and generating image-based evaluation scores. The whole pipeline from images to evaluation results (including the strawman algorithm) can be tested using the provided functionality. The annotation software (Matlab files and executables) is also available upon request.

6. Conclusions

We have discussed the problem of establishing benchmark databases for the development of medical image analysis. We have pointed out the importance of commonly accepted and used databases. We have proposed the framework for constructing benchmark databases and protocols for diabetic retinopathy in medical image analysis. We have built reusable tools needed to solve the important subtasks, including

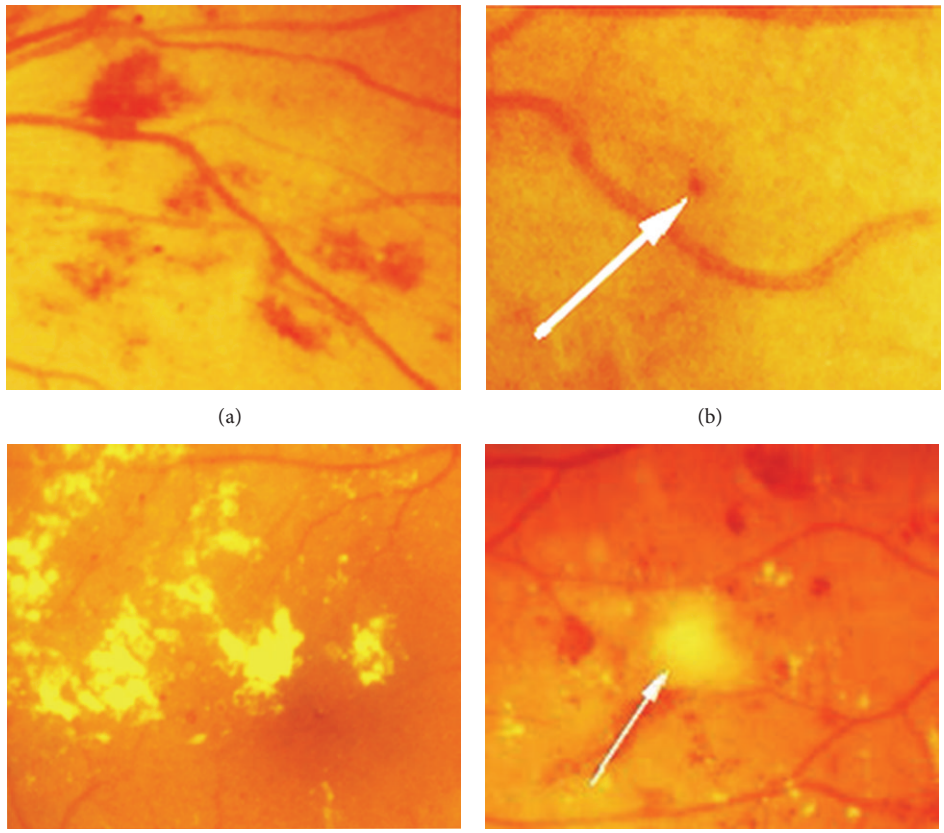


Figure 9: Abnormal retinal findings caused by the diabetes (best viewed in colour): (a) haemorrhages; (b) microaneurysms (marked with an arrow); (c) hard exudates; (d) soft exudate (marked with an arrow) [6].

the annotation tool for collecting the expert knowledge, made our implementations publicly available, and established the diabetic retinopathy database DiaRetDB1 to promote and help other researchers collect and publish their data. We believe that public databases and common evaluation procedures support development of better methods and promote the best methods to be adopted in clinical practice.

Acknowledgments

The authors thank the Finnish Funding Agency for Technology and Innovation (TEKES Project nos. 40430/05 and 40039/07) and the partners of the ImageRet project (http://www2.it.lut.fi/project/imageret/) for their support.

References

[1] T. Kauppi, *Eye fundus image analysis for automatic detection of diabetic retinopathy [Ph.D. thesis]*, Lappeenranta University of Technology, 2010.

[2] World Health Organization, "Definition, diagnosis and classification of diabetes mellitus and its complications: part 1: diagnosis and classification of diabetes mellitus," Tech. Rep., World Health Organization Noncommunicable, Geneva, Switzerland, 1999.

[3] World Health Organization and The International Diabetes Federation, *Diabetes Action Now: An Initiative of the World Health Organization and the International Diabetes Federation*, 2004.

[4] G. von Wendt, *Screening for diabetic retinopathy: aspects of photographic methods [Ph.D. thesis]*, Karolinska Institutet, 2005.

[5] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen et al., "The diaretdb1 diabetic retinopathy database and evaluation protocol," in *Proceedings of the British Machine Vision Conference (BMVC '07)*, pp. 252–261, University of Warwick, 2007.

[6] T. Kauppi, V. Kalesnykiene, J. K. Kamarainen et al., "Diaretdb1 diabetic retinopathy database and evaluation protocol," in *Proceedings of the Medical Image Understanding and Analysis (MIUA '07)*, pp. 61–65, 2007.

[7] T. Kauppi, J.-K. Kamarainen, L. Lensu et al., "A framework for constructing benchmark databases and protocols for retinopathy in medical image analysis," in *Intelligent Science and Intelligent Data Engineering*, J. Yang, F. Fang, and C. Sun, Eds., vol. 7751 of *Lecture Notes in Computer Science*, pp. 832–843, Springer, Berlin, Germany, 2012.

[8] J.-K. Kamarainen, L. L. Lensu, and T. Kauppi, "Combining multiple image segmentations by maximizing expert agreement," in *Machine Learning in Medical Imaging*, F. Wang, D. Shen, P. Yan, and K. Suzuki, Eds., Lecture Notes in Computer Science, pp. 193–200, Springer, Berlin, Germany, 2012.

[9] T. Kauppi, J.-K. Kamarainen, L. Lensu et al., "Fusion of multiple expert annotations and overall score selection for medical

image diagnosis," in *Proceedings of the 16th Scandinavian Conference on Image Analysis (SCIA '09)*, pp. 760–769, Springer, 2009.

[10] N. A. Thacker, A. F. Clark, J. L. Barron et al., "Performance characterization in computer vision: a guide to best practices," *Computer Vision and Image Understanding*, vol. 109, no. 3, pp. 305–334, 2008.

[11] K. H. Zou, "Receiver operating characteristic (roc) literature research," 2002, http://www.spl.harvard.edu/archive/spl-pre2007/pages/ppl/zou/roc.html.

[12] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[13] "Diabetic retinopathy database and evaluation protocol (DIARETDB1)," Electronic material (Online), http://www2.it.lut.fi/project/imageret/diaretdb1_v2_1/.

[14] "Image annotation tool (IMGANNOTOOL)," Electronic material (Online), http://www2.it.lut.fi/project/imageret/.

[15] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, 2002.

[16] P. Paalanen, J.-K. Kamarainen, J. Ilonen, and H. Kälviäinen, "Feature representation and discrimination based on Gaussian mixture model probability densities—practices and algorithms," *Pattern Recognition*, vol. 39, no. 7, pp. 1346–1358, 2006.

[17] "Structured analysis of the retina (STARE)," Electronic material (Online), http://www.clemson.edu/ces/.

[18] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piece-wise threhsold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[19] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 951–958, 2003.

[20] "Digital retinal images for vessel extraction (DRIVE)," Electronic material (Online), http://www.isi.uu.nl/Research/Databases/DRIVE/.

[21] J. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[22] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation a new publicly available database," in *Medical Imaging: Image Processing*, pp. 648–656, 2004.

[23] "Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology (MESSIDOR)," Electronic material (Online), http://messidor.crihan.fr/.

[24] "Collection of multispectral images of the fundus (CMIF)," Electronic material (Online), http://www.cs.bham.ac.uk/research/projects/fundus-multispectral/.

[25] I. B. Styles, A. Calcagni, E. Claridge, F. Orihuela-Espina, and J. M. Gibson, "Quantitative analysis of multi-spectral fundus images," *Medical Image Analysis*, vol. 10, no. 4, pp. 578–597, 2006.

[26] "Retinopathy online challenge (ROC)," Electronic material (Online), http://roc.healthcare.uiowa.edu/.

[27] M. Niemeijer, B. van Ginneken, M. J. Cree et al., "Retinopathy online challenge: automatic of microaneurysms in digital photographs," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 185–195, 2010.

[28] 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.

[29] "Review: retinal vessel image set for estimation of widths (REVIEW)," Electronic material (Online), http://reviewdb.lincoln.ac.uk/.

[30] B. Al-Diri, A. Hunter, D. Steel, M. Habib, T. Hudaib, and S. Berry, "Review—a reference data set for retinal vessel profiles," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2262–2265, Vancouver, BC, Canada, August 2008.

[31] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: a database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1–3, pp. 157–173, 2008.

[32] "Application protocol for electronic data exchange in healthcare environments version," 2.5.1, ANSI Standard, http://www.hl7.org/.

[33] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classfiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[34] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 903–921, 2004.

[35] T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado, "The use of receiver operating characteristic curves in biomedical informatics," *Journal of Biomedical Informatics*, vol. 38, no. 5, pp. 404–415, 2005.

[36] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–11104, 2000.

[37] E. Bailliere, S. Bengio, F. Bimbot et al., "The BANCA database and evaluation protocol," in *Proceedings of the International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA '03)*, pp. 625–638, 2003.

[38] M. Everingham and A. Zisserman, "The pascal visual object classes challenge VOC2006 results," in *Proceedings of the ECCV Workshop of VOC*, 2006.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 250929, 15 pages
http://dx.doi.org/10.1155/2013/250929

Research Article

Comparative Evaluation of Osseointegrated Dental Implants Based on Platform-Switching Concept: Influence of Diameter, Length, Thread Shape, and In-Bone Positioning Depth on Stress-Based Performance

Giuseppe Vairo[1] and Gianpaolo Sannino[2]

[1] *Department of Civil Engineering and Computer Science, University of Rome "Tor Vergata," Via del Politecnico 1, 00133 Rome, Italy*
[2] *Department of Oral Health, University of Rome "Tor Vergata," Viale Oxford, 00133 Rome, Italy*

Correspondence should be addressed to Gianpaolo Sannino; gianpaolo.sannino@uniroma2.it

Received 31 March 2013; Accepted 19 May 2013

Academic Editor: Carlo Cattani

This study aimed to investigate the influence of implant design (in terms of diameter, length, and thread shape), in-bone positioning depth, and bone posthealing crestal morphology on load transfer mechanisms of osseointegrated dental implants based on platform-switching concept. In order to perform an effective multiparametric comparative analysis, 11 implants different in dimensions and in thread features were analyzed by a linearly elastic 3-dimensional finite element approach, under a static load. Implant models were integrated with the detailed model of a maxillary premolar bone segment. Different implant in-bone positioning levels were modeled, considering also different posthealing crestal bone morphologies. Bone overloading risk was quantified by introducing proper local stress measures, highlighting that implant diameter is a more effective design parameter than the implant length, as well as that thread shape and thread details can significantly affect stresses at peri-implant bone, especially for short implants. Numerical simulations revealed that the optimal in-bone positioning depth results from the balance of 2 counteracting effects: cratering phenomena and bone apposition induced by platform-switching configuration. Proposed results contribute to identify the mutual influence of a number of factors affecting the bone-implant loading transfer mechanisms, furnishing useful insights and indications for choosing and/or designing threaded osseointegrated implants.

1. Introduction

In the last three decades and in the field of the prosthetic dentistry, features of dental implants and surgical procedures have been developed and enhanced aiming to ensure predictable results and to improve function and aesthetics in completely or partially edentulous patients [1].

A dental implant is a biocompatible device, surgically placed into mandibular or maxillary bone for supporting a prosthetic tooth crown, and thus allowing the replace of the teeth lost due to caries, periodontal disease, injuries, or other reasons. Worldwide statistics show that a high success rate of dental implants (over 95%) occurs if implants are properly designed and manufactured, and if they are inserted in a bone segment characterized by good quality and quantity (e.g., [2–4]). Nevertheless, success of the prosthetic treatment is widely affected by a number of factors that can change the biomechanichal coupling between implant and bone, such as implant location, mechanical and morphological properties of bone, mechanical and geometrical features of implant, and type and magnitude of the load transferred by the implant to the bone, as well as by host factors such as smoking and bacterial environment [5–7].

A crucial aspect that determines the effectiveness of a dental implantation is identified by the proper development of the osseointegration process at the bone-implant interface. This process is similar to the healing process in bone fracture [7–9] and arises from remodeling mechanisms that involve a number of cellular and extracellular coupled biomechanical features. After the implantation, the gap between the implant

and the host bone is rapidly filled by blood clots that are afterwards substituted by a trabecular network. The latter generally evolves towards the formation of lamellar bone that, in turn, undergoes a maturation process that modifies density and mechanical properties of the tissue [8–11]. At the end of the healing process, the mature bone is directly in contact with the implant surface, leading to an interfacial binding that allows to enhance loading transfer mechanisms from prosthetic crown to the bone [12, 13].

Nevertheless, a proper osseointegration process may be counteracted by the activation of histological resorption mechanisms [9, 14–16] that can induce bone weakening or loss at the peri-implant region. Bone resorption mainly affects the bone region around the implant neck, producing a cratering morphology, and it may be activated by surgical trauma or bacterial infection, as well as by overloading states [4, 5, 14–22]. Under functional or pathological (e.g., induced by bruxism) loads, overloading at the peri-implant bone may occur by a shortcoming in load transfer mechanisms, mainly due to bad occlusion, improper implant use, wrong prosthesis and/or implant design, and improper implant placement. In these cases, high stress concentrations are induced at the bone-implant interfaces, leading to possible physiologically inadmissible strains that activate bone resorption [23, 24]. Clinical trials and follow-up analyses [2–4, 17, 18] have shown that the implant failure may generally occur if the bone resorption process significantly evolves from a crestal initiation. Depending on implant features, positioning and loads, this process may become instable, leading to a progressive increase in stress intensity at the peri-implant interface [19] that, in turn, further contributes to the progressive overload-induced bone loss.

Recent clinical evidence [25–29] suggests that cratering phenomena may be significantly limited when the connection diameter of the abutment is narrower than the implant collar and when an implant subcrestal positioning is applied. In this case, probably due to the different position of the implant/abutment microgap and to the different stress pattern induced at the peri-implant regions with respect to a crestal positioning, remodeling process generally evolves allowing bone apposition on the horizontal implant surface and thus transferring the biological width from the vertical to the horizontal level (platform switching) [30–34].

In order to improve durability and clinical effectiveness of rehabilitations based on such an approach, mechanical and biological factors mainly affecting loading transfer from implant to bone have to be properly identified and quantified. Thereby, optimized implant designing strategies and surgical protocols could be traced, allowing us to minimize overloading risks and marginal bone loss, as well as contributing to ensure predictable clinical results.

In the recent specialized literature many authors have proposed results based on well-established in vivo, in vitro, and in silico approaches, aiming to investigate main biomechanical factors influencing the preservation of the peri-implant marginal bone as well as the stress/strain patterns induced by osseointegrated implants [4, 26–29, 35, 36]. In this context, finite-element method has been widely used in the last years to analyze the influence of implant and prosthesis design [37–40], of magnitude and direction of loads [41–44], and of bone mechanical properties [45–47], as well as for modeling different clinical scenarios [48–54]. Nevertheless, many effects related to the implant design and to the in-bone positioning depth, as well as their mutual influence on the stress-based implant performance, have not yet been completely understood and clarified, especially for implants based on platform-switching concept.

In this study, 11 threaded dental implants, based on platform-switching concept and different for dimensions and thread type, were compared via a multiparametric three-dimensional (3D) finite-element approach. Accurate and convergent bone-implant models, defined by considering a maxillary premolar bone segment, have been solved by employing a linearly elastic displacement-based formulation and considering a static functional loading condition. Stress distributions were numerically evaluated at the peri-implant regions on both compact and cancellous bone, furnishing quantitative risk measures of bone physiological failure. Proposed numerical results highlighted the influence of implant shape, in terms of implant length and diameter as well as in terms of thread features, on possible overloading risks and on mechanisms of load transfer. The influence of implant positioning in bone was also investigated by considering numerical models based on both crestal and subcrestal implant placements. Finally, in the case of a crestal positioning and in order to contribute to the understanding of the biomechanical relationship between mechanical stimuli and marginal bone loss, several numerical simulations were carried out for analyzing the effects of different cratering levels on stress patterns at the peri-implant bone.

2. Material and Methods

Ten threaded dental implants, different in diameter (D), length (L), thread shape, and geometrical concept, were analyzed and compared with each other and with an Ankylos implant (Dentsply Friadent, Mannheim, Germany) characterized by $D = 3.5$ mm and $L = 11.0$ mm. Figure 1 summarizes the main geometrical features of the implants analyzed in this study, introducing also the corresponding notation. Symbols T0/30 and T10/30 refer to the implant thread: T0/30 denotes a saw-tooth thread with the side angled at 120° with respect to the implant axis and with a free thickness of 0.33 mm at the internal diameter; T10/30 denotes a trapezoid-shaped thread with sides angled at 120° and 100° with respect to the implant axis and with a free thickness of 0.25 mm at the internal diameter. Both threads are characterized by two starts with a conical helix having the same anomaly and with an effective pitch of 1.2 mm. Moreover, symbol ST indicates that both starts exhibit the same thread truncation, resulting in a maximum thread depth of 0.38 mm, whereas symbol DT denotes implants with a different thread truncation for each start, resulting in maximum thread depths of 0.19 mm and 0.38 mm, respectively. Implants, except the Ankylos device, have also a helical milling, with the effective pitch equal to the implant threaded length. Depending on width and depth of cut, small and large millings are identified by symbols SM and

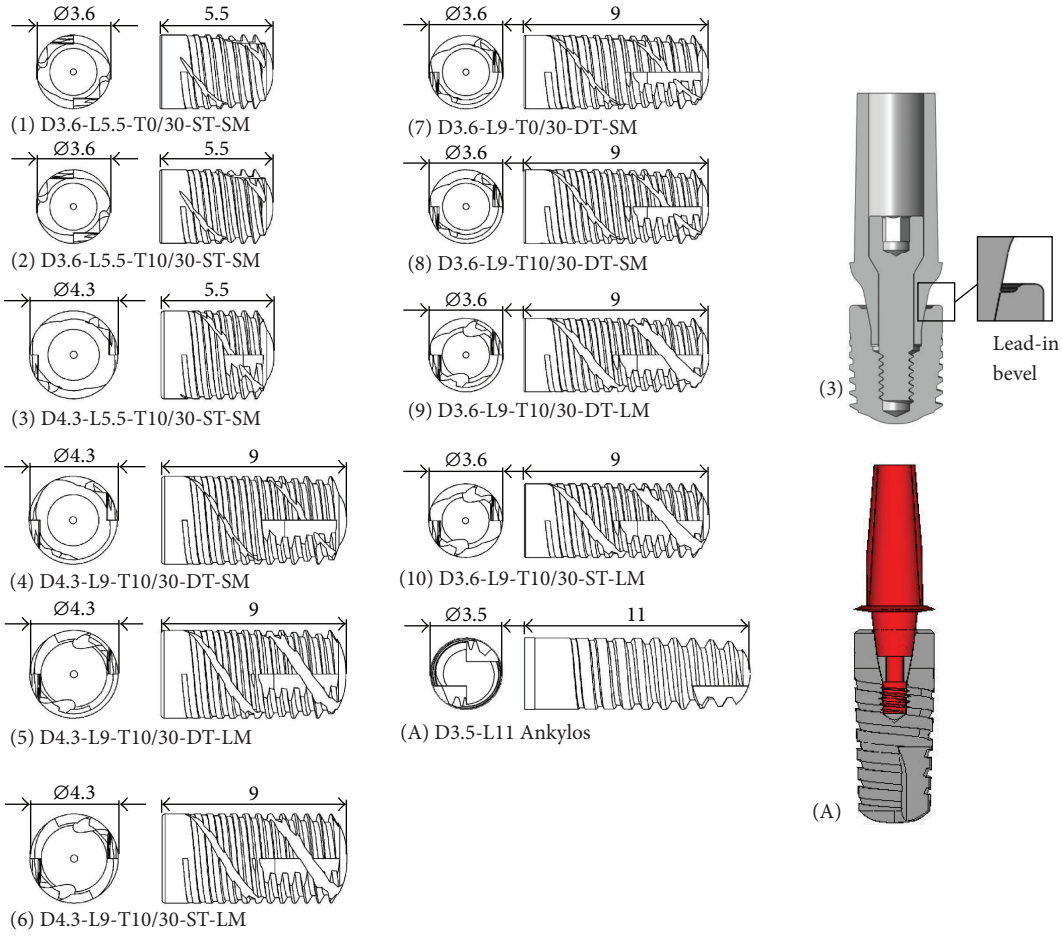


Figure 1: Threaded dental implants analyzed in this study. Notation and examples of implant-abutment coupled systems that allow a platform-switching configuration.

LM, respectively. Implants denoted by 1 to 10 in Figure 1 were characterized by an internal lead-in bevel extending from the outer most diameter of the implant platform into a flattened area or ledge. Moreover, implants analyzed in this study have vertical cutting grooves for self-tapping insertion and have been coupled with abutments characterized by connection diameters narrower than the implant collars, thereby allowing a platform-switching configuration (see Figure 1).

Models of implants and abutments were built up by using a parametric CAD software (SolidWorks 9; Dessault Systèmes, Concord, Mass) and, in order to perform consistent comparisons, they were integrated within the model of a premolar bone segment, obtained by the three-dimensional (3D) model of an edentulous maxilla (Figure 2). The latter was reconstructed starting from multislice computed tomography (MSCT) scans and by using a modeling commercial software (Mimics, Materialise HQ, Leuven, Belgium). Moving from the different hues of gray displayed in the planar CT scans, corresponding to different radiolucency levels of substances with different density values, the software allowed us to distinguish between mineralized and soft tissues, by filtering pixels with a suitable Hounsfield units (HU) [55]. In detail, disregarding gingival soft tissues, the solid model of the maxillary jaw was obtained by a segmentation procedure of voxels identified by HU $>$ 150 (Figure 2(a)) and based on a home-made smoothed linear interpolation algorithm. Cortical and trabecular regions were distinguished, considering $150 < \text{HU} \leq 750$ for the cancellous bone and HU $>$ 750 for the cortical bone. With the aim of improving the model quality, ad hoc local geometry adjustments were performed, ensuring that the cortical bone regions were characterized by a mean thickness of about 2 mm. Starting from the complete maxillary jaw model, the finite-element computations were carried out on a submodel of the second premolar region, defined by considering two coronal sections at the distance of 40 mm along the mesiodistal direction (y, in Figure 2(b)) and positioning implants at the mid-span of the bone segment.

A subcrestal positioning was firstly investigated, by considering implant models positioned with the crestal platform at 1 mm depth with respect to the outer bone surface. As a notation rule, in the foregoing this configuration will be denoted as P1. Moreover, in order to analyze the positioning influence for implants similar in diameter and length, numerical models relevant to the implants D3.6-L9-T10/30-DT-SM and Ankylos (indicated as 8 and A, resp., in Figure 1) were analyzed by considering a crestal positioning (i.e., with the

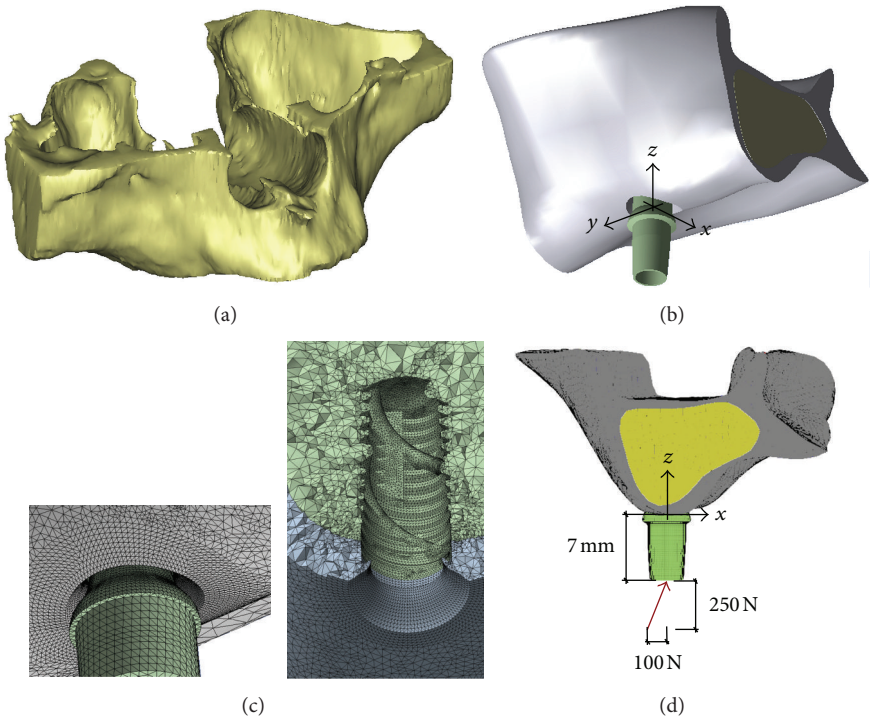


Figure 2: (a) Three-dimensional solid model of the edentulous maxilla considered in this study and obtained by a segmentation process based on multislice computed tomography (MSCT). (b) Submodel of the second premolar maxillary region, defined by considering two coronal sections at the distance of 40 mm along the mesiodistal direction (y axis) and positioning implants at the mid-span of the bone segment. (c) Examples of mesh details. (d) Loading condition.

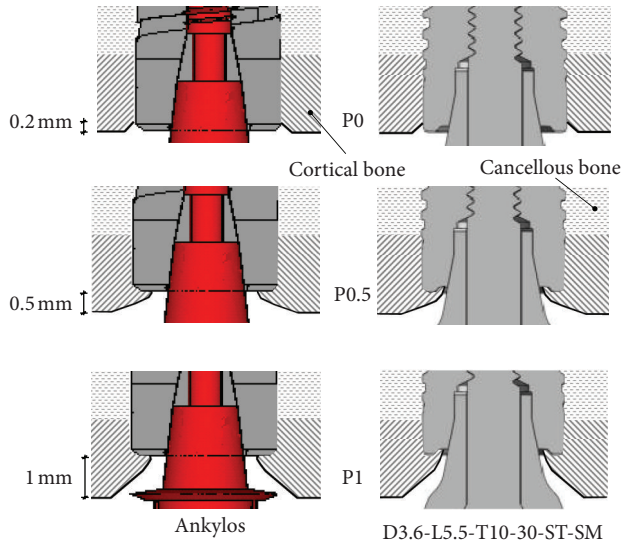


Figure 3: Modeling of crestal bone geometries and different configurations of implant in-bone positioning analyzed in this study. In the case of the configuration P0, a crestal bone loss of about 10% in thickness is depicted.

implant platform at the level of the outer bone surface and denoted as P0); an intermediate subcrestal positioning at 0.5 mm depth (denoted as P05). With the aim of reproducing as realistically as possible the physiological structure of the compact bone arising around a functioning implant after a healing period, different crestal geometries were modeled. In particular, in agreement with well-established clinical evidence [25–27] and modeling approaches [40, 47, 53], and as sketched in Figure 3, a crestal bone apposition at the implant platform of about 0.25 mm in mean thickness was modeled for subcrestal placements (i.e., for models denoted as P1 and P05), whereas a marginal bone loss of 10% in cortical

thickness was modeled for the crestal positioning (P0). For implants 8 and A crestally placed (P0), the influence of different levels of marginal bone loss (0–50% in cortical thickness) was also analyzed.

All the involved materials were modeled as linearly elastic with an isotropic constitutive symmetry, and all material volumes were modeled as homogeneous. Thereby, bone living tissue was described by considering a dry-material model, wherein viscous and fluid-solid interaction effects were neglected. Implants and abutments were assumed to be constituted by a titanium alloy, Ti6Al4V, whose Young's modulus and Poisson's ratio were 114.0 GPa and 0.34, respectively [56]. Bone elastic properties were assumed to approximate type II bone quality [57] and, in agreement with data available in the literature [40, 47, 58], they were set as follows:

(i) Poisson's ratio of the bone tissue (both cortical and trabecular) equal to 0.30;

(ii) Young's modulus of the cortical bone equal to 13.7 GPa;

(iii) Young's modulus of the cancellous bone equal to 0.5 GPa, corresponding to a mean bone density of about 0.5 g·cm^{-3} [59].

Finite-element simulations were carried out considering a static load applied at the top of the abutments without any eccentricity with respect to the implant axis and angled with respect to the occlusal plane of about 68°. The lateral force component along the buccolingual direction (x, in Figure 2) was assumed to be equal to 100 N and the vertical intrusive one (along z, in Figure 2) was 250 N. In order to allow consistent comparisons, abutments were adjusted in such a way that the application points of the load were 7 mm from the bone insertion surface in all numerical models (see Figure 2(d)).

Complete osseous integration between implant and bone tissue was assumed, enforcing the continuity of the displacement field at the bone-implant interface. Furthermore, displacement continuity is imposed between each component of a given prosthetic device. As regards boundary conditions for numerical models describing the coupled bone-implant system, all displacement degrees of freedom were prevented for any boundary node lying on the coronal sections delimiting the bone submodel. In agreement with the theory of elasticity [60], since the distance between submodel boundary sections and the implant location was much greater than the implant's characteristic dimensions, these boundary conditions did not significantly affect stress-based comparative results at the peri-implant regions.

Discrete finite-element meshes were generated by employing elements based on a pure displacement formulation and were analyzed with a commercial solver code (Ansys 13.0; Ansys Inc., Canonsburg, PA). Computational models were obtained by considering 10-node tetrahedral elements [61], with quadratic shape functions and three degrees of freedom per node. In order to ensure suitable accuracy of the numerical finite-element solutions at the peri-implant regions, mesh-size for the bone-implant models was set up as a result of a convergence analysis, based on the coupled estimate within the multiregion computational domain of the displacement error norm and of the energy error norm [61]. In detail, following the numerical procedure proposed by Zienkiewicz and Zhu [62], implemented in the Ansys environment and recently applied for prosthetic dental applications [47], the proposed numerical results were obtained by solving discrete models based on $h_0/D = 0.1$ and $h_i/D = 0.01$, h_0 and h_i being mean mesh-size away from the bone-implant interface and close to the peri-implant regions, respectively. This choice was proved to ensure a good numerical accuracy, resulting for all models analyzed in this study in a value of the energy error norm lower than 5% and in a value of the displacement error norm lower than 0.5%.

Jaw submodel treated by a single-implant prosthesis was numerically compared by analyzing stress distributions arising at the peri-implant regions. The Von Mises equivalent stress (σ_{VM}), often used in well-established numerical dental studies (e.g., [35–54, 63, 64]), was used as a global stress indicator for characterizing load transfer mechanisms of a given implant. Nevertheless, the Von Mises stress measure, always positive in sign, does not allow a distinction between tensile and compressive local stresses. Since experimental evidence [24, 58, 65] confirms that bone physiological failure and overload-induced resorption process are differently activated in traction and compression, more effective and direct risk indications were obtained by analyzing stress measures based on principal stresses (σ_i, with $i = 1, 2, 3$) [44, 47, 53, 63, 64]. In detail, in a given material point P of the computational domain that models the peri-implant bone, the following stress measures were computed:

$$\begin{aligned} \sigma_C(P) &= \min\left\{\sigma_1(P), \sigma_2(P), \sigma_3(P), 0\right\}, \\ \sigma_T(P) &= \max\left\{\sigma_1(P), \sigma_2(P), \sigma_3(P), 0\right\}, \end{aligned} \tag{1}$$

σ_C and σ_T having the meaning of maximum compressive and maximum tensile stress in P, respectively. Therefore, in order to combine effects induced on bone by compressive and tensile local states which are simultaneously present, the bone safety in P against overloading-related failure/resorption process activation was postulated to occur if the following inequality was satisfied:

$$R = \frac{|\sigma_C|}{\sigma_{C0}} + \frac{\sigma_T}{\sigma_{T0}} \leq 1, \tag{2}$$

where symbol $|a|$ denotes the absolute value of the scalar quantity a and where σ_{T0}, σ_{C0} are the admissible stress levels in pure traction and compression, respectively. Accordingly, the dimensionless positive quantity R can be thought of as a quantitative risk indicator, such that the condition $R > 1$ identifies a local critical state of bone with respect to overloading effects. By assuming that overloads occur when ultimate bone strength is reached, in this study it was assumed that $\sigma_{T0} = 180$ MPa and $\sigma_{C0} = 115$ MPa for cortical bone and $\sigma_{T0} = \sigma_{C0} = 5$ MPa for trabecular bone [58, 65].

In order to perform significant numerical comparisons, the previously introduced stress measures and the risk index R were computed for each implant within a control volume Ω, defined by considering a bone layer surrounding the implant

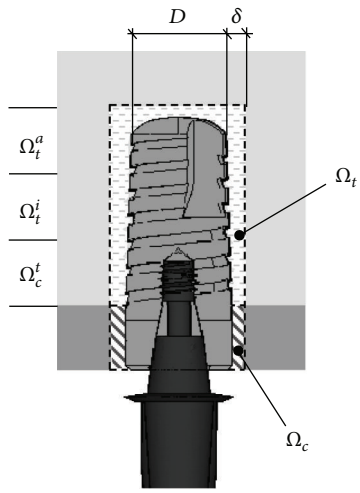


Figure 4: Control regions employed for computing the local stress measures and the overloading risk index R at the bone-implant interface.

with a mean thickness δ. With reference to the sketch in Figure 4, the region Ω has been conveniently considered as subdivided in its complementary parts Ω_c and Ω_t (such that $\Omega = \Omega_c \cup \Omega_t$), representing cortical and trabecular control regions, respectively. In turn, Ω_t has been further subdivided, by 2 planes orthogonal to the implant axis, into 3 complementary control subregions having equal length along the implant axis. These three trabecular regions will be denoted as Ω_t^c (crestal region), Ω_t^i (intermediate region), and Ω_t^a (apex region). Results discussed in the foregoing were obtained by assuming $\delta/D = 0.25$, and they refer to average and peak values of σ_{VM}, σ_C, σ_T, and R over Ω_c, Ω_t^c, Ω_t^i, Ω_t^a. These results were computed via a postprocessing phase carried out by means of a MatLab (The MathWorks, Inc., Natick, MA) home-made procedure, taking as input by the solver code some primary geometrical and topological data (nodes and elements lying in Ω), as well as stress solutions at the finite-element Gauss points within Ω.

3. Results

3.1. Subcrestal Positioning P1. For implants introduced in Figure 1 and considering the subcrestal positioning P1 (see Figure 3), Figures 5 and 6 show Von Mises stress distributions relevant to the loading coronal plane $y = 0$, computed via the present 3D finite-element approach at the peri-implant cortical and trabecular bone regions. Moreover, Figure 7 shows average and peak values over the control volumes Ω_c and Ω_t (see Figure 4) of σ_{VM} and of the principal stress measures defined by (1). Finally, Figure 8 highlights mean and peak values of the overloading risk index R computed at both trabecular and cortical peri-implant bone regions.

By assuming complete osseous integration, the highest stress concentrations were computed at the cortical bone near the implant neck. There, stress patterns were significantly affected by implant diameter (D) and bone-implant interface length (L). In detail, by increasing D and/or by increasing L mean and peak stress values decreased in Ω_c and Ω_t, and stress distributions tended to be more homogenous. Compressive mean and peak values at the cortical peri-implant region always prevailed with respect to the corresponding tensile states. This occurrence was not generally respected at the trabecular interface, wherein tensile stresses were higher at the crestal region (Ω_t^c) and smaller at the implant apex (Ω_t^a) than the compressive stresses. Nevertheless, the highest trabecular stress peaks were associated with the compressive states arising in Ω_t^a (see Figure 7(b)).

Referring to the notation introduced in Figure 1, implants denoted by D4.3-L9 (i.e., labeled as 4, 5, and 6) exhibited the best stress performances, resulting in the smallest values of the stress measures as well as in the smallest values of the overloading risk index R. On the contrary, implants denoted by D3.6-L5.5 (labeled as 1 and 2) numerically experienced the worst loading transmission mechanisms. Moreover, the stress-based performance of the commercial implant Ankylos D3.5-L11 was estimated as fully comparable with that of the threaded implants D3.6-L9 (labeled as 7, 8, 9, and 10), although the greater Ankylos' length induced more favorable stress distributions at the trabecular bone, especially referring to the compressive states arising at the implant apex (see Figure 7(b)).

Proposed results clearly show that the parameter that mainly affects the implant stress-based performances is the diameter D, irrespective of the length L. In fact, by comparing stress results relevant to implant 2 with those of implant 3, that is, by increasing D of about 20% (passing from $D = 3.6$ mm to $D = 4.3$ mm) when $L = 5.5$ mm, compressive (resp., tensile) peak values reduced of about 27% in both Ω_c and Ω_t (resp., 20% in Ω_c and 30% in Ω_t). On the contrary, by comparing stress results relevant to implant 2 with those of implant 9, that is, by increasing L of about 60% (passing from $L = 5.5$ mm to $L = 9$ mm) when $D = 3.6$ mm, compressive peaks reduced only by about 16% (resp., 26%) at the cortical (resp., trabecular) bone, whereas tensile peaks were almost comparable. These considerations are qualitatively applicable also when the overloading risk index R is addressed (see Figure 8), leading to similar conclusions.

Within the limitations of this study, overloading risks were greater in cancellous region than those in cortical, and proposed numerical results highlighted that, under the simulated loading condition, the safety inequality $R < 1$ was everywhere satisfied in bone for all the analyzed implants.

Moreover, the proposed numerical results suggest that thread shape and thread details can induce significant effects on local stress patterns in bone around implants. In particular, the use of the same thread truncation (ST) for both thread starts induced a more uniform local stress distributions than the case characterized by a different thread truncation (DT), since all the threads had practically the same engaged depth. As a result, mean and peak values of σ_T reduced at the cortical bone passing from DT to ST, as it is shown in Figure 7(b) by comparing results relevant to implants 5 and 6 (peaks reduced of about 20% and mean values of about 13%) and to implants 9 and 10 (peaks reduced of about 23% and mean values of about 18%).

The influence of the thread shape may be clearly highlighted by analyzing the stress-based performances of

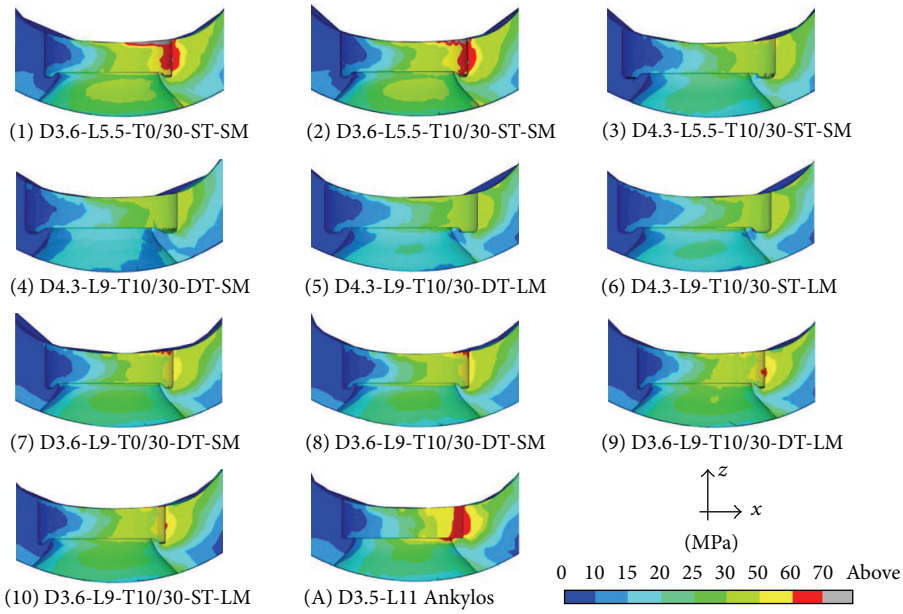


FIGURE 5: Von Mises stress contours (blue: 0; red: 70 MPa) at the coronal section $y = 0$ for implants defined in Figure 1 and in the case of the subcrestal positioning P1 (see Figure 3). Cortical peri-implant bone interface.

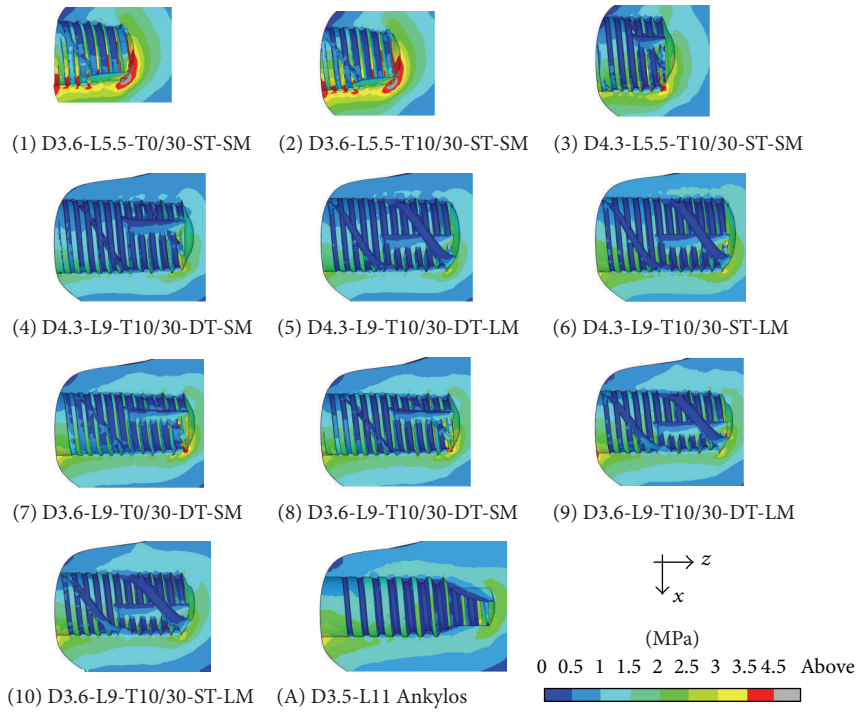


FIGURE 6: Von Mises stress contours (blue: 0; red: 4.5 MPa) at the coronal section $y = 0$ for implants defined in Figure 1 and in the case of the subcrestal positioning P1 (see Figure 3). Trabecular peri-implant bone interface.

implants 1 and 2 and of implants 7 and 8. In particular, trapezoid-shaped thread (labelled as T10/30 in Figure 1) induced more favorable compressive and tensile states at both cortical and trabecular regions than the saw-tooth thread (T0/30), leading to the reduction of the cortical peak values of about 24% for σ_C when the implants D3.6-L5.5 were addressed and of about 35% for σ_T in the case of the implants D3.6-L9. Such an effect is also observable by analyzing the risk index R (see Figure 8). In particular, the thread shape T10/30 induced a significant reduction in R (at both cortical and trabecular regions), especially for short implants.

Finally, indications on the influence of the helical-milling width and depth may be drawn by considering numerical results relevant to implants 4 and 5 and to implants 8 and 9.

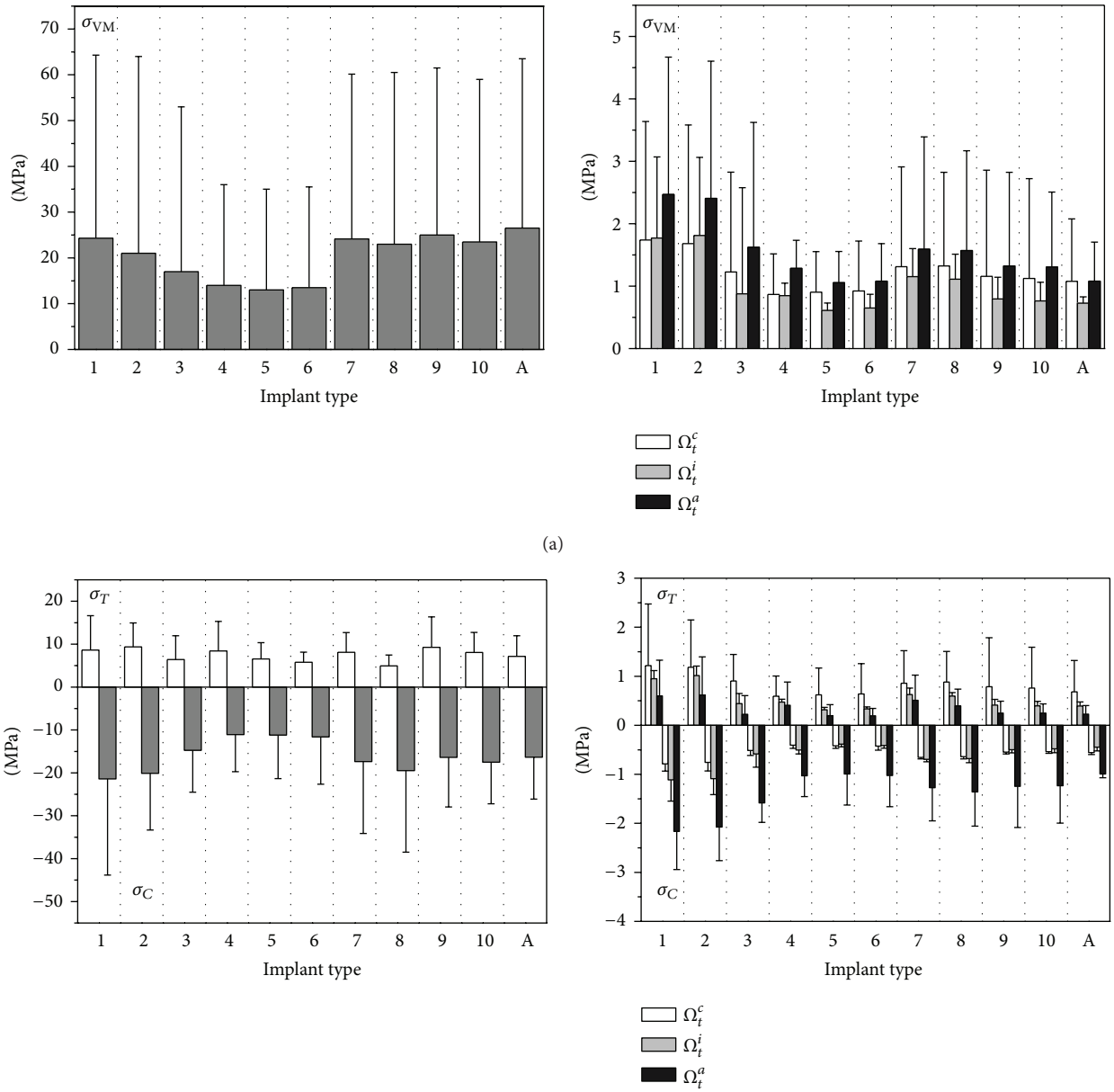


FIGURE 7: Von Mises ((a), σ_{VM}) and principal ((b), σ_T tensile and σ_C compressive) stress measures at cortical (left side) and trabecular (right side) bone-implant interface for implants defined in Figure 1 and in the case of the subcrestal positioning P1 (see Figure 3). Average (bars) and peak (lines) values.

Although almost comparable global stress patterns and local stress measures were experienced passing from SM (small milling) to LM (large milling), the analysis of the index R reveals that large milling shape can induce a reduction of the risk of overloading states at the cancellous bone, especially for small values of L.

3.2. Influence of In-Bone Positioning Depth. In order to analyze the influence of the implant in-bone positioning depth on loading transmission mechanisms, reference has been made to the comparative numerical analyses carried out for the implant D3.6-L9-T10/30-DT-SM and for the implant Ankylos D3.5-L11 (i.e., for implants 8 and A in Figure 1). Addressing the positioning configurations introduced in Figure 3, Figure 9 shows Von Mises stress distributions relevant to the loading coronal plane $y = 0$, computed at cortical and trabecular peri-implant bone regions, and Figure 10 shows mean and peak values of σ_{VM}, σ_T, and σ_C computed over the control volumes Ω_c and Ω_t (see Figure 4). Finally, Figure 11 summarizes mean and peak values of the overloading risk index R computed at both trabecular and cortical bone interfaces. It is worth pointing out that the

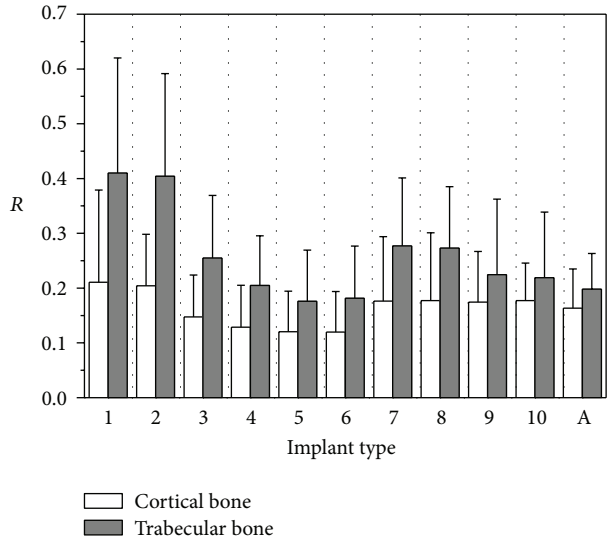


Figure 8: Overloading risk index R computed at cortical and trabecular peri-implant bone for implants defined in Figure 1 and in the case of the subcrestal positioning P1 (see Figure 3). Average (bars) and peak (lines) values.

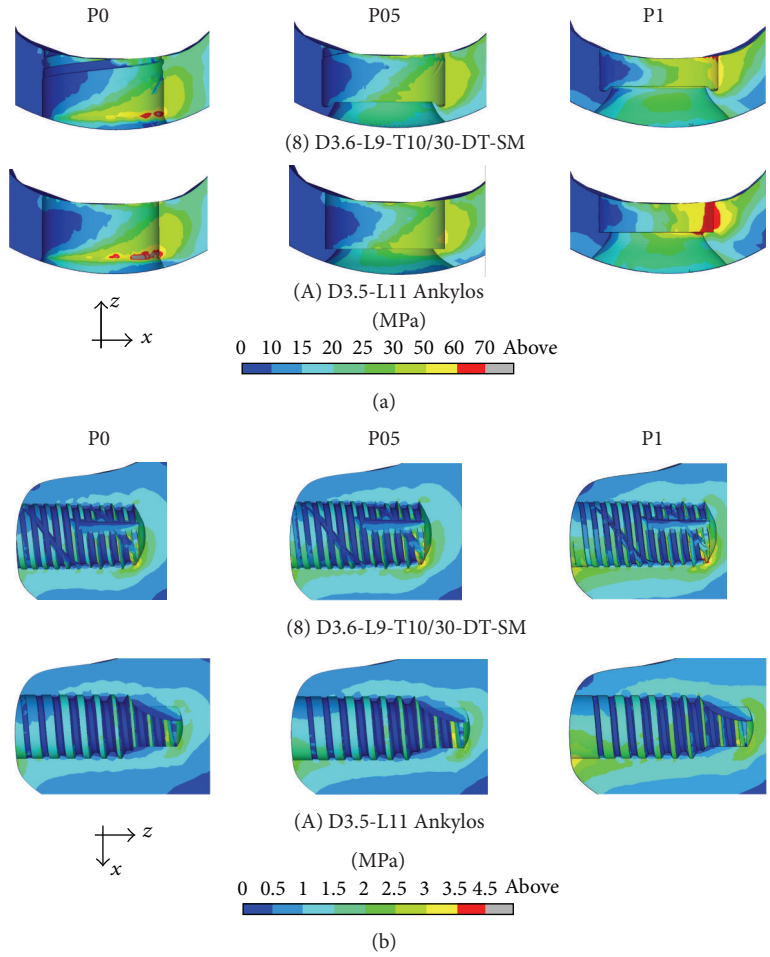


Figure 9: Von Mises stress contours (blue: 0; red: 70 MPa) at the coronal section $y = 0$ for implants 8 and A (see Figure 1) and for different implant in-bone positioning levels (see Figure 3). Cortical (a) and trabecular (b) peri-implant bone interface.

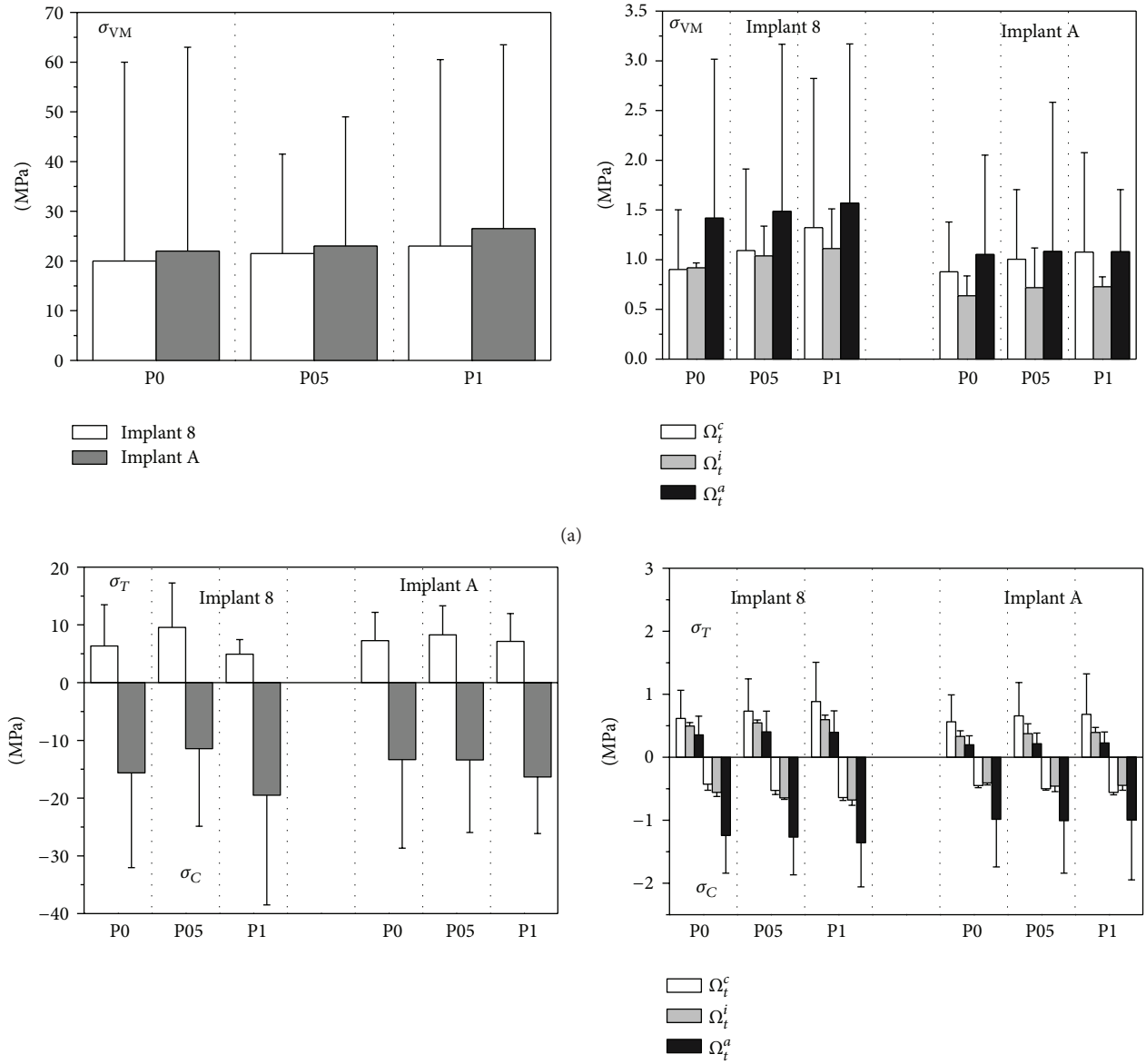


Figure 10: Von Mises ((a), σ_{VM}) and principal ((b), σ_T tensile and σ_C compressive) stress measures at cortical (left side) and trabecular (right side) bone-implant interface for implants 8 and A (see Figure 1) and for different implant in-bone positioning levels (see Figure 3). Average (bars) and peak (lines) values.

results referred to the crestal positioning P0 were computed by modeling a crestal bone loss of about 10% in cortical thickness (see Figure 3).

Proposed numerical results confirmed that the implant Ankylos induced more favorable loading transmission mechanisms than implant 8, also considering different values of in-bone positioning depth. Moreover, the analysis of Von Mises stress distributions as well as of the values of principal-stress-based measures suggests that the crestal positioning (P0) induced significant stress concentrations at the cortical bone around the implant neck. In this case, stress peaks were estimated as comparable with those obtained for the subcrestal positioning P1. When the intermediate subcrestal positioning P05 was analyzed, the lowest compressive peaks at Ω_c were experienced for both implants, although tractions slightly greater than the other positioning configurations occurred. In trabecular bone, stress patterns were computed as almost comparable in the three cases under investigation. Nevertheless, the positioning case P0 induced stress distributions in trabecular regions that were slightly better than P05 and P1.

This evidence is fully confirmed by analyzing the results obtained for the risk index R. In particular, referring to its peak values, overloading risk at the cortical bone for P05 was lower than that for P0 and P1 of about 14% and 19% for implant 8, respectively, and of about 6% and 3% for implant A.

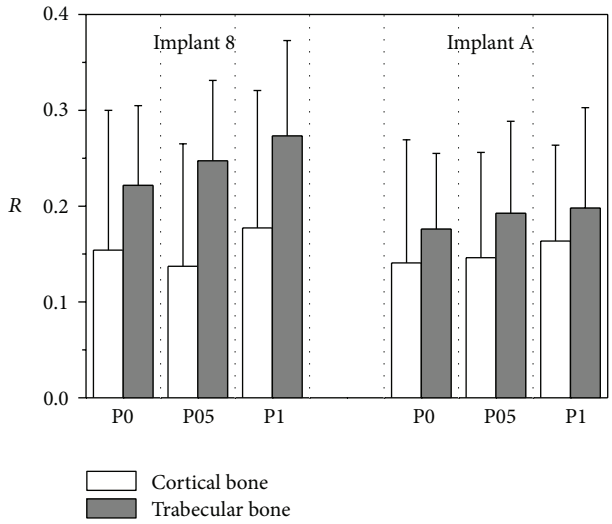


Figure 11: Overloading risk index R computed at cortical and trabecular peri-implant bone for implants 8 and A (see Figure 1) and for different implant in-bone positioning levels (see Figure 3). Average (bars) and peak (lines) values.

On the other hand, values of R for P0 were lower at the trabecular bone than those for P05 and P1 of about 10% and 18% for implant 8, respectively, and of about 10% and 15% for implant A.

3.3. Influence of Marginal Bone Loss in Crestal Positioning. For implants 8 and A (see Figure 1), crestally positioned in agreement with the configuration P0 (see Figure 3), the influence of the amount in crestal bone loss was also analyzed. In particular, numerical simulations were carried out considering three different levels of marginal bone loss, from the ideal case consisting in the absence of cratering effects (bone loss equal to 0% in thickness of the cortical bone layer) up to the case of 50% bone loss. For the sake of compactness, in Figure 12 only peak and mean values of the Von Mises stress measure computed over Ω_c and Ω_t are shown, together with results computed for the overloading risk index R.

Numerical analyses showed that modeling an increase in cratering depth induced an increase in stress levels at both cortical and trabecular peri-implant regions and thereby induced an increase in the risk of overloading. In particular, for both implants, the Von Mises stress peaks relevant to a crestal bone loss of 50% in thickness were greater of about 120% in cortical bone and 105% in trabecular than those in the ideal case of 0% bone loss.

4. Discussion

The 11 dental implants that were analyzed by finite-element simulations exhibited different stress-based biomechanical behaviours, dependent on implant shape and thread, as well as on positioning depth and bone geometry around the implant neck. Simulation results considered functioning implants based on platform-switching concept and were obtained by modeling the crestal bone geometry after a healing and loading period.

Numerical results obtained by considering a subcrestal in-bone positioning 1 mm depth of implants have highlighted the influence of implant length and diameter on load transfer mechanisms. In agreement with numerical findings obtained by other authors [37–41], an increase in implant diameter induced a significant reduction of stress peaks mainly at cortical bone, whereas the variation in implant length produced a certain influence only on stress patterns at the cancellous bone-implant interface. Accordingly, the present numerical results suggest that, in order to control overloading risk, the implant diameter can be considered as a more effective design parameter than the implant length. Similar findings were proposed in [40, 47] and were relevant also to traditional implants crestally positioned. Overloading risk, quantitatively estimated by combining compressive and tensile effects via a principal-stress-based strength criterion for bone, was computed as significant at the cortical region around the implant neck (mainly as a result of dominant compressive states induced by the lateral load component) and/or at crestal (dominant tensile states) or apical (dominant compressive states) trabecular regions (induced by the vertical intrusive load component).

Stress analyses of implants with similar length and diameter allowed us to investigate the influence of some thread features. In particular, the proposed numerical results suggest that thread shape and thread details can induce significant effects on the peri-implant stress patterns. Threads analyzed in this study were characterized by two starts and numerical results have shown that the use of the same thread truncation for both starts induced more uniform local stress distributions than the cases characterized by a different thread truncation. As regards the thread shape, trapezoid-shaped thread produced compressive and tensile states at both cortical and trabecular regions more favorable than those of the saw-tooth thread, leading to reductions in stress values that were significantly affected by implant length and diameter. Moreover, numerical evidence has highlighted that the presence of a wide helical-milling along the implant body does not significantly affect the loading transmission mechanisms, but it can contribute to reduce risks of overloading at the trabecular apical bone, especially when short implants are considered.

Numerical simulations carried out on coupled bone-implant models defined by considering different levels of the implant in-bone positioning depth have shown that a crestal placement, combined with a reduced marginal bone loss, induced great stress values at the crestal cortical regions, confirming the biomechanical relationship between the stress-based mechanical stimuli and the possible activation of bone resorption process at the implant collar [21]. In agreement with clinical evidence and with other numerical studies [4, 18, 19, 25–34, 40, 47, 53], present results confirm also that a subcrestal positioning of implants based on platform-switching concept may contribute to the preservation of the crestal bone as well as can induce more effective and homogeneous stress distributions at the peri-implant regions. In particular, proposed simulation results have shown that,

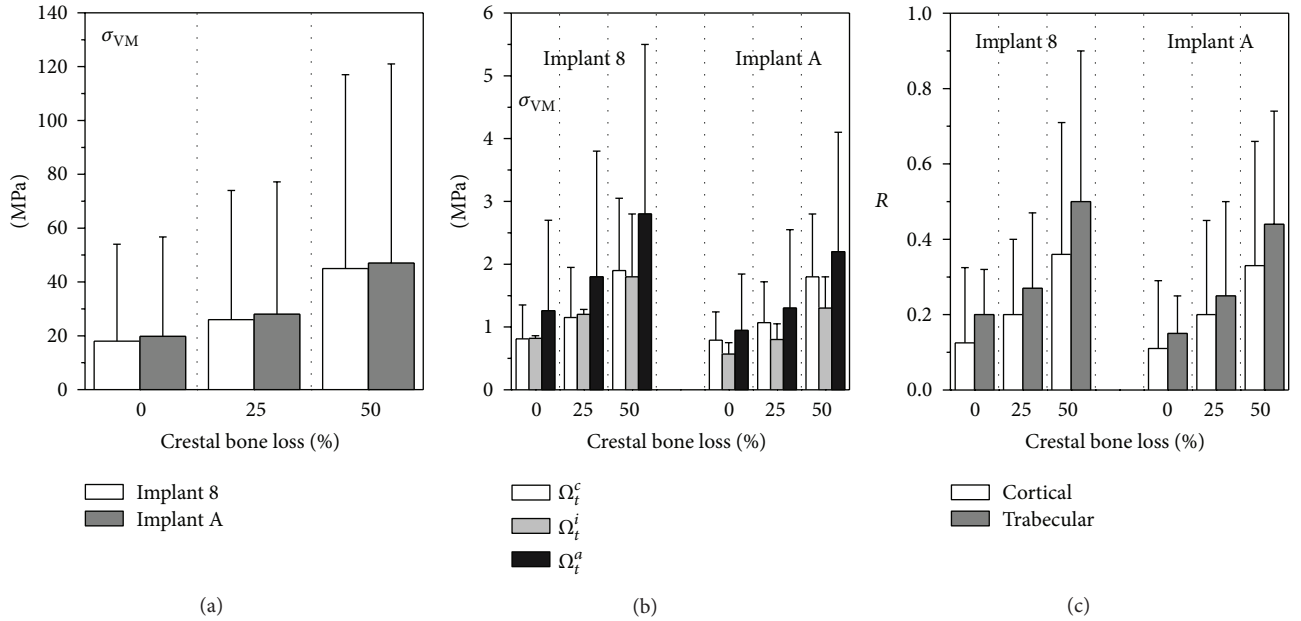


Figure 12: Von Mises stress measure at cortical (a) and trabecular (b) bone-implant interface for implants 8 and A (see Figure 1) and with a crestal positioning characterized by different levels of crestal bone loss. (c) Overloading risk index R. Average (bars) and peak (lines) values.

in the case of subcrestal placements, stress distributions were mainly affected by two counteracting effects. On one hand, when the implant's in-bone positioning depth increases then the vertical thickness of the cortical bone engaged in load transfer mechanisms reduces, tending to generate stress concentrations. But, on the other hand, the horizontal bone apposition induced by the platform-switching configuration in a subcrestal positioning highly contributes to an effective redistribution of the stress field. As a result of a balance condition between previous effects, the best stress-based performance among cases herein analyzed has been experienced considering an in-bone positioning depth of about 25% in cortical thickness.

In the case of crestal positioning, the proposed numerical results have shown that if the crestal bone morphology, affected by possible marginal bone loss, is not properly modeled, then a significant underestimation of stress values and an inaccurate evaluation of loading transfer mechanisms are generally obtained. Moreover, the present finite-element analyses have confirmed that a progressive marginal bone loss can lead to a progressive increase in stress intensity at the peri-implant interface that, in turn, can contribute to a further overload-induced bone loss, jeopardizing clinical effectiveness and durability of the prosthetic treatment. These results are qualitatively in agreement with numerical evidence obtained in [19, 40, 41, 47] although, due to simplified and/or different models used in those studies, quantitative comparisons cannot be made.

It is worth remarking that, contrary to a number of recent numerical approaches [33, 38, 39, 41, 46], the present study accounted for the influence of posthealing crestal bone morphology in functioning implants and was based on a detailed three-dimensional geometrical modeling of the bone segment wherein the implant is inserted. Accordingly, the results herein proposed can be retained as complementary with respect to several previous simplified studies, furnishing more refined and accurate indications for choosing and/or designing threaded dental implants, as well as giving clear insights towards the understanding of main factors affecting the loading transmission mechanisms.

Although in the current study a number of aspects influencing the biomechanical interaction between dental implant and bone have been accounted for, some limitations can be found in modeling assumptions herein employed. In particular, the ideal and unrealistic condition of 100% osseous integration was assumed; stress analyses were performed by simulating static loads and disregarding any muscle-jaw interaction; bone was modeled as a dry isotropic linear elastic material, whose mechanical properties were assumed to be time independent; the space dependence of bone density and mechanical response has been simply described by distinguishing trabecular and cortical homogeneous regions. All these assumptions do not completely describe possible clinical scenarios because of possible osseointegration defects at the peri-implant regions; different patient-dependent loading distributions; much more complex and time-dependent forces and significant muscular effects; anisotropic, inhomogeneous, nonlinear, and inelastic response of living tissues; bone remodeling; and spatially graded tissue properties. Nevertheless, in agreement with other numerical studies [35–54], present assumptions can be accepted in a computational sense in order to deduce significant and clinically useful indications for the comparative stress-based assessment of threaded dental implants.

In order to enhance the present finite-element approach, future studies will be devoted to the modeling of bone

as a nonlinear, anisotropic, viscous, and inhomogeneous regenerative tissue that responds to stress by resorption or regeneration under time-dependent muscular and external loads, accounting also for a more refined correlation between bone density and its mechanical response.

5. Concluding Remarks

Within the limitations of this study, numerical simulations showed that implant design (in terms of implant diameter, length, thread shape), in-bone positioning depth, and crestal bone morphology highly affect the mechanisms of load transmission. Aiming at the minimization of the overloading risks, the implant diameter can be retained as a more effective design parameter than the implant length. In particular, a significant reduction of stress peaks, mainly at the cortical bone, occurred when implant diameter increased. Nevertheless, implant length exhibited a certain influence on the bone-implant mechanical interaction at the cancellous interface, resulting in more effective and homogeneous stress distributions in trabecular bone when the implant length increased. Stress-based performances of dental implants were also found to be significantly affected by thread features. In detail, trapezoid-shaped thread induced compressive and tensile states at both cortical and trabecular regions more favorable than the saw-tooth thread. Moreover, the use of the same thread truncation for different thread starts induced a more uniform local stress distributions than the case of a different thread truncation. In the case of short implants, the presence of a wide helical-milling along the implant body produced a reduction in the overloading risk at the trabecular apical bone. Overloading risks were computed as high around the implant neck (for compressive states) in cortical bone and at the crestal (for tensile states) or apical (for compressive states) trabecular bone. Risk of overloading reduced when small levels of crestal bone loss were considered, as induced by suitable platform-switching strategies.

References

[1] T. D. Taylor, U. Belser, and R. Mericske-Stern, "Prosthodontic considerations," *Clinical oral Implants Research*, vol. 11, pp. 101–107, 2000.

[2] S. E. Eckert and P. C. Wollan, "Retrospective review of 1170 endosseous implants placed in partially edentulous jaws," *Journal of Prosthetic Dentistry*, vol. 79, no. 4, pp. 415–421, 1998.

[3] R. J. Weyant, "Short-term clinical success of root-form titanium implant systems," *Journal of Evidence-Based Dental Practice*, vol. 3, pp. 127–130, 2003.

[4] A. M. Roos-Jansåker, C. Lindahl, H. Renvert, and S. Renvert, "Nine- to fourteen-year follow-up of implant treatment. Part I: implant loss and associations to various factors," *Journal of Clinical Periodontology*, vol. 33, no. 4, pp. 283–289, 2006.

[5] J. B. Brunski, "Biomechanics of dental implants," in *Implants in Dentistry*, M. Block, J. N. Kent, and L. R. Guerra, Eds., pp. 63–71, W.B. Saunders, Philadelphia, Pa, USA, 1997.

[6] J. B. Brunski, D. A. Puleo, and A. Nanci, "Biomaterials and biomechanics of oral and maxillofacial implants: current status and future developments," *International Journal of Oral and Maxillofacial Implants*, vol. 15, no. 1, pp. 15–46, 2000.

[7] J. E. Lemons, "Biomaterials, biomechanics, tissue healing, and immediate-function dental implants," *The Journal of Oral Implantology*, vol. 30, no. 5, pp. 318–324, 2004.

[8] F. Marco, F. Milena, G. Gianluca, and O. Vittoria, "Peri-implant osteogenesis in health and osteoporosis," *Micron*, vol. 36, no. 7-8, pp. 630–644, 2005.

[9] M. Çehreli, S. Şahin, and K. Akça, "Role of mechanical environment and implant design on bone tissue differentiation: current knowledge and future contexts," *Journal of Dentistry*, vol. 32, no. 2, pp. 123–132, 2004.

[10] C. D. C. Lopes and B. K. Júnior, "Histological findings of bone remodeling around smooth dental titanium implants inserted in rabbit's tibias," *Annals of Anatomy*, vol. 184, no. 4, pp. 359–362, 2002.

[11] B. Helgason, E. Perilli, E. Schileo, F. Taddei, S. Brynjólfsson, and M. Viceconti, "Mathematical relationships between bone density and mechanical properties: a literature review," *Clinical Biomechanics*, vol. 23, no. 2, pp. 135–146, 2008.

[12] K. G. Strid, "Radiographic results," in *Tissue-Integrated Prostheses: Osseointegration in Clinical Dentistry*, P. I. Brånemark, G. A. Zarb, and T. Albrektsson, Eds., pp. 187–198, Quintessence, Chicago, Ill, USA, 1985.

[13] L. Sennerby, L. E. Ericson, P. Thomsen, U. Lekholm, and P. Astrand, "Structure of the bone-titanium interface in retrieved clinical oral implants," *Clinical Oral Implants Research*, vol. 2, no. 3, pp. 103–111, 1991.

[14] Y. Ujiie, R. Todescan, and J. E. Davies, "Peri-implant crestal bone loss: a putative mechanism," *International Journal of Dentistry*, vol. 2012, Article ID 742439, 14 pages, 2012.

[15] F. W. Neukam, T. F. Flemmig, C. Bain et al., "Local and systemic conditions potentially compromising osseointegration Consensus report of Working Group 3," *Clinical Oral Implants Research*, vol. 17, no. 2, pp. 160–162, 2006.

[16] S. Şahin, M. C. Çehreli, and E. Yalçin, "The influence of functional forces on the biomechanics of implant-supported prostheses—a review," *Journal of Dentistry*, vol. 30, no. 7-8, pp. 271–282, 2002.

[17] D. P. Callan, A. O'Mahony, and C. M. Cobb, "Loss of crestal bone around dental implants: a Retrospective Study," *Implant Dentistry*, vol. 7, no. 4, pp. 258–266, 1998.

[18] J. S. Hermann, D. L. Cochran, P. V. Nummikoski, and D. Buser, "Crestal bone changes around titanium implants. A radiographic evaluation of unloaded nonsubmerged and submerged implants in the canine mandible," *Journal of Periodontology*, vol. 68, no. 11, pp. 1117–1130, 1997.

[19] K. Akca and M. C. Cehreli, "Biomechanical consequences of progressive marginal bone loss around oral implants: a finite element stress analysis," *Medical and Biological Engineering and Computing*, vol. 44, no. 7, pp. 527–535, 2006.

[20] J. T. Irving, "Factors concerning bone loss associated with periodontal disease," *Journal of Dental Research*, vol. 49, no. 2, pp. 262–267, 1970.

[21] D. R. Carter, M. C. H. Van Der Meulen, and G. S. Beaupré, "Mechanical factors in bone growth and development," *Bone*, vol. 18, no. 1, pp. 5S–10S, 1996.

[22] A. Kozlovsky, H. Tal, B.-Z. Laufer et al., "Impact of implant overloading on the peri-implant bone in inflamed and non-inflamed peri-implant mucosa," *Clinical Oral Implants Research*, vol. 18, no. 5, pp. 601–610, 2007.

[23] S. C. Cowin, *Bone Mechanics Handbook*, CRC Press, Boca Raton, Fla, USA, 2001.

[24] R. B. Martin, D. B. Burr, and N. A. Sharkey, *Skeletal Tissue Mechanics*, Springer, New York, NY, USA, 1998.

[25] B. Assenza, A. Scarano, G. Petrone et al., "Crestal bone remodeling in loaded and unloaded implants and the microgap: a histologic study," *Implant Dentistry*, vol. 12, no. 3, pp. 235–241, 2003.

[26] M. P. Hänggi, D. C. Hänggi, J. D. Schoolfield, J. Meyer, D. L. Cochran, and J. S. Hermann, "Crestal bone changes around titanium implants. Part I: a retrospective radiographic evaluation in humans comparing two non-submerged implant designs with different machined collar lengths," *Journal of Periodontology*, vol. 76, no. 5, pp. 791–802, 2005.

[27] Y.-K. Shin, C.-H. Han, S.-J. Heo, S. Kim, and H.-J. Chun, "Radiographic evaluation of marginal bone level around implants with different neck designs after 1 year," *International Journal of Oral and Maxillofacial Implants*, vol. 21, no. 5, pp. 789–794, 2006.

[28] J. S. Hermann, J. D. Schoolfied, R. K. Schenk, D. Buser, and D. L. Cochran, "Influence of the size of the microgap on crestal bone changes around titanium implants. A histometric evaluation of unloaded non-submerged implants in the canine mandible," *Journal of Periodontology*, vol. 72, no. 10, pp. 1372–1383, 2001.

[29] F. Hermann, H. Lerner, and A. Palti, "Factors influencing the preservation of the periimplant marginal bone," *Implant Dentistry*, vol. 16, no. 2, pp. 165–175, 2007.

[30] L. López-Marí, J. L. Calvo-Guirado, B. Martín-Castellote, G. Gomez-Moreno, and M. López-Marí, "Implant platform switching concept: an updated review," *Medicina Oral, Patologia Oral y Cirugia Bucal*, vol. 14, no. 9, pp. e450–e454, 2009.

[31] D. M. Gardner, "Platform switching as a means to achieving implant esthetics," *The New York State Dental Journal*, vol. 71, no. 3, pp. 34–37, 2005.

[32] R. J. Lazzara and S. S. Porter, "Platform switching: a new concept in implant dentistry for controlling postrestorative crestal bone levels," *International Journal of Periodontics and Restorative Dentistry*, vol. 26, no. 1, pp. 9–17, 2006.

[33] Y. Maeda, J. Miura, I. Taki, and M. Sogo, "Biomechanical analysis on platform switching: is there any biomechanical rationale?" *Clinical Oral Implants Research*, vol. 18, no. 5, pp. 581–584, 2007.

[34] M. Degidi, A. Piattelli, J. A. Shibli, R. Strocchi, and G. Iezzi, "Bone formation around a dental implant with a platform switching and another with a TissueCare Connection. A histologic and histomorphometric evaluation in man," *Titanium*, vol. 1, pp. 10–17, 2009.

[35] J.-P. A. Geng, K. B. C. Tan, and G.-R. Liu, "Application of finite element analysis in implant dentistry: a review of the literature," *Journal of Prosthetic Dentistry*, vol. 85, no. 6, pp. 585–598, 2001.

[36] R. C. Van Staden, H. Guan, and Y. C. Loo, "Application of the finite element method in dental implant research," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 9, no. 4, pp. 257–270, 2006.

[37] H.-J. Chun, S.-Y. Cheong, J.-H. Han et al., "Evaluation of design parameters of osseointegrated dental implants using finite element analysis," *Journal of Oral Rehabilitation*, vol. 29, no. 6, pp. 565–574, 2002.

[38] L. Himmlová, T. Dostálová, A. Kácovsky, and S. Konvicková, "Influence of implant length and diameter on stress distribution: a finite element analysis," *Journal of Prosthetic Dentistry*, vol. 91, pp. 20–25, 2004.

[39] C. S. Petrie and J. L. Williams, "Comparative evaluation of implant designs: influence of diameter, length, and taper on strains in the alveolar crest—a three-dimensional finite-element analysis," *Clinical Oral Implants Research*, vol. 16, no. 4, pp. 486–494, 2005.

[40] L. Baggi, I. Cappelloni, M. Di Girolamo, F. Maceri, and G. Vairo, "The influence of implant diameter and length on stress distribution of osseointegrated implants related to crestal bone geometry: a three-dimensional finite element analysis," *Journal of Prosthetic Dentistry*, vol. 100, no. 6, pp. 422–431, 2008.

[41] D. Bozkaya, S. Muftu, and A. Muftu, "Evaluation of load transfer characteristics of five different implants in compact bone at different load levels by finite elements analysis," *Journal of Prosthetic Dentistry*, vol. 92, no. 6, pp. 523–530, 2004.

[42] H.-J. Chun, H.-S. Shin, C.-H. Han, and S.-H. Lee, "Influence of implant abutment type on stress distribution in bone under various loading conditions using finite element analysis," *International Journal of Oral and Maxillofacial Implants*, vol. 21, no. 2, pp. 195–202, 2006.

[43] I. Alkan, A. Sertgöz, and B. Ekici, "Influence of occlusal forces on stress distribution in preloaded dental implant screws," *Journal of Prosthetic Dentistry*, vol. 91, no. 4, pp. 319–325, 2004.

[44] G. Sannino, G. Marra, L. Feo, G. Vairo, and A. Barlattani, "3D finite element non linear analysis on the stress state at the bone-implant interface in dental osteointegrated implants," *Oral & Implantology*, vol. 3, no. 3, pp. 26–37, 2010.

[45] C.-L. Lin, Y.-C. Kuo, and T.-S. Lin, "Effects of dental implant length and bone quality on biomechanical responses in bone around implants: a 3-D non-linear finite element analysis," *Biomedical Engineering*, vol. 17, no. 1, pp. 44–49, 2005.

[46] T. Kitagawa, Y. Tanimoto, K. Nemoto, and M. Aida, "Influence of cortical bone quality on stress distribution in bone around dental implant," *Dental Materials Journal*, vol. 24, no. 2, pp. 219–224, 2005.

[47] L. Baggi, I. Cappelloni, F. Maceri, and G. Vairo, "Stress-based performance evaluation of osseointegrated dental implants by finite-element simulation," *Simulation Modelling Practice and Theory*, vol. 16, no. 8, pp. 971–987, 2008.

[48] F. Chen, K. Terada, K. Hanada, and I. Saito, "Anchorage effects of a palatal osseointegrated implant with different fixation: a finite element study," *Angle Orthodontist*, vol. 75, no. 4, pp. 593–601, 2005.

[49] H.-J. Chun, D.-N. Park, C.-H. Han, S.-J. Heo, M.-S. Heo, and J.-Y. Koak, "Stress distributions in maxillary bone surrounding overdenture implants with different overdenture attachments," *Journal of Oral Rehabilitation*, vol. 32, no. 3, pp. 193–205, 2005.

[50] A. N. Natali, P. G. Pavan, and A. L. Ruggero, "Evaluation of stress induced in peri-implant bone tissue by misfit in multi-implant prosthesis," *Dental Materials*, vol. 22, no. 4, pp. 388–395, 2006.

[51] M. Bevilacqua, T. Tealdo, M. Menini et al., "The influence of cantilever length and implant inclination on stress distribution in maxillary implant-supported fixed dentures," *Journal of Prosthetic Dentistry*, vol. 105, no. 1, pp. 5–13, 2011.

[52] C. M. Bellini, D. Romeo, F. Galbusera et al., "A finite element analysis of tilted versus nontilted implant configurations in the edentulous Maxilla," *International Journal of Prosthodontics*, vol. 22, no. 2, pp. 155–157, 2009.

[53] L. Baggi, S. Pastore, M. Di Girolamo, and G. Vairo, "Implant-bone load transfer mechanisms in complete-arch prostheses supported by four implants: a three-dimensional finite element approach," *Journal of Prosthetic Dentistry*, vol. 109, pp. 9–21, 2013.

[54] G. Sannino and A. Barlattani, "Mechanical evaluation of an implant-abutment self-locking taper connection: finite element analysis and experimental tests," *International Journal of Oral & Maxillofacial Implants*, vol. 28, no. 1, pp. e17–e26, 2013.

[55] J. Y. Rho, M. C. Hobatho, and R. B. Ashman, "Relations of mechanical properties to density and CT numbers in human bone," *Medical Engineering and Physics*, vol. 17, no. 5, pp. 347–355, 1995.

[56] J. E. Lemon and F. Dietsh-Misch, "Biomaterials for dental implants," in *Contemporary Implant Dentistry*, C. E. Misch, Ed., pp. 271–302, Mosby, St. Louis, Mo, USA, 2nd edition, 1999.

[57] U. Lekholm and G. A. Zarb, "Patient selection and preparation," in *Tissue-Integrated Prostheses: Osseointegration in Clinical Dentistry*, P. I. Branemark, G. A. Zarb, and T. Albrektsson, Eds., pp. 199–209, Quintessence, Chicago, Ill, USA, 1985.

[58] A. N. Natali, R. T. Hart, P. G. Pavan, and I. Knets, "Mechanics of bone tissue," in *Dental Biomechanics*, A. N. Natali, Ed., pp. 1–19, Taylor & Francis, London, UK, 2003.

[59] J. Y. Rho, R. B. Ashman, and H. Turner, "Young's modulus of trabecular and cortical bone material: ultrasonic and microtensile measurements," *Journal of Biomechanics*, vol. 26, no. 2, pp. 111–119, 1993.

[60] C. Truesdell and R. A. Toupin, "The classical field theories," in *Handbuch Der Physik*, S. Flügge, Ed., vol. 3, Springer, Berlin, Germany, 1960.

[61] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, McGraw-Hill, New York, NY, USA, 4th edition, 1998.

[62] O. C. Zienkiewicz and J. Z. Zhu, "Simple error estimator and adaptive procedure for practical engineering analysis," *International Journal for Numerical Methods in Engineering*, vol. 24, no. 2, pp. 337–357, 1987.

[63] F. Maceri, M. Martignoni, and G. Vairo, "Mechanical behaviour of endodontic restorations with multiple prefabricated posts: a finite-element approach," *Journal of Biomechanics*, vol. 40, no. 11, pp. 2386–2398, 2007.

[64] F. Maceri, M. Martignoni, and G. Vairo, "Optimal mechanical design of anatomical post-systems for endodontic restoration," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 12, no. 1, pp. 59–71, 2009.

[65] X. E. Guo, "Mechanical properties of cortical and cancellous bone tissue," in *Bone Mechanics Handbook*, S. C. Cowin, Ed., pp. 10.1–10.23, CRC Press, Boca Raton, Fla, USA, 2nd edition, 2001.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 295986, 7 pages
http://dx.doi.org/10.1155/2013/295986

Research Article

Effect of Pilates Training on Alpha Rhythm

Zhijie Bian,[1] Hongmin Sun,[2] Chengbiao Lu,[1] Li Yao,[3] Shengyong Chen,[4] and Xiaoli Li[3]

[1] *Institute of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China*
[2] *College of Physical Education, Yanshan University, Qinhuangdao 066004, China*
[3] *National Lab of Cognitive Neuroscience and Learning, Beijing Normal University, Xin Jie Kou Wai Avenue, Haidian District, Beijing 100875, China*
[4] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Xiaoli Li; xiaoli@bnu.edu.cn

Received 13 April 2013; Accepted 26 May 2013

Academic Editor: Carlo Cattani

In this study, the effect of Pilates training on the brain function was investigated through five case studies. Alpha rhythm changes during the Pilates training over the different regions and the whole brain were mainly analyzed, including power spectral density and global synchronization index (GSI). It was found that the neural network of the brain was more active, and the synchronization strength reduced in the frontal and temporal regions due to the Pilates training. These results supported that the Pilates training is very beneficial for improving brain function or intelligence. These findings maybe give us some line evidence to suggest that the Pilates training is very helpful for the intervention of brain degenerative diseases and cogitative dysfunction rehabilitation.

1. Introduction

Pilates was created in the 1920s by physical trainer Joseph H. Pilates and has been developed based on the Eastern and Western health preservation methods, such as Yoga and Taichi. This exercise is suitable for all the people and may be one of the most attractive fitness trainings [1, 2]. Pilates exercise was found to be able to correct body posture, relax the waist and neck, solve the problem of shoulder, and reduce fat of arm and abdomen [3–5]. Pilates can improve the blood circulation and cardiopulmonary function as the exercise is dominated by the rhythmic breath, particularly the lateral thoracic breathing that can effectively promote the exchange of oxygen. The Pilates has been proven to impact personal autonomy [6], pain control [1], improved muscle strength [7], flexibility [8], and motor skills [9]. Physical activity can be considered as an approach to improve organic conditions and prevent physical degeneration [10]. Further studies suggest that Pilates can release the stress of mind, increase brain's oxygen supply, and enhance brain function [11, 12], and studies in aged samples also suggest that Pilates is beneficial to mental state, including sleep quality, emotion, and self-confidence [2].

However, the direct evidence of Pilates on brain activity such as electroencephalographic (EEG) is lacking. In this study, we recorded resting-state EEG signals before and after Pilates exercise. We concentrated on the analysis of alpha rhythm (8–13 Hz) changes of the EEG, which is associated with the intelligence. The aim is to demonstrate whether or not Pilates can impact the brain functions or intelligence.

2. Methods

2.1. Subjects. After providing informed consent, five healthy postgraduate girls (mean age 24 ± 1 years) voluntarily participated in this study. They were free to withdraw from the experiments at any time. All subjects included in this experiment were right-handed, nonathletes, and had never been suffering from neurological and psychiatric disorders. The study was approved by the local ethics committee, and all participants gave written informed consent for this study.

2.2. Pilates Training. The five girls were trained with Pilates four sessions a week (Monday, Tuesday, Thursday, and Friday) in a well-ventilated room, at least 90 minutes per session. For the first three weeks, they were taught Pilates movements

step by step, and they reviewed the former movements in each training session and were corrected by the coach after learning the new ones. After they were taught a total of 24 movements, they practiced for 4–6 times in each session, and they were instructed to perform the sequences as accurately and smoothly coupled with breathing. The training lasted for 10 weeks. And the resting-state EEG rhythms were recorded with eyes closed before Pilates training and after each two weeks training.

2.3. Data Acquisition. EEG recordings were performed at six different time points. The first recording was performed just prior to the onset of training week (week 0). After each two weeks training, there was one recording, such as week 2, week 4, week 6, week 8, and week 10. During recordings, the subjects were asked to close their eyes and sit in a comfortable armchair, who were relaxed and awake in a dim room for 5 minutes during each recording.

The EEG data acquisition was performed with Neuroscan EEG/ERP recording system amplifiers (SynAmps2) with 64 Ag/AgCl surface electrodes, which were fixed in a cap at the standard positions according to the extended international 10–20 system, and with 32 bit SCAN4.5 acquisition system that could also be used to continuously view the EEG recordings. A reference electrode was placed between Cz and CPz, and ground electrode was placed between FPz and Fz. Horizontal and vertical electrooculograms (EOG) were recorded as well. The EEG was recorded with unipolar montages except for the EOG with bipolar montages. The impedances of all electrodes were <10 kΩ. During the recording, the data was band-pass filtered in the frequency range 0.05–200 Hz and sampled at 2 KHz. Digital conversion of the measured analog signals was accomplished with a 24 bit digitizer.

2.4. Data Analysis. In this study, the alpha rhythm (8–13 Hz) in the EEG recordings was concentrated on. In order to detect the alpha rhythm's changes over different regions, the brain was divided into five regions: frontal, left temporal, central, right temporal, and posterior (see Figure 1). Power spectral density and global synchronization index (GSI) at the alpha frequency band were computed in all regions.

2.4.1. Preprocessing for EEG. The raw EEG data was analyzed offline using EEGLAB (http://sccn.ucsd.edu/eeglab/ [13]). It was rereferenced to M1 (left mastoid process) and M2 (right mastoid process), the two EOG channels were extracted, the band-pass filter (8–13 Hz) was initially used to include the frequency band of interest, and then the data was resampled to 250 Hz for further analysis.

2.4.2. Spectral Analysis. After preprocessing, we chose EEG data of 4 minutes for analysis. Power spectral density (PSD) was estimated using pwelch method, which has a better noise performance compared with other power spectra estimation methods. The PSD was calculated using 10s epochs for each signal. Each epoch was divided into overlapping segments using periodic 10-s hamming window with 50% overlap. And then the peak power and peak power frequency were calculated for the alpha band in each epoch. Outliers rejection was performed using generalized extreme studentized deviate (GESD) [14] for all epochs in each channel. The remained epochs were averaged.

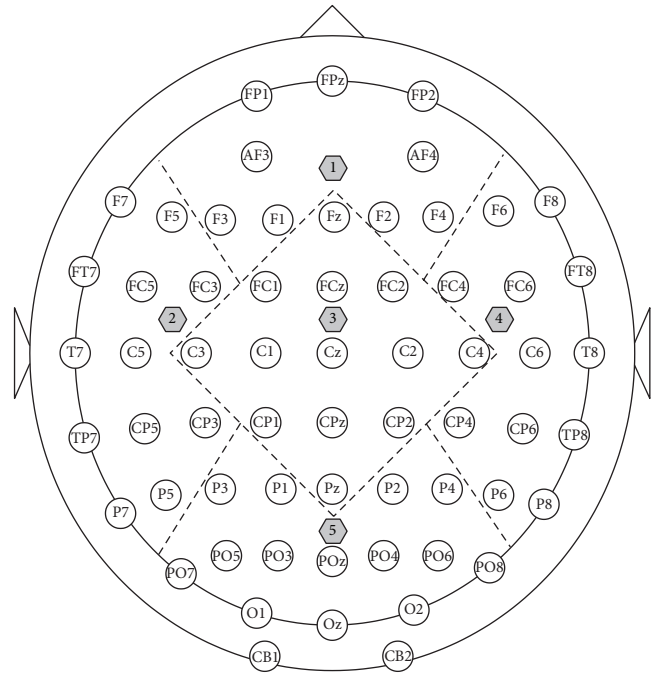


Figure 1: Extended 10–20 electrodes system and area electrodes' partition. The dotted lines divided the whole into 5 regions: the numbers 1, 2, 3, 4, and 5 separately denote the frontal, left temporal, central, right temporal, and posterior regions, respectively.

The PSD for each channel in all frequency bands was obtained. In order to estimate the changes of peak power and corresponding frequency during the Pilates training over different regions and the whole brain, the PSD was averaged over each region and the whole brain.

2.4.3. GSI. Synchronization is known as a key feature to evaluate the information process in the brain. For long EEG data, global synchronization index (GSI) can reveal the true synchronization features of multivariable EEG sequences better than other methods [15].

To eliminate the effect of amplitude, the EEG signals preprocessed need to be normalized by

$$Z = \{z_i(n)\} \quad (i = 1, \ldots, M; n = 1, \ldots, T),$$
$$x_i(n) = \frac{(z_i(n) - \langle Z_i \rangle)}{\sigma_i}, \tag{1}$$
$$X = \{x_i(n)\},$$

where Z is considered as the multivariate EEG data, M is the number of channels, n is the number of data points in time window T, $x_i(n)$ is the normalized signal, and X is a vector of $x_i(n)$, and $\langle Z_i \rangle$ and σ_i are the mean and standard deviation of $z_i(n)$, respectively.

Table 1: Comparisons of global changes before training (BT) and after training (AT) for each case.

Persons	Changes					
	Alpha peak power		Alpha peak frequency		GSI	
	BT (μV^2/Hz)	AT (μV^2/Hz)	BT (Hz)	AT (Hz)	BT	AT
First	209.26	213.47 ± 32.79	10.05	10.02 ± 0.06	0.53	0.43 ± 0.03
Second	6.53	9.67 ± 1.27	9.23	9.76 ± 0.09	0.37	0.31 ± 0.03
Third	3.55	3.91 ± 0.52	11.89	11.48 ± 0.25	0.32	0.28 ± 0.02
Forth	45.06	65.95 ± 10.97	10.23	9.61 ± 0.08	0.35	0.32 ± 0.05
Fifth	44.28	57.34 ± 9.25	10.06	10.06 ± 0.06	0.34	0.29 ± 0.02
Average	61.74	70.07 ± 10.96	10.29	10.18 ± 0.11	0.38	0.33 ± 0.03

To calculate the GSI of multivariate EEG data, a phase correlation matrix **C** was constructed. The phase of the each EEG series is estimated using continuous wavelet transform. The phase difference of two EEG traces is defined by

$$\Delta\varphi^{w}_{x_i x_k}(s,\tau) = \varphi^{w}_{x_i}(s,\tau) - \varphi^{w}_{x_k}(s,\tau) \quad (k = 1,\ldots,M). \tag{2}$$

Then, the phase synchronization is calculated by

$$\gamma_{ik} = \left|\left\langle e^{j\Delta\varphi^{w}_{x_i x_k}(s,\tau)}\right\rangle_T\right| \in [0,1], \tag{3}$$

where $\langle\cdot\rangle_T$ indicates the average of the time window T. γ_{ik} indicates the phase synchronization of signals $x_i(n)$ and $x_k(n)$. For all EEG series, a phase correlation matrix can be written as $\mathbf{C} = \{\gamma_{ik}\}$.

Then, the eigenvalue decomposition of **C** is defined as follows:

$$\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \tag{4}$$

where eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_M$ are in increasing order and $\mathbf{v}_i, i = 1,\ldots,M$ are the corresponding eigenvectors.

In order to reduce the "bias" caused by the algorithm and length of data, amplitude adjusted Fourier transformed (AAFT) surrogate method [16] was used in this study. Based on the surrogate series X_{surr}, the normalized phase surrogate correlation matrix **R** was calculated, and the $\lambda^s_1 \le \lambda^s_2 \le \cdots \le \lambda^s_M$ were the eigenvalues of surrogate correlation matrix **R**. The distribution of the surrogate eigenvalues can reflect the random synchronization of the multivariate time series. To reduce the effects of the random components in the total synchronization, the eigenvalues were divided by the averaged surrogate eigenvalues. The GSI was calculated by

$$\lambda^g_i = \frac{\lambda_i/\overline{\lambda^s_i}}{\sum_{i=1}^{M}\lambda_i/\overline{\lambda^s_i}} \quad (i = 1,\ldots,M),$$
$$\text{GSI} = 1 + \frac{\sum_{i=1}^{M}\lambda^g_i \log(\lambda^g_i)}{\log(M)}, \tag{5}$$

where $\overline{\lambda^s_i}$ is the averaged eigenvalues of the surrogate series.

Calculating the GSI used 10 s epochs with 50% overlap for the alpha rhythm over the five regions and the whole brain. Outlier's rejection [14] was also used, and then the remained epochs were averaged. Average of GSI over different regions and the whole brain was obtained as well.

2.4.4. Calculation of the Relative Variable Ratio. In order to estimate the changes during the Pilates training, the relative variable ratio may be calculated by

$$r^{(k)}_{ji} = \frac{y^{(k)}_{ji} - y^{(k)}_{j1}}{y^{(k)}_{j1}} \tag{6}$$
$$(i = 1,\ldots,N, N = 6; j = 1,\ldots K, K = 5; k = 1,2,3),$$

where N is the number of tests, K is the number of subjects, and $r^{(k)}_{ji}$ is the relative variable ratio to the first test. $y^{(k)}_{ji}$ is the feature value of EEG recordings. When $k = 1, r^{(k)}_{ji}$ presents the changes of the peak power; when $k = 2, r^{(k)}_{ji}$ presents the changes of the peak frequency; when $k = 3, r^{(k)}_{ji}$ presents the changes of GSI. All changes were over the Pilates training.

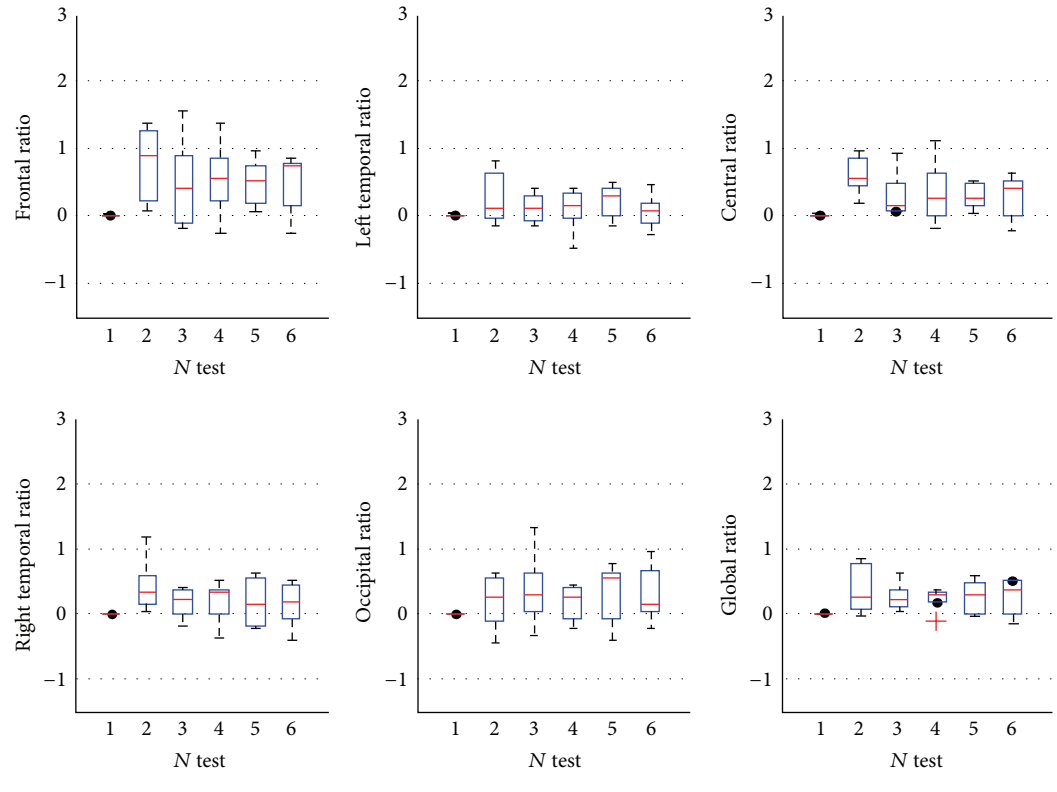
If the variables increased over the Pilates training, $r^{(k)}_{ji}$ will be greater than zero; if they decreased, $r^{(k)}_{ji}$ will be less than zero; if there are no changes, $r^{(k)}_{ji}$ will be approximate to zero. For the limited numbers of only five subjects, boxplot is used to describe the changes over the Pilates training duration.

3. Results

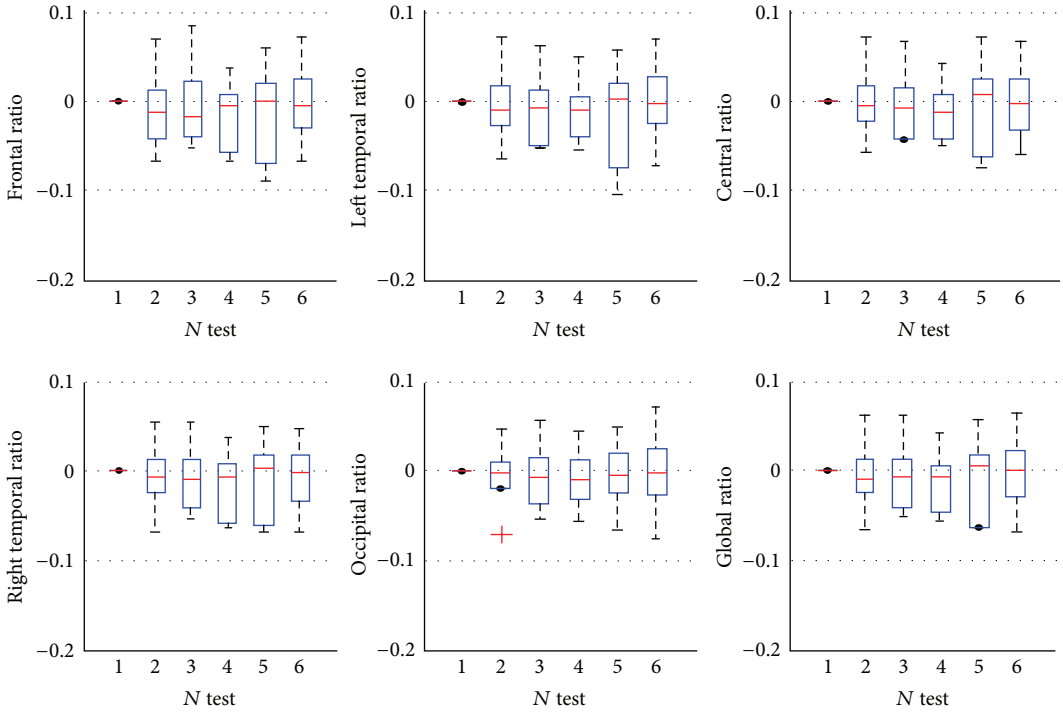
3.1. Spectral Analysis. The results of alpha peak power and alpha peak frequency in each region and over the whole brain were shown in Figure 2. The comparisons of global changes before training (BT) and after training (AT) for each case were shown in Table 1.

The alpha peak powers were different among the five cases. The power that is in the first case was the largest. A relative lower peak power was observed in the second and the third cases. There may be individual difference, but the trend of changes was the same. Table 1 presented that the alpha peak power increased in all cases and the average value increased as well (61.74 to 70.07 ± 10.96) (Table 1). The changes of alpha peak frequencies varied among different individuals: decreased in three cases, increased in one case, and unchanged in one case, and the average value was slightly decreased (10.29 to 10.18 ± 0.11) (Table 1).

The ratios of alpha peak power and alpha peak frequency could eliminate the effect of individual factor (see Figure 2). The ratios were obtained to investigate the two indicators' changes during Pilates training. Figure 2(a) showed that alpha peak power was increased in various regions and



(a) Alpha Peak Power



(b) Alpha Peak Frequency

Figure 2: Relative changes of alpha peak power (a) and peak frequency (b) during the Pilates training. Alpha peak power increased in the five regions and the whole brain as (a) shows. As (b) shows, most of the median of alpha peak frequency decreased but was not significant. One box represented one test in (a) and (b).

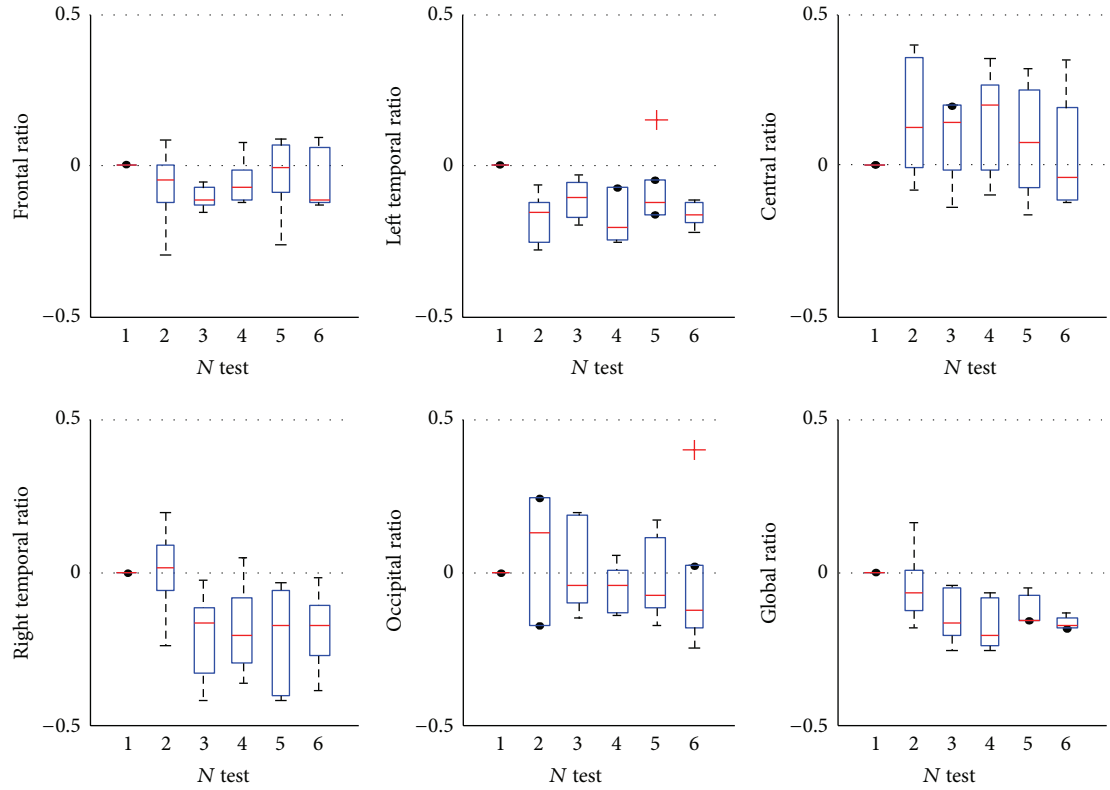


Figure 3: Relative changes of GSI for alpha rhythm during the Pilates training. The GSI in the frontal and temporal regions was decreased, but it almost increased in the central region, and the changes in the occipital region were not obvious. The GSI over the whole brain decreased obviously. One box represented one test.

the whole brain. The median of ratios was greater than zero. The ratios of alpha peak power versus alpha peak frequency were increased by about 30% to 90%, (especially in the second test, which was two weeks after Pilates training), 10% to 30%, 10% to 60%, and 20% to 40%, for the frontal, temporal, central, occipital, and the whole brain, respectively. The alpha peak frequency decreased in small degree during Pilates training, and the changes were not statistically significant (see Figure 2(b)).

3.2. GSI. The GSI changes of the whole brain before and after pilates training in individuals and the average value of the five subjects were listed in Table 1. The GSI values were decreased during the Pilates training significantly.

The time-dependent changes of GSI during the Pilates training in different regions and over the whole brain were also studied. Figure 3 plotted the relative variable ratios of GSI. For the frontal region, the GSI has decreased by about 0–10%, 8%–10%, and 5% after two, four, and six weeks training, respectively, but increased in some subjects after eight weeks training. For the left temporal region, the GSI decreased at least by 5–25% after two weeks training. For the right temporal region, the GSI decreased at least by 5–40% after four weeks training, but there was inconsistent variation after the two weeks training. For the central region, the GSI increased in varying degrees after two weeks training. For the occipital region, there were no consistent changes during Pilates training. For the whole area of the brain, the GSI decreased slightly after two weeks training but decreased at least by 5% after four weeks training.

4. Discussions

In this study, we used the resting-state EEG recording to investigate the effects of the Pilates training on the brain EEG. The results showed that the Pilates training could increase the power of the brain alpha rhythm and reduce the synchronization strength of alpha rhythm in the frontal and temporal regions. These findings may support that the Pilates training maybe beneficial for improving brain function because the alpha rhythm and its synchronization are associated with the human brain higher function such as intelligence. These results suggest that Pilates training may be helpful for the intervention of brain degenerative diseases and cogitative dysfunction rehabilitation. Future study will demonstrate this hypothesis.

Human EEG activity reflects the synchronization of cortical pyramidal neurons. Alpha rhythm in the spontaneous EEG signals is an important predictor of the efficacy of cortical information processing during cognitive and sensorimotor demand [17]. Alpha rhythm is often considered as one of the indicators of the brain function and has a significant correlation with performance on memory tasks [18], and the alpha power is considered as an important parameter to represent neural activities and processing mechanisms [19]. Although the exact mechanisms of alpha

rhythm generation and its functional significance are not understood completely so far, there is increasing evidence that synchronized oscillatory activity in the cerebral cortex is essential for spatiotemporal coordination and integration of activity of anatomically distributed but functionally related neural elements [20]. Alpha power was positively correlated with intelligence variables, while some lower frequency bands negatively correlated with them [21]. The higher the absolute amplitude or power of the EEG, the stronger the background neural synchronization, then the better the cognitive performance [22], and the higher the IQ [23]. Lower alpha power is associated with many diseases, such as obsessive-compulsive disorder [24], Down's syndrome [25], Alzheimer' [26], and restless legs syndrome [27]. Patients with these diseases showed intelligence, memory loss, and alpha rhythm abnormalities [26]. There is also a correlation between alpha power and intelligence [21]. Cortical neural synchronization at the basis of eye-closed resting-state EEG rhythms was enhanced in elite karate athletes [28]. In this study, the alpha peak power was increased during the Pilates training, which suggests the increased neural network activity and perhaps the intelligence during the Pilates training.

Previous study found that right postcentral gyrus and bilateral supramarginal gyrus were sensitive to the motor skill training [29], and the functional connectivity in the right postcentral gyrus and right supramarginal gyrus strengthened from week 0 to week 2 and decreased from week 2 to week 4. The findings in these case studies are very similar to the above results, and the functional connectivity changes based on the resting-state EEG recordings are associated with motor skill learning. Another similar study also demonstrates that the frontoparietal network connectivity increased one week after two brief motor training sessions in a dynamic balancing task [30], and there is an association between structural grey matter alterations and functional connectivity changes in prefrontal and supplementary motor areas. The GSI is a synchronization method of reflecting the multichannel synchronization strength. As shown in Figure 3, the GSI values of the alpha rhythm decreased in varying degrees over the frontal and temporal regions, increased over the central region, and decreased over the whole brain for all cases after two weeks training. The frontal and temporal regions are associated with cognition (i.e., attention and planning), and the central region is motor related. Because the Pilates can improve the balance, control, and muscle strength [7], the GSI of alpha rhythm in the frontal and temporal regions decreased when the subjects were in the resting state, in which the subjects were in a very relaxed condition, without attention and planning procession. The reduction of the synchronization strength in those regions can support what is mentioned above. This study demonstrates that the Pilates training may improve the function of control.

Acknowledgments

This research was funded in part by the National Science Fund for Distinguished Young Scholars (61025019) and by the National Natural Science Foundation of China (81271422).

References

[1] K. Caldwell, M. Harrison, M. Adams, and N. T. Triplett, "Effect of Pilates and taiji quan training on self-efficacy, sleep quality, mood, and physical performance of college students," *Journal of Bodywork and Movement Therapies*, vol. 13, no. 2, pp. 155–163, 2009.

[2] V. Gladwell, S. Head, M. Haggar, and R. Beneke, "Does a program of pilates improve chronic non-specific low back pain?" *Journal of Sport Rehabilitation*, vol. 15, no. 4, pp. 338–350, 2006.

[3] N. H. Turner, "Simple Pilates techniques for back and abdomen muscles," Exercise: Pilates & Yoga, 2009, http://www.helium.com/.

[4] K. S. Keays, S. R. Harris, J. M. Lucyshyn, and D. L. MacIntyre, "Effects of pilates exercises on shoulder range of motion, pain, mood, and upper-extremity function in women living with breast cancer: a pilot study," *Physical Therapy*, vol. 88, no. 4, pp. 494–510, 2008.

[5] D. Curnow, D. Cobbin, J. Wyndham, and S. T. B. Choy, "Altered motor control, posture and the Pilates method of exercise prescription," *Journal of Bodywork and Movement Therapies*, vol. 13, no. 1, pp. 104–111, 2009.

[6] E. G. Johnson, A. Larsen, H. Ozawa, C. A. Wilson, and K. L. Kennedy, "The effects of Pilates-based exercise on dynamic balance in healthy adults," *Journal of Bodywork and Movement Therapies*, vol. 11, no. 3, pp. 238–242, 2007.

[7] J. M. Schroeder, J. A. Crussemeyer, and S. J. Newton, "Flexibilty and heart rate response to an acute Pilates reformer session," *Medicine and Science in Sports and Exercise*, vol. 34, no. 5, article S258, 2002.

[8] N. A. Segal, J. Hein, and J. R. Basford, "The effects of pilates training on flexibility and body composition: an observational study," *Archives of Physical Medicine and Rehabilitation*, vol. 85, no. 12, pp. 1977–1981, 2004.

[9] C. Lange, V. B. Unnithan, E. Larkam, and P. M. Latta, "Maximizing the benefits of Pilates-inspired exercise for learning functional motor skills," *Journal of Bodywork and Movement Therapies*, vol. 4, no. 2, pp. 99–108, 2000.

[10] B. J. May, "Mobility training for the older adult," *Topics in Geriatric Rehabilitation*, vol. 19, no. 3, pp. 191–198, 2003.

[11] W. McNeill, "Decision making in Pilates," *Journal of Bodywork and Movement Therapies*, vol. 15, no. 1, pp. 103–107, 2011.

[12] W. McNeill, "Neurodynamics for Pilates teachers," *Journal of Bodywork and Movement Therapies*, vol. 16, no. 3, pp. 353–358, 2012.

[13] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[14] J. E. Seem, "Using intelligent data analysis to detect abnormal energy consumption in buildings," *Energy and Buildings*, vol. 39, no. 1, pp. 52–58, 2007.

[15] D. Cui, X. Liu, Y. Wan, and X. Li, "Estimation of genuine and random synchronization in multivariate neural series," *Neural Networks*, vol. 23, no. 6, pp. 698–704, 2010.

[16] K. T. Dolan and M. L. Spano, "Surrogate for nonlinear time series analysis," *Physical Review E*, vol. 64, no. 4, part 2, Article ID 046128, 6 pages, 2001.

[17] V. K. Lim, J. P. Hamm, W. D. Byblow, and I. J. Kirk, "Decreased desychronisation during self-paced movements in frequency

bands involving sensorimotor integration and motor functioning in Parkinson's disease," *Brain Research Bulletin*, vol. 71, no. 1–3, pp. 245–251, 2006.

[18] E. A. Golubeva, *Individual Characteristics of Human Memory: A Psychophysiological Study*, Pedagogika, Moscow, Russia, 1980.

[19] T. Liu, J. Shi, D. Zhao, and J. Yang, "The relationship between EEG band power, cognitive processing and intelligence in school-age children," *Psychology Science Quarterly*, vol. 50, no. 2, pp. 259–268, 2008.

[20] A. Anokhin and F. Vogel, "EEG α rhythm frequency and intelligence in normal adults," *Intelligence*, vol. 23, no. 1, pp. 1–14, 1996.

[21] R. G. Schmid, W. S. Tirsch, and H. Scherb, "Correlation between spectral EEG parameters and intelligence test variables in school-age children," *Clinical Neurophysiology*, vol. 113, no. 10, pp. 1647–1656, 2002.

[22] W. Klimesch, "EEG α and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Research Reviews*, vol. 29, no. 2-3, pp. 169–195, 1999.

[23] R. W. Thatcher, D. North, and C. Biver, "EEG and intelligence: relations between EEG coherence, EEG phase delay and power," *Clinical Neurophysiology*, vol. 116, no. 9, pp. 2129–2141, 2005.

[24] Y. W. Shin, T. H. Ha, S. Y. Kim, and J. S. Kwon, "Association between EEG α power and visuospatial function in obsessive-compulsive disorder," *Psychiatry and Clinical Neurosciences*, vol. 58, no. 1, pp. 16–20, 2004.

[25] O. Devinsky, S. Sato, R. A. Conwit, and M. B. Schapiro, "Relation of EEG α background to cognitive function, brain atrophy, and cerebral metabolism in Down's syndrome. Age-specific changes," *Archives of Neurology*, vol. 47, no. 1, pp. 58–62, 1990.

[26] D. Arnaldi, G. Rodriguez, and A. Picco, "Brain functional network in Alzheimer's disease: diagnostic markers for diagnosis and monitoring," *International Journal of Alzheimer's Disease*, vol. 2011, Article ID 481903, 10 pages, 2011.

[27] Ş. Akpinar, "The primary restless legs syndrome pathogenesis depends on the dysfunction of EEG α activity," *Medical Hypotheses*, vol. 60, no. 2, pp. 190–198, 2003.

[28] C. Babiloni, N. Marzano, M. Iacoboni et al., "Resting state cortical rhythms in athletes: a high-resolution EEG study," *Brain Research Bulletin*, vol. 81, no. 1, pp. 149–156, 2010.

[29] L. Ma, S. Narayana, D. A. Robin, P. T. Fox, and J. Xiong, "Changes occur in resting state network of motor system during 4 weeks of motor skill learning," *NeuroImage*, vol. 58, no. 1, pp. 226–233, 2011.

[30] M. Taubert, G. Lohmann, D. S. Margulies, A. Villringer, and P. Ragert, "Long-term effects of motor training on resting-state networks and underlying brain structure," *NeuroImage*, vol. 57, no. 4, pp. 1492–1498, 2011.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 106867, 14 pages
http://dx.doi.org/10.1155/2013/106867

Research Article

Fast Discriminative Stochastic Neighbor Embedding Analysis

Jianwei Zheng, Hong Qiu, Xinli Xu, Wanliang Wang, and Qiongfang Huang

School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Correspondence should be addressed to Jianwei Zheng; zjw@zjut.edu.cn

Received 9 February 2013; Accepted 22 March 2013

Academic Editor: Carlo Cattani

Feature is important for many applications in biomedical signal analysis and living system analysis. A fast discriminative stochastic neighbor embedding analysis (FDSNE) method for feature extraction is proposed in this paper by improving the existing DSNE method. The proposed algorithm adopts an alternative probability distribution model constructed based on its K-nearest neighbors from the interclass and intraclass samples. Furthermore, FDSNE is extended to nonlinear scenarios using the kernel trick and then kernel-based methods, that is, KFDSNE1 and KFDSNE2. FDSNE, KFDSNE1, and KFDSNE2 are evaluated in three aspects: visualization, recognition, and elapsed time. Experimental results on several datasets show that, compared with DSNE and MSNP, the proposed algorithm not only significantly enhances the computational efficiency but also obtains higher classification accuracy.

1. Introduction

In recent years, dimensional reduction which can reduce the curse of dimensionality [1] and remove irrelevant attributes in high-dimensional space plays an increasingly important role in many areas. It promotes the classification, visualization, and compression of the high dimensional data. In machine learning, dimension reduction is used to reduce the dimension by mapping the samples from the high-dimensional space to the low-dimensional space. There are many purposes of studying it: firstly, to reduce the amount of storage, secondly, to remove the influence of noise, thirdly, to understand data distribution easily, and last but not least, to achieve good results in classification or clustering.

Currently, many dimensional reduction methods have been proposed, and they can be classified variously from different perspectives. Based on the nature of the input data, they are broadly categorized into two classes: linear subspace methods which try to find a linear subspace as feature space so as to preserve certain kind of characteristics of observed data, and nonlinear approaches such as kernel-based techniques and geometry-based techniques; from the class labels' perspective, they are divided into supervised learning and unsupervised learning; furthermore, the purpose of the former is to maximize the recognition rate between classes while the latter is for making the minimum of information loss. In addition, judging whether samples utilize local information or global information, we divide them into local method and global method.

We briefly introduce several existing dimensional reduction techniques. In the main linear techniques, principal component analysis (PCA) [2] aims at maximizing the variance of the samples in the low-dimensional representation with a linear mapping matrix. It is global and unsupervised. Different from PCA, linear discriminant analysis (LDA) [3] learns a linear projection with the assistance of class labels. It computes the linear transformation by maximizing the amount of interclass variance relative to the amount of intraclass variance. Based on LDA, marginal fisher analysis (MFA) [4], local fisher discriminant analysis (LFDA) [5], and maxmin distance analysis (MMDA) [6] are proposed. All of the three are linear supervised dimensional reduction methods. MFA utilizes the intrinsic graph to characterize the intraclass compactness and uses meanwhile the penalty graph to characterize interclass separability. LFDA introduces the locality to the LFD algorithm and is particularly useful for samples consisting of intraclass separate clusters. MMDA considers maximizing the minimum pairwise samples of interclass.

To deal with nonlinear structural data, which can often be found in biomedical applications [7–10], a number of nonlinear approaches have been developed for dimensional reduction. Among these kernel-based techniques and geometry-based techniques are two hot issues. Kernel-based techniques

attempt to obtain the linear structure of nonlinearly distributed data by mapping the original inputs to a high-dimensional feature space. For instance, kernel principal component analysis (kernel PCA) [11] is the extension of PCA using kernel tricks. Geometry-based techniques, in general, are known as manifold learning techniques such as isometric mapping (ISOMAP) [12], locally linear embedding (LLE) [13], Laplacian eigenmap (LE) [14], Hessian LLE (HLLE) [15], and local tangent space alignment (LTSA) [16]. ISOMAP is used for manifold learning by computing the pairwise geodesic distances for input samples and extending multidimensional scaling. LLE exploits the linear reconstructions to discover nonlinear structure in high-dimensional space. LE first constructs an undirected weighted graph, and then recovers the structure of manifold by graph manipulation. HLLE is based on sparse matrix techniques. As for LTSA, it begins by computing the tangent space at every point and then optimizes to find an embedding that aligns the tangent spaces.

Recently, stochastic neighbor embedding (SNE) [17] and extensions thereof have become popular for feature extraction. The basic principle of SNE is to convert pairwise Euclidean distances into probabilities of selecting neighbors to model pairwise similarities. As extension of SNE, t-SNE [18] uses Student's t-distribution to model pairwise dissimilarities in low-dimensional space and it alleviates the optimization problems and the crowding problem of SNE by the methods below: (1) it uses a symmetrized version of the SNE cost function with simpler gradients that was briefly introduced by Cook et al. [19], and (2) it employs a heavy-tailed distribution in the low-dimensional space. Subsequently, Yang et al. [20] systematically analyze the characteristics of the heavy-tailed distribution and the solutions to crowding problem. More recently, Wu et al. [21] explored how to measure similarity on manifold more accurately and proposed a projection approach called manifold-oriented stochastic neighbor projection (MSNP) for feature extraction based on SNE and t-SNE. MSNP employs Cauchy distribution rather than standard Student's t-distribution used in t-SNE. In addition, for the purpose of learning the similarity on manifold with high accuracy, MSNP uses geodesic distance for characterizing data similarity. Though MSNP has many advantages in terms of feature extraction, there is still a drawback in it: MSNP is an unsupervised method and lacks the idea of class label, so it is not suitable for pattern identification. To overcome the disadvantage of MSNP, we have done some preliminary work and presented a method called discriminative stochastic neighbor embedding analysis (DSNE) [22]. DSNE effectively resolves the problems above, but since it selects all the training samples as their reference points, it has high computational cost and is thus computationally infeasible for the large-scale classification tasks with high-dimensional features [23, 24]. On the basis of our previous research, we present a method called fast discriminative stochastic neighbor embedding analysis (FDSNE) to overcome the disadvantages of DSNE in this paper.

The rest of this paper is organized as follows: in Section 2, we introduce in detail the proposed FDSNE and briefly compare it with MSNP and DSNE in Section 3. Section 4 gives the nonlinear extension of FDSNE. Furthermore, experiments on various databases are presented in Section 5. Finally, Section 6 concludes this paper and several issues for future works are described.

2. Fast Discriminative Stochastic Neighbor Embedding Analysis

Consider a labeled data samples matrix as

$$\mathbf{X} = \left\{\mathbf{x}_1^1, \ldots, \mathbf{x}_{N_1}^1, \mathbf{x}_1^2, \ldots, \mathbf{x}_{N_2}^2, \ldots, \mathbf{x}_1^C, \ldots, \mathbf{x}_{N_C}^C\right\}, \quad (1)$$

where $\mathbf{x}_i^c \in R^d$ is a d-dimensional sample and means the ith sample in the cth class. C is the number of sample classes, N_c is the number of samples in the cth class, and $N = N_1 + N_2 + \cdots + N_C$.

In fact, the basic principle of FDSNE is the same as t-SNE which is to convert pairwise Euclidean distances into probabilities of selecting neighbors to model pairwise similarities [18]. Since the DSNE selects all the training samples as its reference points, it has high computational cost and is thus computationally infeasible for the large-scale classification tasks with high-dimensional features. So according to the KNN classification rule, we propose an alternative probability distribution function which makes the label of target sample determined by its first K-nearest neighbors in FDSNE. In this paper, $\mathrm{NH}_l(\mathbf{x}_i)$ and $\mathrm{NM}_l(\mathbf{x}_i)$ are defined. They, respectively, denote the lth-nearest neighbor of $\mathbf{x}_i$ from the same class and the different classes in the transformed space. Mathematically, the joint probability p_{ij} is given by

$$p_{ij} = \begin{cases} \dfrac{\exp\left(-d_{ij}^2/2\lambda^2\right)}{\sum_{t\in H_m}\exp\left(-d_{mt}^2/2\lambda^2\right)} & \forall j \in H_i \\ \dfrac{\exp\left(-d_{ij}^2/2\lambda^2\right)}{\sum_{t\in M_m}\exp\left(-d_{mt}^2/2\lambda^2\right)} & \forall j \in M_i \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

In formula (2), $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)}$ is the Euclidian distance between two samples $\mathbf{x}_i$ and $\mathbf{x}_j$, the parameter λ is the variance parameter of Gaussian which determines the value of p_{ij}, $H_i = \{j \mid 1 \le j \le N, 1 \le i \le N, \mathbf{x}_j = \mathrm{NH}_k(\mathbf{x}_i) \text{ and } 1 \le k \le K_1\}$, $H_m = \{t \mid 1 \le t \le N, 1 \le m \le N, \mathbf{x}_t = \mathrm{NH}_k(\mathbf{x}_m) \text{ and } 1 \le k \le K_1\}$, $M_i = \{j \mid 1 \le j \le N, 1 \le i \le N, \mathbf{x}_j = \mathrm{NH}_k(\mathbf{x}_i) \text{ and } 1 \le k \le K_2\}$, and $M_m = \{t \mid 1 \le t \le N, 1 \le m \le N, \mathbf{x}_t = \mathrm{NH}_k(\mathbf{x}_m) \text{ and } 1 \le k \le K_2\}$, and then the denominator in formula (2) means all of the reference points under selection from the same class or the different classes. In particular, the joint probability p_{ij} not only keeps symmetrical characteristics of the probability distribution matrix but also makes the probability value of interclass data to be 1 and the same for intraclass data.

For low-dimensional representations, FDSNE uses counterparts $\mathbf{y}_i$ and $\mathbf{y}_j$ of the high-dimensional datapoints $\mathbf{x}_i$ and

$\mathbf{x}_j$. It is possible to compute a similar joint probability via the following expression:

$$q_{ij} = \begin{cases} \dfrac{\left(1 + d_{ij}^2(\mathbf{A})\right)^{-1}}{\sum_{t \in H_m} \left(1 + d_{mt}^2(\mathbf{A})\right)^{-1}} & \forall j \in H_i \\ \dfrac{\left(1 + d_{ij}^2(\mathbf{A})\right)^{-1}}{\sum_{t \in M_m} \left(1 + d_{mt}^2(\mathbf{A})\right)^{-1}} & \forall j \in M_i \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

In what follows, we introduce the transformation by a linear projection: $\mathbf{y}_i = \mathbf{A}\mathbf{x}_i$ ($\mathbf{A} \in \mathbf{R}^{r \times d}$) so that $d_{ij}(\mathbf{A}) = \|\mathbf{y}_i - \mathbf{y}_j\| = \|\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{x}_j\| = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{A}^T \mathbf{A} (\mathbf{x}_i - \mathbf{x}_j)}$. Then by simple algebra formulation, formula (3) has the following equivalent expression:

$$q_{ij} = \begin{cases} \dfrac{\left(1 + \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_i - \mathbf{x}_j\right)\right)^{-1}}{\sum_{t \in H_m} \left(1 + \left(\mathbf{x}_m - \mathbf{x}_t\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_m - \mathbf{x}_t\right)\right)^{-1}} & \forall j \in H_i \\ \dfrac{\left(1 + \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_i - \mathbf{x}_j\right)\right)^{-1}}{\sum_{t \in M_m} \left(1 + \left(\mathbf{x}_m - \mathbf{x}_t\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_m - \mathbf{x}_t\right)\right)^{-1}} & \forall j \in M_i \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Note that all data have the intrinsic geometry distribution and there is no exception for intraclass samples and interclass samples. Then the same distribution is required to hold in feature space. Since the Kullback-Leiber divergence [25] is wildly used to quantify the proximity of two probability distributions, we choose it to build our penalty function here. Based on the above definition, the function can be formulated as:

$$\min C(\mathbf{A}) = \sum_{\forall j \in H_i} p_{ij} \log \frac{p_{ij}}{q_{ij}} + \sum_{\forall j \in M_i} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \tag{5}$$

In this work, we use the conjugate gradient method to minimize $C(\mathbf{A})$. In order to make the derivation less cluttered, we first define four auxiliary variables w_{ij}, u_{ij}, u_{ij}^H, and u_{ij}^M as:

$$\begin{aligned} w_{ij} &= \left[1 + \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_i - \mathbf{x}_j\right)\right]^{-1}, \\ u_{ij} &= \left(p_{ij} - q_{ij}\right) w_{ij}, \\ u_{ij}^H &= \begin{cases} u_{ij} & \forall j \in H_i \\ 0 & \text{otherwise,} \end{cases} \\ u_{ij}^M &= \begin{cases} u_{ij} & \forall j \in M_i \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{6}$$

Then differentiating $C(\mathbf{A})$ with respect to the transformation matrix $\mathbf{A}$ gives the following gradient, which we adopt for learning:

$$\begin{aligned} &\frac{dC(\mathbf{A})}{d(\mathbf{A})} \\ &= \sum_{\forall j \in H_i} \frac{p_{ij}}{q_{ij}} \left(q_{ij}\right)' + \sum_{\forall j \in M_i} \frac{p_{ij}}{q_{ij}} \left(q_{ij}\right)' \\ &= 2\mathbf{A} \left[\sum_{\forall j \in H_i} p_{ij} \frac{\left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T}{1 + \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_i - \mathbf{x}_j\right)} \right] \\ &\quad - 2\mathbf{A} \left[\sum_{\forall j \in H_i} p_{ij} \left(\sum_{t \in H_m} \left(1 + \left(\mathbf{x}_m - \mathbf{x}_t\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_m - \mathbf{x}_t\right)\right)^{-2} \right. \right. \\ &\qquad \left. \times \left(\mathbf{x}_m - \mathbf{x}_t\right)\left(\mathbf{x}_m - \mathbf{x}_t\right)^T \right) \\ &\qquad \times \left(\sum_{t \in H_m} \left(1 + \left(\mathbf{x}_m - \mathbf{x}_t\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_m - \mathbf{x}_t\right)\right)^{-1} \right)^{-1} \Bigg] \\ &\quad + 2\mathbf{A} \left[\sum_{\forall j \in M_i} p_{ij} \frac{\left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T}{1 + \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_i - \mathbf{x}_j\right)} \right] \\ &\quad - 2\mathbf{A} \left[\sum_{\forall j \in M_i} p_{ij} \left(\sum_{t \in M_m} \left(1 + \left(\mathbf{x}_m - \mathbf{x}_t\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_m - \mathbf{x}_t\right)\right)^{-2} \right. \right. \\ &\qquad \left. \times \left(\mathbf{x}_m - \mathbf{x}_t\right)\left(\mathbf{x}_m - \mathbf{x}_t\right)^T \right) \\ &\qquad \times \left(\sum_{t \in M_m} \left(1 + \left(\mathbf{x}_m - \mathbf{x}_t\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_m - \mathbf{x}_t\right)\right)^{-1} \right)^{-1} \Bigg] \\ &= 2\mathbf{A} \left[\sum_{\forall j \in H_i} p_{ij} w_{ij} \left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \right. \\ &\qquad \left. - \sum_{t \in H_m} q_{mt} w_{mt} \left(\mathbf{x}_m - \mathbf{x}_t\right)\left(\mathbf{x}_m - \mathbf{x}_t\right)^T \right] \\ &\quad + 2\mathbf{A} \left[\sum_{\forall j \in M_i} p_{ij} w_{ij} \left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \right. \\ &\qquad \left. - \sum_{t \in M_m} q_{mt} w_{mt} \left(\mathbf{x}_m - \mathbf{x}_t\right)\left(\mathbf{x}_m - \mathbf{x}_t\right)^T \right] \\ &= 2\mathbf{A} \left[\sum_{\forall j \in H_i} u_{ij} \left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \right. \\ &\qquad \left. + \sum_{\forall j \in M_i} u_{ij} \left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \right]. \end{aligned} \tag{7}$$

Let $\mathbf{U}^H$ be the N order matrix with element u_{ij}^H, and let $\mathbf{U}^M$ be the N order matrix with element u_{ij}^M. Note that $\mathbf{U}^H$ and $\mathbf{U}^M$ are symmetric matrices; therefore, $\mathbf{D}^H$ can be defined as a diagonal matrix that each entry is column (or row) sum of $\mathbf{U}^H$ and the same for $\mathbf{D}^M$, that is, $\mathbf{D}_{ii}^H = \sum_j \mathbf{U}_{ij}^H$ and $\mathbf{D}_{ii}^M = \sum_j \mathbf{U}_{ij}^M$. With this definition, the gradient expression (7) can be reduced to

$$\begin{aligned}
\frac{dC(\mathbf{A})}{d(\mathbf{A})} &= 2\mathbf{A}\left\{\sum_{\forall j\in H_i} u_{ij}\left(\mathbf{x}_i-\mathbf{x}_j\right)\left(\mathbf{x}_i-\mathbf{x}_j\right)^T \right.\\
&\qquad \left. +\sum_{\forall j\in M_i} u_{ij}\left(\mathbf{x}_i-\mathbf{x}_j\right)\left(\mathbf{x}_i-\mathbf{x}_j\right)^T\right\}\\
&= 2\mathbf{A}\left\{\left(\sum_{\forall j\in H_i} u_{ij}\mathbf{x}_i\mathbf{x}_i^T+\sum_{\forall j\in H_i} u_{ij}\mathbf{x}_j\mathbf{x}_j^T\right.\right.\\
&\qquad \left.-\sum_{\forall j\in H_i} u_{ij}\mathbf{x}_i\mathbf{x}_j^T-\sum_{\forall j\in H_i} u_{ij}\mathbf{x}_j\mathbf{x}_i^T\right)\\
&\qquad +\left(\sum_{\forall j\in M_i} u_{ij}\mathbf{x}_i\mathbf{x}_i^T+\sum_{\forall j\in M_i} u_{ij}\mathbf{x}_j\mathbf{x}_j^T\right.\\
&\qquad \left.\left.-\sum_{\forall j\in M_i} u_{ij}\mathbf{x}_i\mathbf{x}_j^T-\sum_{\forall j\in M_i} u_{ij}\mathbf{x}_j\mathbf{x}_i^T\right)\right\}\\
&= 4\mathbf{A}\left\{\left(\mathbf{X}\mathbf{D}^H\mathbf{X}^T-\mathbf{X}\mathbf{U}^H\mathbf{X}^T\right)\right.\\
&\qquad \left.+\left(\mathbf{X}\mathbf{D}^M\mathbf{X}^T-\mathbf{X}\mathbf{U}^M\mathbf{X}^T\right)\right\}\\
&= 4\mathbf{A}\left\{\mathbf{X}\left(\mathbf{D}^H-\mathbf{U}^H+\mathbf{D}^M-\mathbf{U}^M\right)\mathbf{X}^T\right\}.
\end{aligned} \tag{8}$$

Once the gradient is calculated, our optimal problem (5) can be solved by an iterative procedure based on the conjugate gradient method. The description of FDSNE algorithm can be given by the following.

Step 1. Collect the sample matrix $\mathbf{X}$ with class labels, and set K-nearest neighborhood parameter K_1, K_2, the variance parameter λ, and the maximum iteration times Mt.

Step 2. Compute the pairwise Euclidian distance for $\mathbf{X}$ and compute the joint probability p_{ij} by utilizing formula (2) and class labels.

Step 3 (set $t = 1 : Mt$). We search for the solution in loop: firstly, compute the joint probability q_{ij} by utilizing formula (4); then, compute gradient $dC(\mathbf{A})/d(\mathbf{A})$ by utilizing formula (8); finally, update $\mathbf{A}^t$ based on $\mathbf{A}^{t-1}$ by conjugate gradient operation.

Step 4. Judge whether $C^t - C^{t-1} < \varepsilon$ (in this paper, we take $\varepsilon = 1e-7$) converges to a stable solution or t reaches the maximum value Mt. If these prerequisites are met, Step 5 is performed; otherwise, we repeat Step 3.

Step 5. Output $\mathbf{A} = \mathbf{A}^t$.

Hereafter, we call the proposed method as fast discriminative stochastic neighbor embedding analysis (FDSNE).

3. Comparison with MSNP and DSNE

MSNP is derived from SNE and t-SNE, and it is a linear method and has nice properties, such as sensitivity to nonlinear manifold structure and convenience for feature extraction. Since the structure of MSNP is closer to that of FDSNE, we briefly compare FDSNE with MSNP and DSNE in this section.

FDSNE, MSNP, and DSNE use different probability distributions to determine the reference points. The difference can be explained in the following aspects.

Firstly, MSNP learns the similarity relationship of the high-dimensional samples by estimating neighborhood distribution based on geodesic distance metric, and the same distribution is required in feature space. Then the linear projection matrix $\mathbf{A}$ is used to discover the underlying structure of data manifold which is nonlinear. Finally, the Kullback-Leibler divergence objective function is used to keep pairwise similarities in feature space. So the probability distribution function of MSNP and its gradient used for learning are respectively given by

$$\begin{aligned}
p_{ij} &= \frac{\exp\left(-D_{ij}^{\text{geo}}/2\right)}{\sum_{k\neq i}\exp\left(-D_{ik}^{\text{geo}}/2\right)},\\
q_{ij} &= \frac{\left[\gamma^2+\left(\mathbf{x}_i-\mathbf{x}_j\right)^T\mathbf{A}^T\mathbf{A}\left(\mathbf{x}_i-\mathbf{x}_j\right)\right]^{-1}}{\sum_{k\neq l}\left[\gamma^2+\left(\mathbf{x}_k-\mathbf{x}_l\right)^T\mathbf{A}^T\mathbf{A}(\mathbf{x}_k-\mathbf{x}_l)\right]^{-1}},\\
\min C(\mathbf{A}) &= \sum_{i,j} p_{ij}\log\frac{p_{ij}}{q_{ij}},
\end{aligned} \tag{9}$$

where D_{ij}^{geo} is the geodesic distance for $\mathbf{x}_i$ and $\mathbf{x}_j$ and γ is the freedom degree parameter of Cauchy distribution.

DSNE selects the joint probability to model the pairwise similarities of input samples with class labels. It also introduces the linear projection matrix $\mathbf{A}$ as MSNP. The cost function is constructed to minimize the intraclass Kullback-Leibler divergence as well as to maximize the interclass KL divergences. Its probability distribution function and gradient are, respectively, given as by

$$p_{ij} = \begin{cases} \dfrac{\exp\left(-\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2/2\lambda^2\right)}{\sum_{c_k=c_l}\exp\left(-\left\|\mathbf{x}_k-\mathbf{x}_l\right\|^2/2\lambda^2\right)} & \text{if } c_i=c_j\\ \dfrac{\exp\left(-\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2/2\lambda^2\right)}{\sum_{c_k\neq c_m}\exp\left(-\left\|\mathbf{x}_k-\mathbf{x}_m\right\|^2/2\lambda^2\right)} & \text{else} \end{cases}$$

$$q_{ij} = \begin{cases} \dfrac{\left(1 + \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_i - \mathbf{x}_j\right)\right)^{-1}}{\sum_{c_k = c_l} \left(1 + \left(\mathbf{x}_k - \mathbf{x}_l\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_k - \mathbf{x}_l\right)\right)^{-1}} & \text{if } c_i = c_j \\ \dfrac{\left(1 + \left(\mathbf{x}_i - \mathbf{x}_j\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_i - \mathbf{x}_j\right)\right)^{-1}}{\sum_{c_k \neq c_m} \left(1 + \left(\mathbf{x}_k - \mathbf{x}_m\right)^T \mathbf{A}^T \mathbf{A} \left(\mathbf{x}_k - \mathbf{x}_m\right)\right)^{-1}} & \text{else,} \end{cases}$$

$$\min C(\mathbf{A}) = \sum_{c_i = c_j} p_{ij} \log \frac{p_{ij}}{q_{ij}} + \sum_{c_i \neq c_k} p_{ik} \log \frac{p_{ik}}{q_{ik}}. \tag{10}$$

Note that on the basis of the DSNE, FDSNE makes full use of class label which not only keeps symmetrical characteristics of the probability distribution matrix but also makes the probability value of interclass data and intraclass data to be 1, and it can effectively overcome large interclass confusion degree in the projected subspace.

Secondly, it is obvious that the selection of reference point in MSNP or DSNE is related to all training samples, while FDSNE only uses the first K-nearest neighbors of each sample from all classes. In other words, we propose an alternative probability distribution function to determine whether $\mathbf{x}_i$ would pick $\mathbf{x}_j$ as its reference point or not. Actually, the computation of gradient during the optimization process mainly determines the computational cost of MSNP and DSNE. So their computational complexity can be written as $O(2rNd+N^2d)$ in each iteration. Similarly, the computational complexity of FDSNE is $O(2rNd + KNd)$ in each iteration, where $K = K_1 + K_2$. It is obvious that $K \ll N$. Therefore, FDSNE is faster than MSNP and DSNE during each iteration.

4. Kernel FDSNE

As a bridge from linear to nonlinear, kernel method emerged in the early beginning of the 20th century and its applications in pattern recognition can be traced back to 1964. In recent years, kernel method has attracted wide attention and numerous researchers have proposed various theories and approaches based on it.

The principle of kernel method is a mapping of the data from the input space R^d to a high-dimensional space F, which we will refer to as the *feature space*, by nonlinear function. Data processing is then performed in the feature space, and this can be expressed solely in terms of inner product in the feature space. Hence, the nonlinear mapping need not be explicitly constructed but can be specified by defining the form of the inner product in terms of a Mercer kernel function κ.

Obviously, FDSNE is a linear feature dimensionality reduction algorithm. So the remainder of this section is devoted to extend FDSNE to a nonlinear scenario using techniques of kernel methods. Let

$$\kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left\langle \varphi\left(\mathbf{x}_i\right), \varphi\left(\mathbf{x}_j\right) \right\rangle \tag{11}$$

which allows us to compute the value of the inner product in F without having to carry out the map.

It should be noted that we use φ_i to denote $\varphi(\mathbf{x}_i)$ for brevity in the following. Next, we express the transformation $\mathbf{A}$ with

$$\mathbf{A} = \left[\sum_{i=1}^{N} b_i^{(1)} \varphi_i, \ldots, \sum_{i=1}^{N} b_i^{(r)} \varphi_i\right]^T. \tag{12}$$

We define $\mathbf{B} = [b^{(1)}, \ldots, b^{(r)}]^T$ and $\mathbf{\Phi} = [\varphi_1, \ldots, \varphi_N]^T$, and then $\mathbf{A} = \mathbf{B}\mathbf{\Phi}$. Based on above definition, the Euclidian distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ in the F space is

$$\begin{aligned} d_{ij}^F(\mathbf{A}) &= \left\|\mathbf{A}\left(\varphi_i - \varphi_j\right)\right\| = \left\|\mathbf{B}\mathbf{\Phi}\left(\varphi_i - \varphi_j\right)\right\| \\ &= \left\|\mathbf{B}\left(K_i - K_j\right)\right\| = \sqrt{\left(K_i - K_j\right)^T \mathbf{B}^T \mathbf{B} \left(K_i - K_j\right)}, \end{aligned} \tag{13}$$

where $K_i = [\kappa(\mathbf{x}_1, \mathbf{x}_i), \ldots, \kappa(\mathbf{x}_N, \mathbf{x}_i)]^T$ is a column vector. It is clear that the distance in the kernel embedding space is related to the kernel function and the matrix $\mathbf{B}$.

In this section, we propose two methods to construct the objective function. The first strategy makes $\mathbf{B}$ parameterize the objective function. Firstly, we replace $d_{ij}(\mathbf{A})$ with $d_{ij}^F(\mathbf{A})$ in formula (3) so that p_{ij}^1, q_{ij}^1 which are defined to be applied in the high dimensional space F can be written as

$$p_{ij}^1 = \begin{cases} \dfrac{\exp\left(-\left(K_{ii} + K_{jj} - 2K_{ij}\right)/2\lambda^2\right)}{\sum_{t \in H_m} \exp\left(-\left(K_{mm} + K_{tt} - 2K_{mt}\right)/2\lambda^2\right)} & \forall j \in H_i \\ \dfrac{\exp\left(-\left(K_{ii} + K_{jj} - 2K_{ij}\right)/2\lambda^2\right)}{\sum_{t \in M_m} \exp\left(-\left(K_{mm} + K_{tt} - 2K_{mt}\right)/2\lambda^2\right)} & \forall j \in M_i \\ 0 & \text{otherwise,} \end{cases}$$

$$q_{ij}^1 = \begin{cases} \dfrac{\left(1 + \left(K_i - K_j\right)^T \mathbf{B}^T \mathbf{B} \left(K_i - K_j\right)\right)^{-1}}{\sum_{t \in H_m} \left(1 + \left(K_m - K_t\right)^T \mathbf{B}^T \mathbf{B} \left(K_m - K_t\right)\right)^{-1}} & \forall j \in H_i \\ \dfrac{\left(1 + \left(K_i - K_j\right)^T \mathbf{B}^T \mathbf{B} \left(K_i - K_j\right)\right)^{-1}}{\sum_{t \in M_m} \left(1 + \left(K_m - K_t\right)^T \mathbf{B}^T \mathbf{B} \left(K_m - K_t\right)\right)^{-1}} & \forall j \in M_i \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Then, we denote $C(\mathbf{B})$ by modifying $C(\mathbf{A})$ via substituting $\mathbf{A}$ with $\mathbf{B}$ into the regularization term of formula (5). Finally,

Figure 1: Sample images from COIL-20 dataset.

Figure 2: Samples of the cropped images from USPS dataset.

by the same argument as formula (7), we give the following gradient:

$$\begin{aligned}\frac{dC(\mathbf{B})}{d(\mathbf{B})} &= \sum_{\forall j\in M_i}\frac{p_{ij}^1}{q_{ij}^1}\left(q_{ij}^1\right)' + \sum_{\forall j\in H_i}\frac{p_{ij}^1}{q_{ij}^1}\left(q_{ij}^1\right)' \\ &= 2\mathbf{B}\left[\sum_{\forall j\in H_i}u_{ij}^1\left(K_i-K_j\right)\left(K_i-K_j\right)^T\right. \\ &\quad \left.+\sum_{\forall j\in M_i}u_{ij}^1\left(K_i-K_j\right)\left(K_i-K_j\right)^T\right].\end{aligned} \tag{15}$$

In order to make formula (15) easy to be comprehended, w_{ij}^1, u_{ij}^1, u_{ij}^{1H}, and u_{ij}^{1M} are given by

$$\begin{aligned} w_{ij}^1 &= \left[1+\left(K_i-K_j\right)^T\mathbf{B}^T\mathbf{B}\left(K_i-K_j\right)\right]^{-1}, \\ u_{ij}^1 &= \left(p_{ij}-q_{ij}\right)w_{ij}^1, \\ u_{ij}^{1H} &= \begin{cases} u_{ij}^1 & \forall j\in H_i \\ 0 & \text{otherwise,}\end{cases} \\ u_{ij}^{1M} &= \begin{cases} u_{ij}^1 & \forall j\in M_i \\ 0 & \text{otherwise.}\end{cases}\end{aligned} \tag{16}$$

Meanwhile, the gradient expression (15) can be reduced to

$$\begin{aligned}\frac{dC(\mathbf{B})}{d(\mathbf{B})} &= 2\mathbf{B}\left\{\sum_{\forall j\in H_i}u_{ij}^1\left(K_i-K_j\right)\left(K_i-K_j\right)^T\right. \\ &\quad \left.+\sum_{\forall j\in M_i}u_{ij}^1\left(K_i-K_j\right)\left(K_i-K_j\right)^T\right\}\end{aligned}$$

Figure 3: Sample face images from ORL dataset.

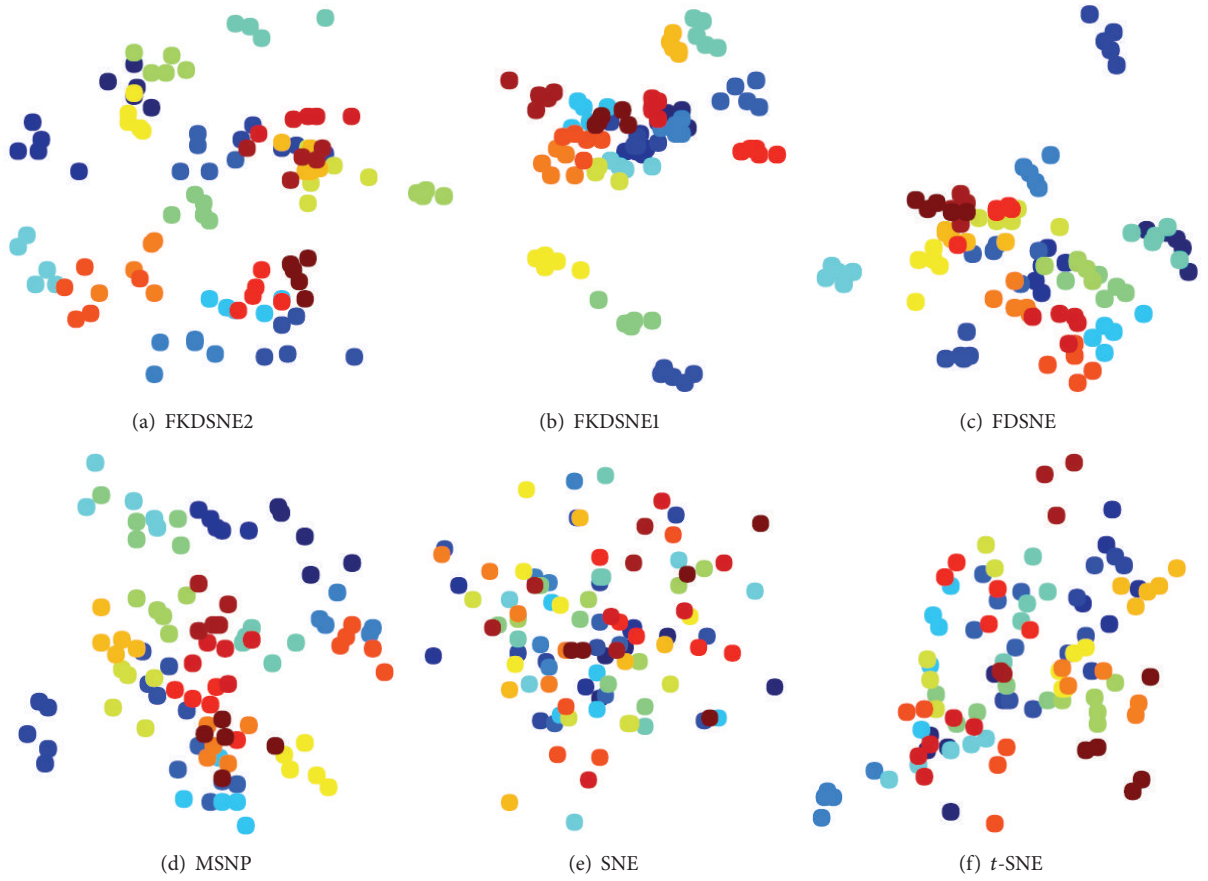
$$\begin{aligned} &= 4\mathbf{B}\left\{\left(\mathbf{K}\mathbf{D}^{1H}\mathbf{K}^T-\mathbf{K}\mathbf{U}^{1H}\mathbf{K}^T\right)\right. \\ &\quad \left.+\left(\mathbf{K}\mathbf{D}^{1M}\mathbf{K}^T-\mathbf{K}\mathbf{U}^{1M}\mathbf{K}^T\right)\right\} \\ &= 4\mathbf{B}\left\{\mathbf{K}\left(\mathbf{D}^{1H}-\mathbf{U}^{1H}+\mathbf{D}^{1M}-\mathbf{U}^{1M}\right)\mathbf{K}^T\right\},\end{aligned} \tag{17}$$

where $\mathbf{U}^{1H}$ is the N order matrix with element u_{ij}^{1H}, and $\mathbf{U}^M$ is the N order matrix with element u_{ij}^{1M}. Note that $\mathbf{U}^{1H}$ and $\mathbf{U}^{1M}$ are symmetric matrices; therefore, $\mathbf{D}^{1H}$ can be defined as a diagonal matrix that each entry is column (or row) sum of $\mathbf{U}^{1H}$ and the same for $\mathbf{D}^{1M}$, that is, $\mathbf{D}_{ii}^{1H} = \sum_j \mathbf{U}_{ij}^{1H}$ and $\mathbf{D}_{ii}^{1M} = \sum_j \mathbf{U}_{ij}^{1M}$.

For convenience, we name this kernel method as FKD-SNE1.

Another strategy is that we let $C^F(\mathbf{A})$ be the objective function in the embedding space F. So its gradient can be written as

$$\begin{aligned}&\frac{dC^F(\mathbf{A})}{d(\mathbf{A})} \\ &= \sum_{\forall j\in M_i}\frac{p_{ij}^1}{q_{ij}^1}\left(q_{ij}^1\right)' + \sum_{\forall j\in H_i}\frac{p_{ij}^1}{q_{ij}^1}\left(q_{ij}^1\right)' \\ &= 2\left[\sum_{\forall j\in H_i}p_{ij}^1\frac{\mathbf{B}\left(K_i-K_j\right)\left(\boldsymbol{\varphi}_i-\boldsymbol{\varphi}_j\right)^T}{\left(1+\left(K_i-K_j\right)^T\mathbf{B}^T\mathbf{B}\left(K_i-K_j\right)\right)}\right] \\ &\quad -2\left[\sum_{\forall j\in H_i}p_{ij}^1\left(\sum_{t\in H_m}\left(1+\left(K_m-K_t\right)^T\mathbf{B}^T\mathbf{B}\left(K_m-K_t\right)\right)^{-2}\right.\right. \\ &\qquad \left.\times\mathbf{B}\left(K_m-K_t\right)\left(\boldsymbol{\varphi}_m-\boldsymbol{\varphi}_t\right)^T\right) \\ &\qquad \left.\times\left(\sum_{t\in H_m}\left(1+\left(K_m-K_t\right)^T\mathbf{B}^T\mathbf{B}\left(K_m-K_t\right)\right)^{-1}\right)^{-1}\right] \\ &\quad +2\left[\sum_{\forall j\in M_i}p_{ij}^1\frac{\mathbf{B}\left(K_i-K_j\right)\left(\boldsymbol{\varphi}_i-\boldsymbol{\varphi}_j\right)^T}{1+\left(K_i-K_j\right)^T\mathbf{B}^T\mathbf{B}\left(K_i-K_j\right)}\right] \\ &\quad -2\left[\sum_{\forall j\in M_i}p_{ij}^1\left(\sum_{t\in M_m}\left(1+\left(K_m-K_t\right)^T\mathbf{B}^T\mathbf{B}\left(K_m-K_t\right)\right)^{-2}\right.\right. \\ &\qquad \left.\left.\times\mathbf{B}\left(K_m-K_t\right)\left(\boldsymbol{\varphi}_m-\boldsymbol{\varphi}_t\right)^T\right)\right.\end{aligned}$$



(a) FKDSNE2 (b) FKDSNE1 (c) FDSNE

(d) MSNP (e) SNE (f) t-SNE

Figure 4: Visualization of 100 images from COIL-20 images dataset.

$$\times \left(\sum_{t \in M_m} \left(1 + (K_m - K_t)^T \mathbf{B}^T \mathbf{B} (K_m - K_t)\right)^{-1} \right)^{-1} \Bigg]$$

$$= 2 \left[\sum_{\forall j \in H_i} p_{ij}^1 w_{ij}^1 \mathbf{B} Q_{ij}^{(K_i - K_j)} - \sum_{t \in H_m} q_{mt}^1 w_{mt}^1 \mathbf{B} Q_{mt}^{(K_m - K_t)} \right] \mathbf{\Phi}$$

$$+ 2 \left[\sum_{\forall j \in M_i} p_{ij}^1 w_{ij}^1 \mathbf{B} Q_{ij}^{(K_i - K_j)} - \sum_{t \in M_m} q_{mt}^1 w_{mt}^1 \mathbf{B} Q_{mt}^{(K_m - K_t)} \right] \mathbf{\Phi}$$

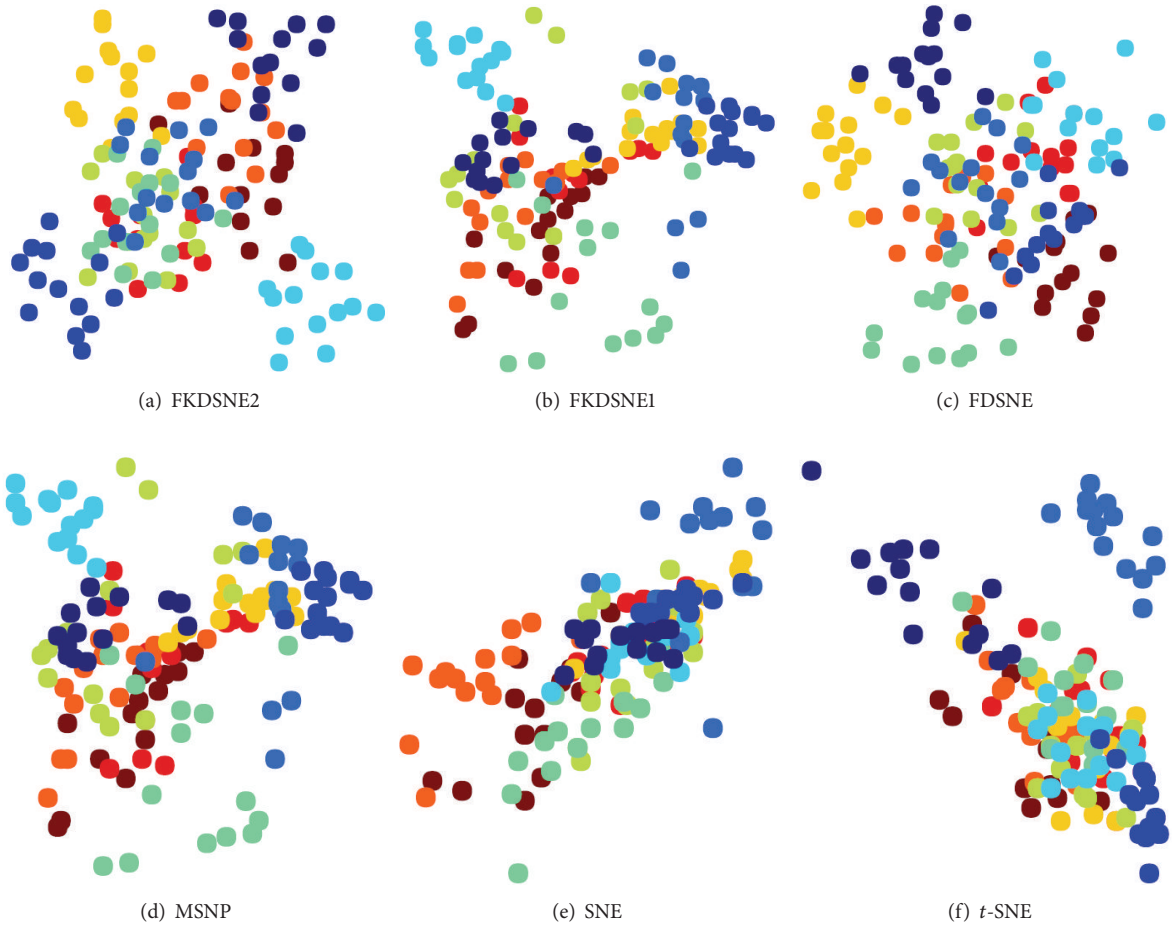
$$= 2 \left[\sum_{\forall j \in H_i} u_{ij}^1 \mathbf{B} Q_{ij}^{(K_i - K_j)} + \sum_{\forall j \in M_i} u_{ij}^1 \mathbf{B} Q_{ij}^{(K_i - K_j)} \right] \mathbf{\Phi} \tag{18}$$

in this form, $Q_{ij}^{(K_i - K_j)}$ can be regard as the $N \times N$ matrix with vector $K_i - K_j$ in the ith column, and vector $K_j - K_i$ in the jth column and the other columns are all zeros.

This method is termed as FKDSNE2. Note that $\mathbf{\Phi}$ is a constant matrix. Furthermore, the observations of formula (18) make us know that updating the matrix $\mathbf{A}$ in the optimization only means updating the matrix $\mathbf{B}$. Additionally, $\mathbf{\Phi}$ does not need to be computed explicitly. Therefore, we do not need to explicitly perform the nonlinear map $\varphi(\mathbf{x})$ to minimize the objective function $C^F(\mathbf{A})$. The computational complexity of FKDSNE1 and FKDSNE2, is respectively, $O(2rN^2 + rNK)$ and $O(2rKN + rN^2)$ in each iteration. Hence, it is obvious that FKDSNE2 is faster than FKDSNE1 during each iteration.

5. Experiments

In this section, we evaluate the performance of our FDSNE, FKDSNE1, and FKDSNE2 methods for feature extraction. Three sets of experiments are carried out on Columbia Object Image Library (COIL-20) (http://www1.cs.columbia.edu/CAVE/software/softlib/coil-20.php), US Postal Service (USPS) (http://www.cs.nyu.edu/~roweis/data.html), and ORL (http://www.cam-orl.co.uk) face datasets to demonstrate their good behavior on visualization, accuracy, and elapsed time. In the first set of experiments, we focus on the visualization of the proposed methods which are compared with that of the relevant algorithms, including SNE [17], t-SNE [18], and MSNP [21]. In the second set of experiments, we apply our methods to recognition task to verify their feature extraction capability and compare them with MSNP and DSNE [22]. Moreover, the elapsed time of FDSNE, FKDSNE1, FKDSNE2, and DSNE is compared in the third set of experiments. In particular, the Gaussian RBF kernel $\kappa(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / 2\sigma^2)$ is chosen as the kernel function of FKDSNE1 and FKDSNE2, where σ is set as the variance of the training sample set of $\mathbf{X}$.



(a) FKDSNE2 (b) FKDSNE1 (c) FDSNE

(d) MSNP (e) SNE (f) t-SNE

Figure 5: Visualization of 140 images from USPS handwritten digits dataset.

5.1. COIL-20, USPS, and ORL Datasets. The datasets used in our experiments are summarized as follows.

COIL-20 is a dataset of gray-scale images of 20 objects. The images of each object were taken 5 degrees apart as the object is rotated on a turntable and each object has 72 images. The size of each image is 40×40 pixels. Figure 1 shows sample images from COIL-20 images dataset.

USPS handwritten digit dataset includes 10 digit characters and 1100 samples in total. The original data format is of 16×16 pixels. Figure 2 shows samples of the cropped images from USPS handwritten digits dataset.

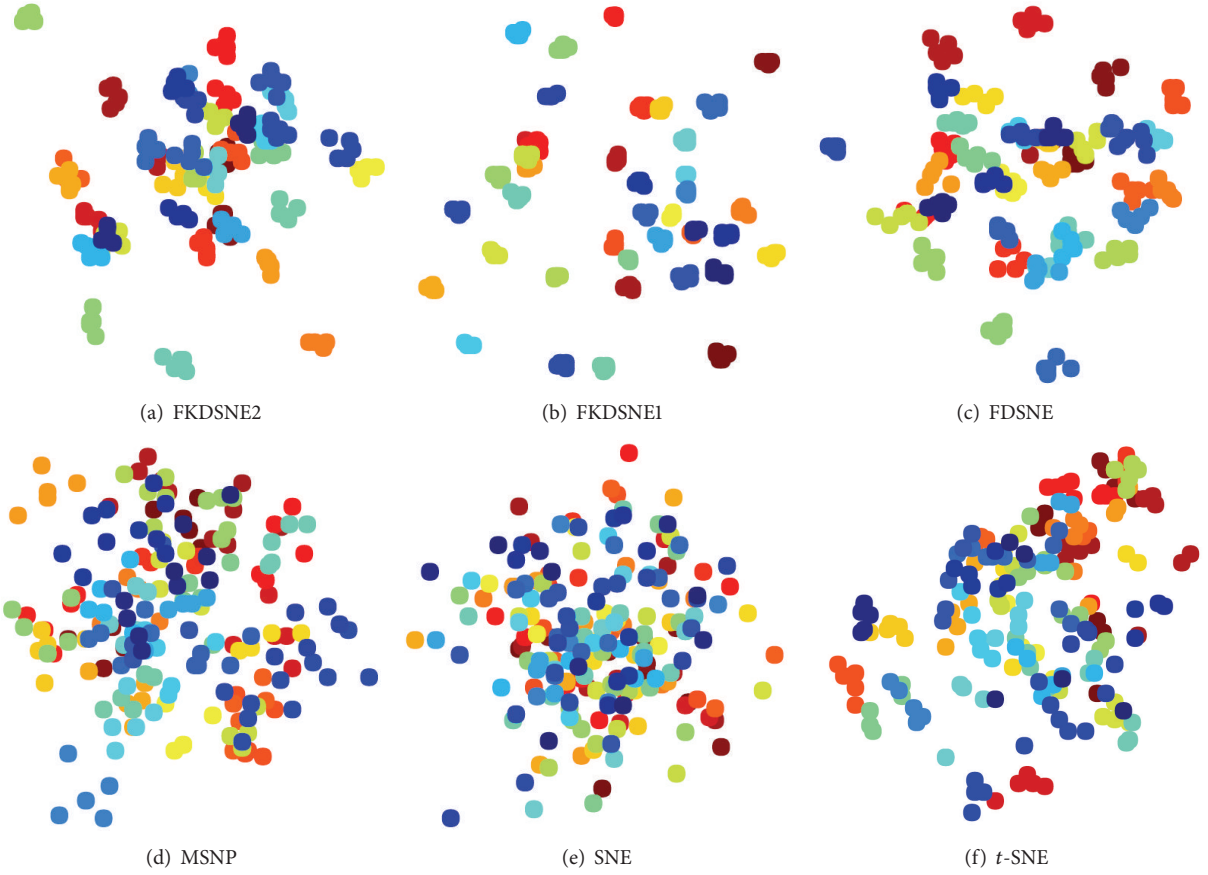
ORL consists of gray images of faces from 40 distinct subjects, with 10 pictures for each subject. For every subject, the images were taken with varied lighting condition and different facial expressions. The original size of each image is 112×92 pixels, with 256 gray levels per pixel. Figure 3 illustrates a sample subject of ORL dataset.

5.2. Visualization Using FDSNE, FKDSNE1, and FKDSNE2. We apply FDSNE, FKDSNE1, and FKDSNE2 to visualization task to evaluate their capability of classification performance. The experiments are carried out, respectively, on COIL-20, USPS, and ORL datasets. For the sake of computational efficiency as well as noise filtering, we first adjust the size of each image to 32×32 pixels on ORL, and then we select five samples from each class on COIL-20, fourteen samples from each class on USPS, and five samples from each class on ORL.

The experimental procedure is to extract a 20-dimensional feature for each image by FDSNE, FKDSNE1, and FKDSNE2, respectively. Then to evaluate the quality of features through visual presentation of the first two-dimensional feature.

FDSNE, FKDSNE1, and FKDSNE2 are compared with three well known visualization methods for detecting classification performance: (1) SNE, (2) t-SNE, and (3) MSPN. The parameters are set as follows: the K-nearest neighborhood parameter of FDSNE, FKDSNE1, and FKDSNE2 methods is $K_1 = h - 1$ (let h denote the number of training samples in each class), $K_2 = 40$; for SNE and t-SNE, the perplexity parameter is perp $= 20$ and the iteration number is $Mt = 1000$; for MSNP, the degree freedom of Cauchy distribution is $\gamma = 4$ and the iteration number is 1000 as well.

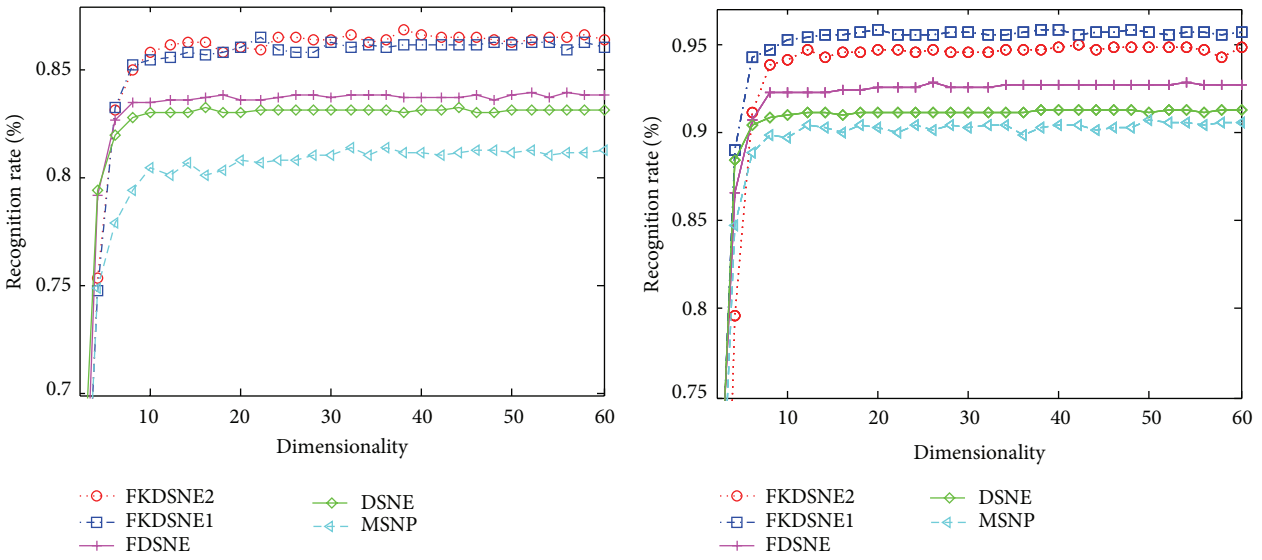
Figures 4, 5, and 6 show the visual presentation results of FDSNE, FKDSNE1, FKDSNE2, SNE, t-SNE, and MSNP, respectively, on COIL-20, USPS, and ORL datasets. The visual presentation is represented as a scatterplot in which a different color determines different class information. The figures reveal that the three nearest-neighbor-based methods,



(a) FKDSNE2 (b) FKDSNE1 (c) FDSNE

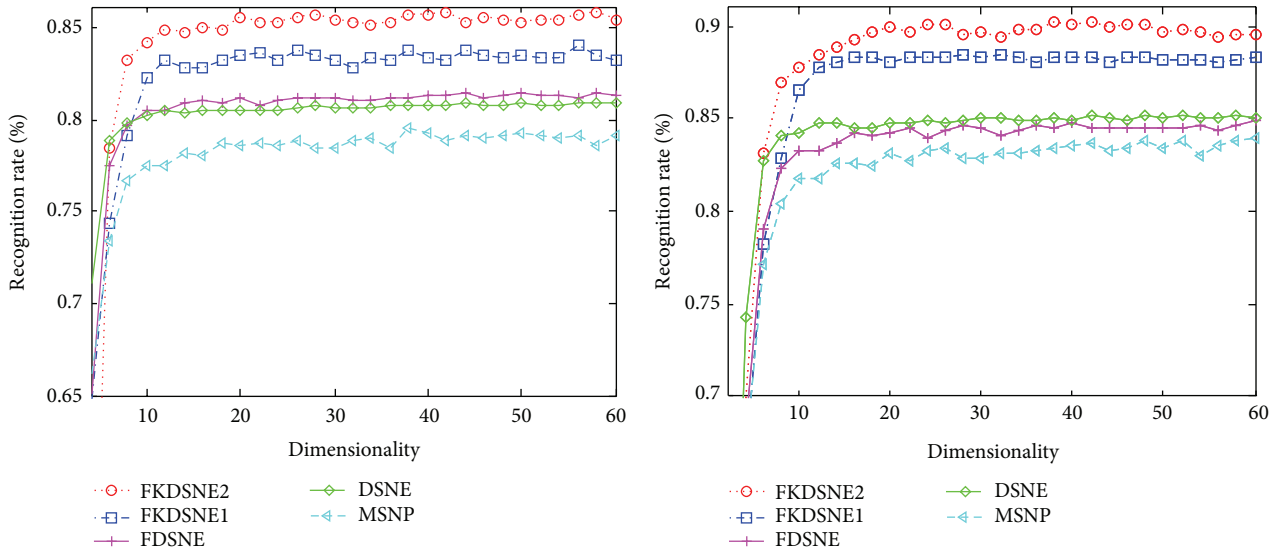
(d) MSNP (e) SNE (f) t-SNE

Figure 6: Visualization of 200 face images from ORL faces dataset.



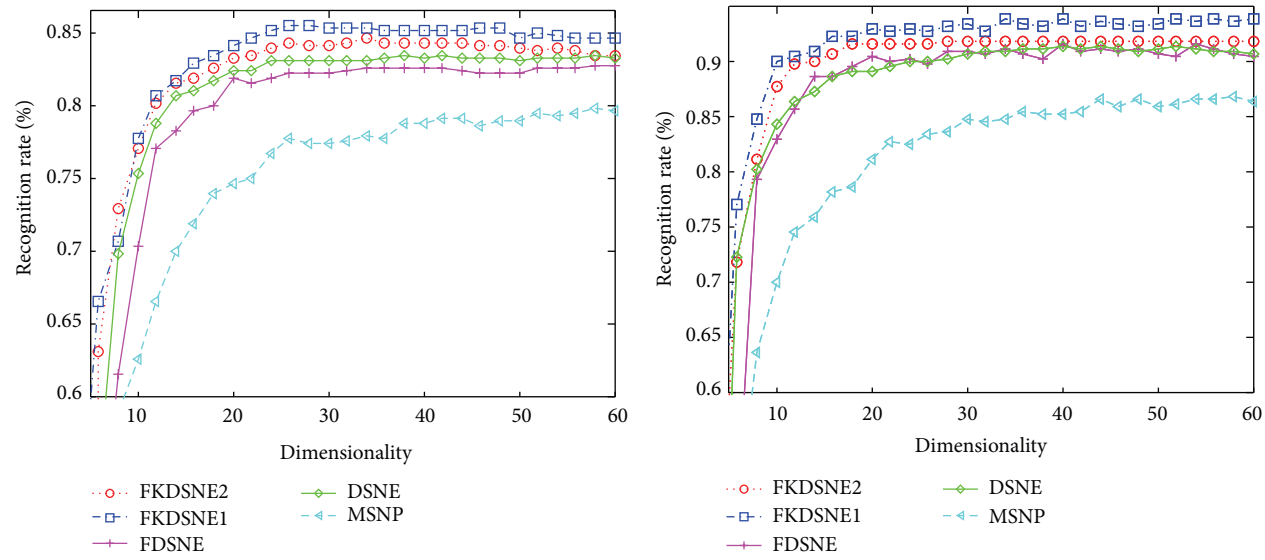
(a) $h = 5$ (b) $h = 10$

Figure 7: Recognition rate (%) versus subspace dimension on COIL-20.



(a) $h = 14$ (b) $h = 25$

Figure 8: Recognition rate (%) versus subspace dimension on USPS.



(a) $h = 3$ (b) $h = 5$

Figure 9: Recognition rate (%) versus subspace dimension on ORL.

that is, FDSNE, FKDSNE1, and FKDSNE2, give considerably better classification result than SNE, t-SNE, and MSNP on all datasets, for the separation between classes is quite obvious. In particular, SNE and t-SNE not only get less separation for the interclass data but also produce larger intraclass scatter. For MSNP, it has smaller intraclass scatter, but there exists an overlapping phenomenon among classes. With regard to FDSNE, FKDSNE1, and FKDSNE2, we can find from the figures that FKDSNE1 shows the best classification performance among all the algorithms on ORL face dataset, while not on the other two datasets COIL-20 and USPS; thereinto, the classification performance of FKDSNE1 is inferior to FDSNE on COIL-20 while on USPS it is inferior to FKDSNE2. In addition, the clustering qualities and separation degree of FKDSNE1 and FKDSNE2 are obviously better than that of FDSNE.

5.3. Recognition Using FDSNE, FKDSNE1, and FKDSNE2. In this subsection, we apply FDSNE, FKDSNE1, and FKDSNE2 to recognition task to verify their feature extraction capability. Nonlinear dimensional reduction algorithms such as SNE and t-SNE lack explicit projection matrix for the out-of-sample data, which means they are not suitable for recognition. So we compare the proposed methods with DSNE and

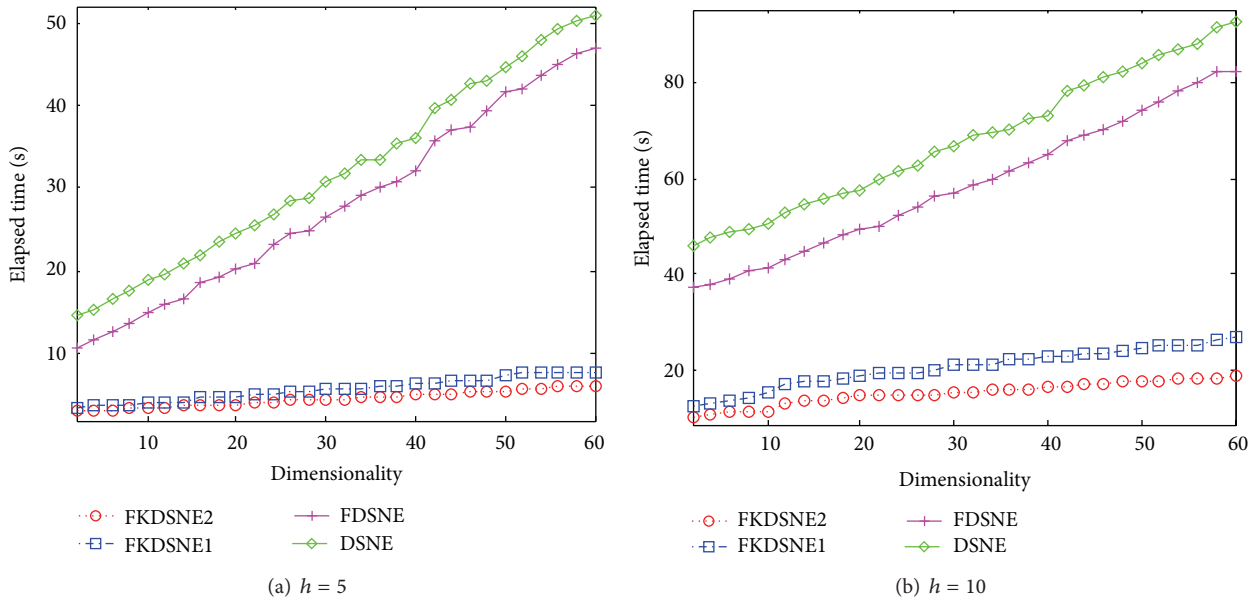


Figure 10: Elapsed time (seconds) versus subspace dimension on COIL-20.

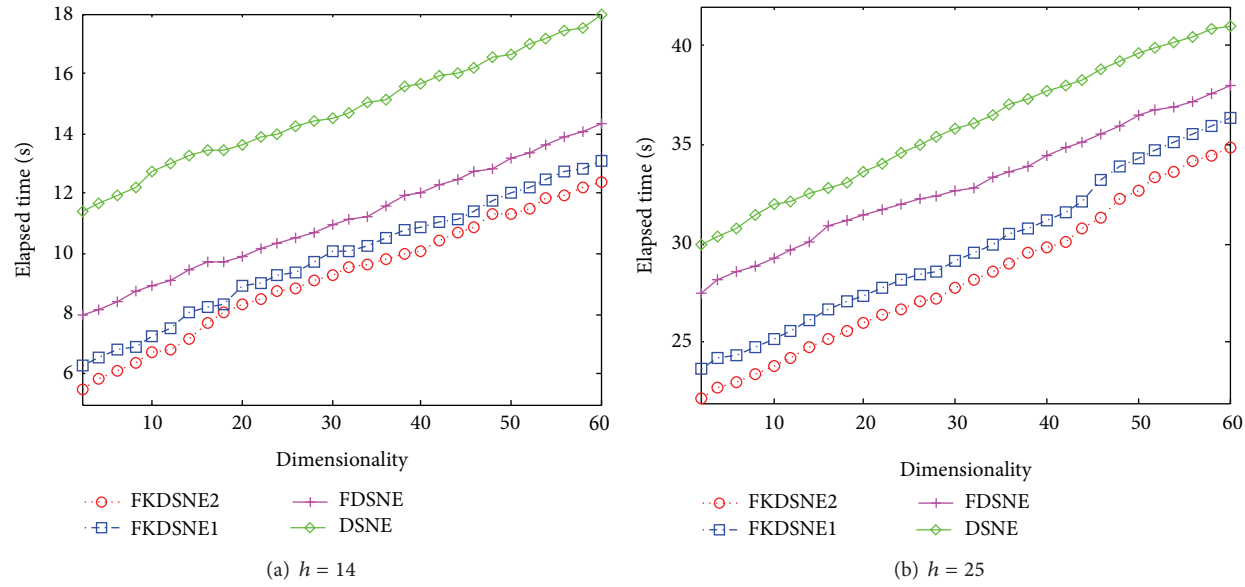
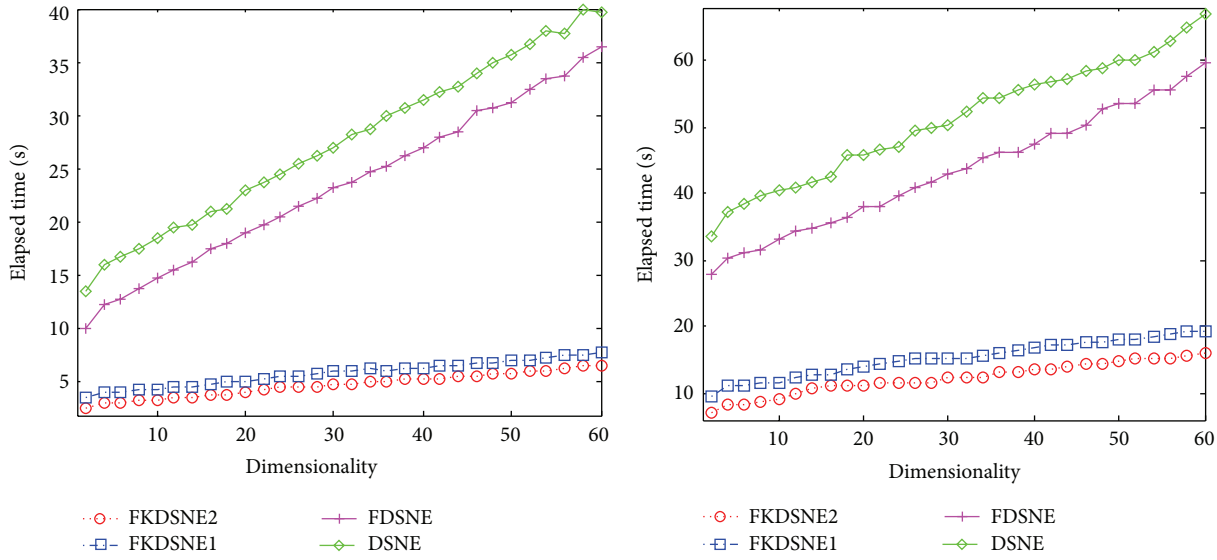


Figure 11: Elapsed time (seconds) versus subspace dimension on USPS.

MSNP, both of them are linear methods and were proved to be better than existing feature extraction algorithms such as SNE, t-SNE, LLTSA, LPP, and so on in [21, 22]. The procedure of recognition is described as follows: firstly, divide dataset into training sample set $\mathbf{X}_{\text{train}}$ and testing sample set $\mathbf{X}_{\text{test}}$ randomly; secondly, the training process for the optimal matrix $\mathbf{A}$ or $\mathbf{B}$ is taken for FDSNE, FKDSNE1 and FKDSNE2; thirdly, feature extraction is accomplished for all samples using $\mathbf{A}$ or $\mathbf{B}$; finally, a testing image is identified by a nearest neighbor classifier. The parameters are set as follows: the K-nearest neighborhood parameter K_1, K_2 in FDSNE, FKDSNE1, and FKDSNE2 is $K_1 = h - 1$, $K_2 = 40$; for DSNE, the perplexity parameter is $\lambda = 0.1$ and the iteration number is $Mt = 1000$; for MSNP, the freedom degree γ of Cauchy distribution in MSNP is determined by cross validation and the iteration number is 1000 as well.

Figure 7 demonstrates the effectiveness of different subspace dimensions for COIL-20 ((a): $h = 5$, (b): $h = 10$). Figure 8 is the result of the experiment in USPS ((a): $h = 14$, (b): $h = 25$), and Figure 9 shows the recognition rate versus subspace dimension on ORL ((a): $h = 3$, (b): $h = 5$). The maximal recognition rate of each method and the corresponding dimension are given in Table 1, where the number in bold stands for the highest recognition rate. From Table 1,



(a) $h = 3$ (b) $h = 5$

Figure 12: Elapsed time (seconds) versus subspace dimension on ORL.

Table 1: The maximal recognition rates (%) versus the subspace dimension.

	COIL-20 $h = 5$	COIL-20 $h = 10$	USPS $h = 14$	USPS $h = 25$	ORL $h = 3$	ORL $h = 5$
MSNP	0.8149 (32)	0.9063 (50)	0.7958 (38)	0.8395 (58)	0.7989 (59)	0.8690 (58)
DSNE	0.8325 (36)	0.9130 (54)	0.8093 (50)	0.8522 (42)	0.8357 (42)	0.9150 (39)
FDSNE	0.8396 (52)	0.9277 (54)	0.8150 (58)	0.8489 (59)	0.8279 (58)	0.9160 (39)
FKDSNE1	0.8651 (22)	**0.9575 (20)**	0.8409 (26)	0.8848 (26)	**0.8550 (26)**	**0.9405 (24)**
FKDSNE2	**0.8689 (28)**	0.9491 (22)	**0.8585 (22)**	**0.9021 (28)**	0.8470 (24)	0.9193 (20)

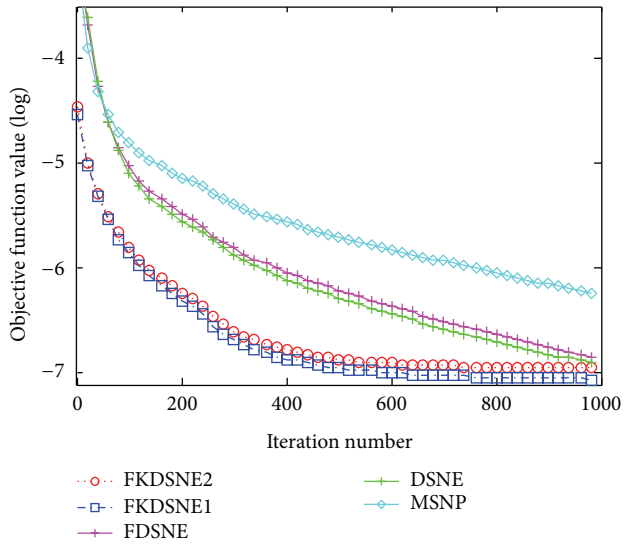


Figure 13: Objective function value (log) versus iterative number on ORL dataset.

we can find that FKDSNE1 and FKDSNE2 outperform MSNP, DSNE, and FDSNE on COIL-20, USPS, and ORL. As can be seen, FKDSNE1 and FKDSNE2 enhance the maximal recognition rate for at least 2% compared with other three methods. Besides, FKDSNE1 and FKDSNE2 achieve considerable recognition accuracy when feature dimension is 20 on the three datasets. It indicates that FKDSNE1 and FKDSNE2 grasp the key character of face images relative to identification with a few features. Though the maximal recognition rate of DSNE and FDSNE is closer to that of FKDSNE1 and FKDSNE2 on ORL dataset, the corresponding dimension of FKDSNE1 and FKDSNE2 is 20 while that of DSNE and FDSNE exceeds 30. From the essence of dimensional reduction, this result demonstrates that FDSNE and DSNE are inferior to FKDSNE1 and FKDSNE2.

5.4. Analysis of Elapsed Time. In this subsection, we further compare the computational efficiency of DSNE, FKDSNE, FKDSNE1, and FKDSNE2. The algorithm MSPN is not compared since its recognition rate is obviously worse than other algorithms. The parameters of the experiment are the same to Section 5.3. Figures 10, 11, and 12, respectively, show the elapsed time of four algorithms under different subspace dimensions on the three datasets. It can be observed from the figures that FKDSNE2 has the lowest computational cost among the four algorithms while DSNE is much inferior to other nearest-neighbor-based algorithms on all datasets. Particularly, on the COIL-20 dataset, the elapsed time of FKDSNE2 is more than 2 times faster than DSNE. As for DSNE

and FDSNE, the former is obviously slower than the latter. Besides, for the two kernel methods, FKDSNE2 is notably faster than FKDSNE1, which confirms our discussion in Section 4.

Furthermore, kernel-based algorithms FKDSNE1 and FKDSNE2 can effectively indicate the linear structure on high-dimensional space. Their objective function can achieve better values on desirable dimensions. For instance, Figure 13 illustrates the objective function value of MSNP, DSNE, FKDSNE, FKDSNE1, and FKDSNE2 versus iterative number on ORL dataset. It can be found that FKDSNE2 and FKDSNE1 is close to the convergence value $1e-7$ while FDSNE and DSNE only achieve $1e-6$ and MSNP achieves $1e-5.4$ when the iterative number is 400. It means that FKDSNE1 and FKDSNE2 can get the more precise objective function value with less iterative number compared with DSNE and FDSNE; that is to say that, FKDSNE1 and FKDSNE2 can achieve the same value by using forty percent of the elapsed time of DSNE and FDSNE.

6. Conclusion

On the basis of DSNE, we present a method called fast discriminative stochastic neighbor embedding analysis (FDSNE) which chooses the reference points in K-nearest neighbors of the target sample from the same class and the different classes instead of the total training samples and thus has much lower computational complexity than that of DSNE. Furthermore, since FDSNE is a linear feature dimensionality reduction algorithm, we extend FDSNE to a nonlinear scenario using techniques of kernel trick and present two kernel-based methods: FKDSNE1 and FKDSNE2. Experimental results on COIL-20, USPS, and ORL datasets show the superior performance of the proposed methods. Our future work might include further empirical studies on the learning speed and robustness of FDSNE by using more extensive, especially large-scale, experiments. It also remains important to investigate acceleration techniques in both initialization and long-run stages of the learning.

Acknowledgment

This project was partially supported by Zhejiang Provincial Natural Science Foundation of China (nos. LQ12F03011 and LQ12F03005).

References

[1] E. Cherchi and C. A. Guevara, "A Monte Carlo experiment to analyze the curse of dimensionality in estimating random coefficients models with a full variance-covariance matrix," *Transportation Research B*, vol. 46, no. 2, pp. 321–332, 2012.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[3] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[5] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

[6] W. Bian and D. Tao, "Max-min distance analysis by using sequential SDP relaxation for dimension reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 1037–1050, 2011.

[7] Z. Teng, J. He et al., "Critical mechanical conditions around neovessels in carotid atherosclerotic plaque may promote intraplaque hemorrhage," *Atherosclerosis*, vol. 223, no. 2, pp. 321–326, 2012.

[8] Z. Teng, A. J. Degnan, U. Sadat et al., "Characterization of healing following atherosclerotic carotid plaque rupture in acutely symptomatic patients: an exploratory study using in vivo cardiovascular magnetic resonance," *Journal of Cardiovascular Magnetic Resonance*, vol. 13, article 64, 2011.

[9] C. E. Hann, I. Singh-Levett, B. L. Deam, J. B. Mander, and J. G. Chase, "Real-time system identification of a nonlinear four-story steel frame structure-application to structural health monitoring," *IEEE Sensors Journal*, vol. 9, no. 11, pp. 1339–1346, 2009.

[10] A. Segui, J. P. Lebaron, and R. Leverge, "Biomedical engineering approach of pharmacokinetic problems: computer-aided design in pharmacokinetics and bioprocessing," *IEE Proceedings D*, vol. 133, no. 5, pp. 217–225, 1986.

[11] F. Wu, Y. Zhong, and Q. Y. Wu, "Online classification framework for data stream based on incremental kernel principal component analysis," *Acta Automatica Sinica*, vol. 36, no. 4, pp. 534–542, 2010.

[12] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[13] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[14] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[15] H. Li, H. Jiang, R. Barrio, X. Liao, L. Cheng, and F. Su, "Incremental manifold learning by spectral embedding methods," *Pattern Recognition Letters*, vol. 32, no. 10, pp. 1447–1455, 2011.

[16] P. Zhang, H. Qiao, and B. Zhang, "An improved local tangent space alignment method for manifold learning," *Pattern Recognition Letters*, vol. 32, no. 2, pp. 181–189, 2011.

[17] G. Hinton and S. Roweis, "Stochastic neighbor embedding," *Advances in Neural Information Processing Systems*, vol. 15, pp. 833–840, 2002.

[18] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[19] J. A. Cook, I. Sutskever, A. Mnih, and G. E. Hinton, "Visualizing similarity data with a mixture of maps," in *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics*, vol. 2, pp. 67–74, 2007.

[20] Z. R. Yang, I. King, Z. L. Xu, and E. Oja, "Heavy-tailed symmetric stochastic neighbor embedding," *Advances in Neural Information Processing Systems*, vol. 22, pp. 2169–2177, 2009.

[21] S. Wu, M. Sun, and J. Yang, "Stochastic neighbor projection on manifold for feature extraction," *Neurocomputing*, vol. 74, no. 17, pp. 2780–2789, 2011.

[22] J. W. Zheng, H. Qiu, Y. B. Jiang, and W. L. Wang, "Discriminative stochastic neighbor embedding analysis method," *Computer-Aided Design & Computer Graphics*, vol. 24, no. 11, pp. 1477–1484, 2012.

[23] C. Cattani, R. Badea, S. Chen, and M. Crisan, "Biomedical signal processing and modeling complexity of living systems," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 298634, 2 pages, 2012.

[24] X. Zhang, Y. Zhang, J. Zhang et al., "Unsupervised clustering for logo images using singular values region covariance matrices on Lie groups," *Optical Engineering*, vol. 51, no. 4, Article ID 047005, 8 pages, 2012.

[25] P. J. Moreno, P. Ho, and N. Vasconcelos, "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications," *Advances in Neural Information Processing Systems*, vol. 16, pp. 1385–1393, 2003.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 347238, 6 pages
http://dx.doi.org/10.1155/2013/347238

Research Article

Fractal Analysis of Elastographic Images for Automatic Detection of Diffuse Diseases of Salivary Glands: Preliminary Results

Alexandru Florin Badea,[1] Monica Lupsor Platon,[2] Maria Crisan,[3] Carlo Cattani,[4] Iulia Badea,[5] Gaetano Pierro,[6] Gianpaolo Sannino,[7] and Grigore Baciut[1]

[1] *Department of Cranio-Maxillo-Facial Surgery, University of Medicine and Pharmacy "Iuliu Haţieganu", Cardinal Hossu Street 37, 400 029 Cluj-Napoca, Romania*
[2] *Department of Clinical Imaging, University of Medicine and Pharmacy "Iuliu Haţieganu", Croitorilor Street 19-21, 400 162 Cluj-Napoca, Romania*
[3] *Department of Histology, Pasteur 5-6 University of Medicine and Pharmacy "Iuliu Haţieganu", 400 349 Cluj-Napoca, Romania*
[4] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*
[5] *Department of Dental Prevention, University of Medicine Pharmacy "Iuliu Haţieganu", Victor Babes Street, 400 012 Cluj-Napoca, Romania*
[6] *Department of System Biology, Phd School, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*
[7] *Department of Oral Health, University of Rome Tor Vergata, Viale Oxford, 00100 Rome, Italy*

Correspondence should be addressed to Maria Crisan; mcrisan7@yahoo.com

Received 10 March 2013; Accepted 12 April 2013

Academic Editor: Shengyong Chen

The geometry of some medical images of tissues, obtained by elastography and ultrasonography, is characterized in terms of complexity parameters such as the fractal dimension (FD). It is well known that in any image there are very subtle details that are not easily detectable by the human eye. However, in many cases like medical imaging diagnosis, these details are very important since they might contain some hidden information about the possible existence of certain pathological lesions like tissue degeneration, inflammation, or tumors. Therefore, an automatic method of analysis could be an expedient tool for physicians to give a faultless diagnosis. The fractal analysis is of great importance in relation to a quantitative evaluation of "real-time" elastography, a procedure considered to be operator dependent in the current clinical practice. Mathematical analysis reveals significant discrepancies among normal and pathological image patterns. The main objective of our work is to demonstrate the clinical utility of this procedure on an ultrasound image corresponding to a submandibular diffuse pathology.

1. Introduction

In some recent papers [1–4], the fractal nature of nucleotide distribution in DNA has been investigated in order to classify and compare DNA sequences and to single out some particularities in the nucleotide distribution, sometimes in order to be used as markers for the existence of certain pathologies [5–9]. Almost all these papers are motivated by the hypothesis that changes in the fractal dimension might be taken as markers for the existence of pathologies since it is universally accepted nowadays that bioactivity and the biological systems are based on some fractal nature organization [3, 4, 10–13]. From a mathematical point of view, this could be explained by the fact that the larger the number of interacting individuals, the more complex the corresponding system of interactions is. These hidden rules that lead to this complex fractal topology could be some simple recursive rules, typical of any fractal-like structure, which usually requires a large number of recursions in order to fill the space.

In recent years, many papers [3–6, 9, 14, 15] have investigated the multi-fractality of biological signals such as DNA and the possible influence of the fractal geometry on

the functionality of DNA from a biological-chemical point of view. Almost all these papers concerning the multifractality of biological signals are based on the hypothesis that the functionality and the evolution of tissues/cells/DNA are related to and measured by the evolving fractal geometry (complexity), so that malfunctions and pathologies can be linked with the degeneracy of the geometry during its evolution time [5–7, 16–18].

From a mathematical point of view, a fractal is a geometric object mainly characterized by the noninteger dimension and self-similarity so that a typical pattern repeats itself cyclically at different scales. A more complex definition of a fractal is based on the four properties: self-similarity, fine structure, irregularities, and noninteger dimension [19]. The fractal dimension is a parameter which measures the relationship between the geometric un-smoothness of the object and its underlying metric space. Since it is a noninteger value, it is usually taken as a measure of the unsmoothness, thus being improperly related to the level of complexity or disorder. Fractality has been observed and measured in several fields of specialization in biology, similar to those in pathology and cancer models [20, 21]. However, only recently have been made some attempts to investigate the structural importance of the "fractal nature" of the DNA. It has been observed in some recent papers that the higher FD corresponds to the higher information complexity and thus to the evolution towards a pathological state [3, 4].

In the following, we will analyse the particularities of the fractal dimension focused on the pathological aspects of some tissues, more specific those belonging to a submandibular gland. For the first time, the FD is computed on images obtained by the new technology of elastographic imaging focused on this salivary gland.

2. Materials and Methods

2.1. Material. A 55-year-old woman presented herself in the emergency room of the Maxilo-Facial Surgery Department for acute pain and enlargement of the left submandibular gland and was selected for ultrasound evaluation. The ultrasound examination was performed using the ACUSON S2000 (Siemens) ultrasound equipment, where the ARFI (acoustic radiation force impulse) and real-time elastography technique were implemented. The ACUSON S2000 is a powerful, non-invasive, ultrasound based device, which gives very accurate B mode and Doppler images of tissues. It has been profitably used for the analysis of abdominal, breast, cardiac, obstetrical, and gynaecological imaging and also for small parts such as thyroid and vascular imaging.

The patient was placed laying down and facing up, while the transducer was placed in contact with skin on the area of the right and then the left submandibular gland successively. The shear wave velocity within the right and the left submandibular gland parenchyma was determined for each submandibular gland (in meters/second); colour elastographic images were also acquired. A colour map was used where stiff tissues were coded in blue and soft tissues in red. These images were studied afterwards for fractal analysis.

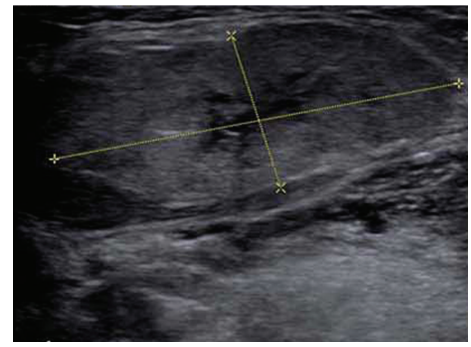


Figure 1: Gray scale ultrasonography of the submandibular gland (right side). The gland is enlarged (total volume around 12 cmc) with well-defined delineation, inhomogeneous structure, hypoechoic area in the center (belongs to the hilum of the gland), and hyperechoic areas under the capsule (belong to the parenchyma).

Figure 1 represents a 2D ultrasound evaluation in a "grey scale" mode, and Figure 2 represents a combination between 2D ultrasonography and "colour flow map" (CFM, or "duplex sonography"). From the first viewing, we can easily detect, by its enlargement, the gland swelling (Figure 1) and the hyper vascular pattern (Figure 2), both of these pieces of information being highly suggestive for the inflammation diagnosis. The combined clinical and ultrasound evaluation is conclusive for an acute inflammation of the submandibular gland. Figures 3 and 5 (obtained on the right salivary swollen gland) and Figures 4 and 6 (obtained on the left side, normal gland) represent elastography in quantitative mode (Figures 3 and 4), color mode (Figures 5 and 6) (ARFI tissue imaging mapping color).

2.2. Methods. Concerning the fractal analysis in this section, we will summarize some definitions already given in [3].

2.3. Parameters for the Analysis of Complexity and Fractal Geometry. As a measure of the complexity and fractal geometry, we will consider only the fractal dimension and regression analysis (Shannon information entropy, lacunarity, and succolarity will be considered in a forthcoming paper).

Let $p_x(n)$ be the probability to find the value x at the position n, the fractal dimension is given by [3, 4, 22]

$$D = \frac{1}{N}\sum_{n=2}^{N}\frac{\log\ p_x(n)}{\log\ n}. \tag{1}$$

In order to compute the FD, we will make use of the gliding box method on a converted black and white image. Let S_N be a given black and white image (BW) with 1 and 0 in correspondence with respectively, black and white pixels, we can consider a gliding box of r-length, so that

$$\mu_r(k) = \sum_{s=k}^{k+r-1} \nu_{\text{sh}}^{*} \tag{2}$$

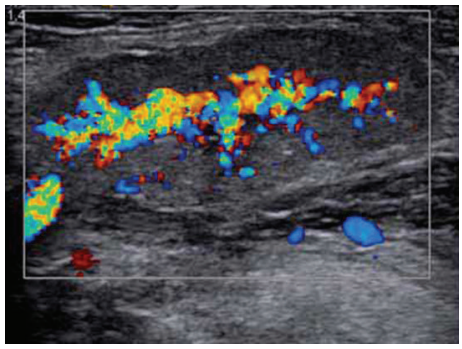


Figure 2: Colour coded Doppler ultrasonography (same case as Figure 1). In the central part of the gland there are vessels (blue and red according to the direction of the blood flow in relation to the transducer). The amplitude and extension of the colour signal are suggestive of hyperaemia (in this case it was an acute inflammation of the submandibular salivary gland).

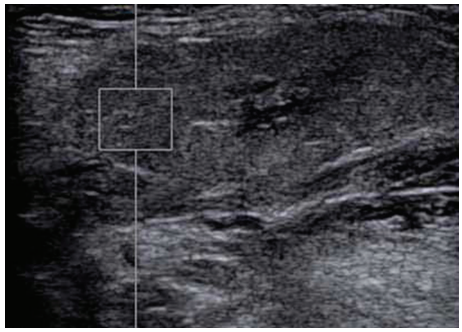


Figure 3: Elastogram of the submandibular gland (on the right side, inflamed gland) using the ARFI procedure. The measurements are made in an area of glandular parenchyma, in a predefined rectangular area, vessel free. The ultrasound speed is 2,55 m/sec.

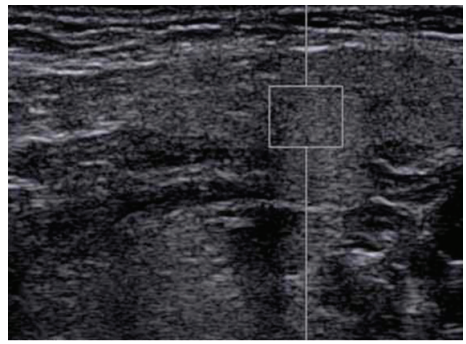


Figure 4: Elastogram of the submandibular gland (left side, normal gland) by means of ARFI procedure. The sample rectangle is positioned subscapular, in a similar position as it was on the right side gland. The ultrasound speed in the measured area is 1,36 m/sec.

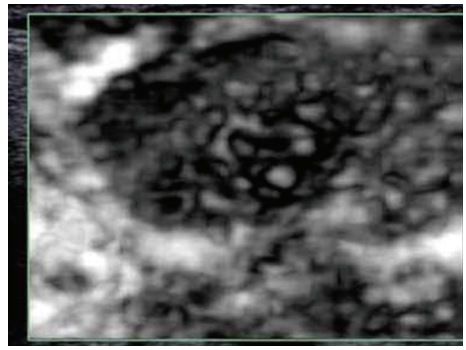


Figure 5: Qualitative (black and white coded; black is rigid; white is soft) elastogram (ARFI procedure) of the submandibular inflamed gland (right side). The pathological area inside the gland is well defined. This area presents a high rigidity index in relation to the amplitude of the pathological process.

is the frequency of "1" within the box. The corresponding probability is

$$p_r(k) = \frac{1}{r} \sum_{s=k}^{k+r-1} v_{\text{sh}}^{*}. \tag{3}$$

Then the box moves to the next position $k+1$ so that we obtain the probability distribution

$$\{p_r(k)\}_{k=1,\ldots,N}, \tag{4}$$

so that we can compute the frequency of "1" within the box. The FD is computed on such gliding boxes through (1).

3. Results

3.1. Fractal Dimension for 2D Ultrasound and Elastographic Images. Concerning the fractal dimension of the elastographic images, as given by (1), we can see (Table 1) that the highest FD is shown by Figure 7 and lowest by the Figure 8.

The images were analyzed in 8-bit using the Image J software (tools box counting).

The figures are referred to a patient with an acute inflammation of the submandibular gland.

Figure 1 shows a 2D ultrasound evaluation in grey scale. Figure 2 shows a 2D colour flow map evaluation (duplex sonography). Figures 3 and 4 were obtained by using the method elastography ARFI-Siemens, and they display quantitative information. The values of fractal dimension (FD) of Figures 3 and 4 are similar, and it is not possible to distinguish between pathological (Figure 3) and normal (Figure 4) states. The Figures 5 and 6 are obtained through elastography ARFI with qualitative information. From the fractal analysis by the box counting method, we have noticed that the value of Fd is lower (1.650) in Figure 5 (pathological condition) than Figure 6 (normal state). Figures 7 (pathological state) and 8 (normal state) were obtained through real time elastography.

From the computations, we can note that the higher value of Fd belongs to the pathological state (1.907), thus suggesting that the Fd increases during the evolution of the pathology (increasing degeneracy). Therefore, from Fd, analysis is possible to distinguish between pathological state and normal state of tissues by real time elastography because it is the better method to discriminate Fd values in a clear, sharp way.

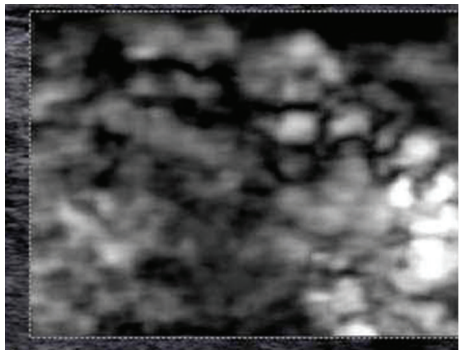


Figure 6: Qualitative (black and white coded; black is rigid; white is soft) elastogram (ARFI procedure) of the normal gland (considered to be the "witness," on the left side). The dispersion of the vectors of speed is obvious. There is no obvious compact hard parenchyma as in the right pathological gland (Figure 5).

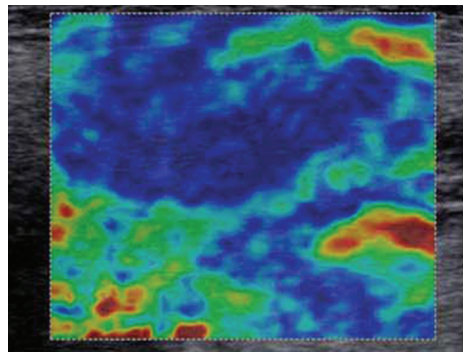


Figure 7: Real-time elastography (qualitative colour coded elastography; blue is rigid; red is soft) obtained by the compression of the right submandibular gland. The blue colour is in direct relation to the rigid parenchyma which is considered to be pathological.

Table 1: Fractal values.

Type of image	Fractal value
2D evaluation ultrasound grey scale	1.777
Duplex sonography	1.754
ARFI (quantitative)—Ps	1.771
ARFI (quantitative)—Ns	1.796
ARFI (qualitative)—Ps	1.650
ARFI (qualitative)—Ns	1.701
Real-time elastography—Ps	1.907
Real-time elastography—Ns	1.543

Ps: pathological state, Ns: normal situation.

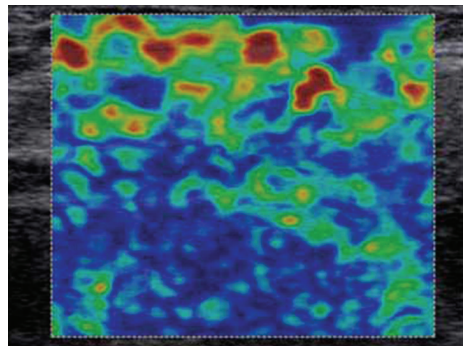


Figure 8: Real-time elastography (qualitative colour coded elastography; blue is rigid; red is soft) obtained by the compression of the left submandibular gland (normal). This is a normal pattern for the gland, suggestive of parts of different elasticity.

4. Discussion

Elastography is an ultrasonographic technique which appreciates tissue stiffness either by evaluating a colour map [23, 24] or by quantifying the shear wave velocity generated by the transmission of an acoustic pressure into the parenchyma (ARFI technique) [25–27]. In the first situation, the visualization of the tissue stiffness implies a "real-time" representation of the colour mode elastographic images overlapped on the conventional gray-scale images, each value (from 1 to 255) being attached to a color. The system uses a color map (red-green-blue) in which stiff tissues are coded in dark blue, intermediate ones in shades of green, softer tissues in yellow and the softest in red, but the color scale may be reversed in relation to how the equipment is calibrated. Depending on the color and with the help of a special software, several elasticity scores that correlate with the degree of tissue stiffness can be calculated [23]. Numerous clinical applications using these procedures were introduced into routine practice, many of them being focused on the detection of tumoral tissue in breast, thyroid, and prostate.

In the last years, a new elastographic method, based on the ARFI technique (acoustic radiation force impulse imaging), is available on modern ultrasound equipment. The ARFI technique consists in a mechanical stimulation of the tissue on which it is applied by the transmission of a short time acoustic wave (<1 ms) in a region of interest, determined by the examiner, perpendicular on the direction of the pressure waves, and leading to a micronic scale "dislocation" of the tissues. Therefore, in contrast with the usual ultrasonographic examination, where the sound waves have an axial orientation, the shear waves do not interact directly with the transducer. Furthermore, the shear waves are attenuated 10.000 faster than the conventional ultrasound waves and therefore need a higher sensitivity in order to be measured [25–29]. Detection waves, which are simultaneously generated, have a much lower intensity than the pressure acoustic wave (1 : 1000). The moment when the detection waves interact with the shear waves represents the time passed from the moment the shear waves were generated until they crossed the region of interest. The shear waves are registered in different locations at various moments and thus the shear wave velocity is automatically calculated, the stiffer the organ the higher the velocity of the shear waves. Therefore, the shear wave velocity is actually considered to be an intrinsic feature of the tissue [25–29]. In current clinical practice, the same transducer is used both to generate the pressure acoustic wave and to register the tissue dislocation. Since the technique is implemented

in the ultrasound equipment through software changes, B mode ultrasound examination, color Doppler interrogation and ARFI images are all possible on the same machine [30].

Currently, elastography is widely studied in relation to different clinical applications: breast, thyroid, liver, colon and prostate [29, 31–36]. The application in salivary gland pathology has been singularly considered at least in our literature database. Some reports present the utility of elastography in a better delineation of tumors of these glands. Applications on diffuse disease are few although the importance of this kind of pathology is important! Inflammations of salivary glands occur in many conditions and the incidence is significant. There is a need for accurate diagnosis, staging, and prognosis. The occurrence of complications is also very important! Elastography represents a "virtual" way of palpation reproductive and with possibility of quantification.

Although there are several improvements, the main limitation of elastography is the dependency of the procedure to the operator's experience. This characteristic makes elastography vulnerable with a quite high amount of variations of elastographic results and interpretation. A more accurate analysis of the elastographic picture based on very precise evaluation as fractal analysis is an obvious step forward. In our preliminary study, the difference between normal and pathologic submandibular tissue using the fractal analysis was demonstrated. Because of the very new technologies accessible in practice as elastography is, and because of the mathematical instruments available as fractal analysis of the pictures, we are encouraged to believe that the ultrasound procedure might become operator independent and more confident for subtle diagnosis. However, a higher number of pictures coming from different patients with diffuse diseases in different stages of evolution are needed.

5. Conclusion

In this work, the multi-fractality of 2D and elastographic images of diffuse pathological states in submandibular glands has been investigated. The corresponding FD has been computed and has shown that images with the highest FD correspond to the existence of pathology. The extension of this study with incrementing the number of ultrasound images and patients is needed to demonstrate the practical utility of this procedure.

Conflict of Interests

The authors declare that there is no conflict of interests concerning the validity of this research with respect to some possible financial gain.

References

[1] V. Anh, G. Zhi-Min, and L. Shun-Chao, "Fractals in DNA sequence analysis," *Chinese Physics*, vol. 11, no. 12, pp. 1313–1318, 2002.

[2] S. V. Buldyrev, N. V. Dokholyan, A. L. Goldberger et al., "Analysis of DNA sequences using methods of statistical physics," *Physica A*, vol. 249, no. 1–4, pp. 430–438, 1998.

[3] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.

[4] G. Pierro, "Sequence complexity of Chromosome 3 in Caenorhabditis elegans," *Advances in Bioinformatics*, vol. 2012, Article ID 287486, 12 pages, 2012.

[5] 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.

[6] D. P. Ferro, M. A. Falconi, R. L. Adam et al., "Fractal characteristics of May-Grünwald-Giemsa stained chromatin are independent prognostic factors for survival in multiple myeloma," *PLoS ONE*, vol. 6, no. 6, Article ID e20706, 2011.

[7] K. Metze, R. L. Adam, and R. C. Ferreira, "Robust variables in texture analysis," *Pathology*, vol. 42, no. 6, pp. 609–610, 2010.

[8] K. Metze, "Fractal characteristics of May Grunwald Giemsa stained chromatin are independent prognostic factors for survival in multiple myeloma," *PLoS One*, vol. 6, no. 6, pp. 1–8, 2011.

[9] P. Dey and T. Banik, "Fractal dimension of chromatin texture of squamous intraepithelial lesions of cervix," *Diagnostic Cytopathology*, vol. 40, no. 2, pp. 152–154, 2012.

[10] 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.

[11] R. F. Voss, "Long-range fractal correlations in DNA introns and exons," *Fractals*, vol. 2, no. 1, pp. 1–6, 1992.

[12] C. A. Chatzidimitriou-Dreismann and D. Larhammar, "Long-range correlations in DNA," *Nature*, vol. 361, no. 6409, pp. 212–213, 1993.

[13] A. Fukushima, M. Kinouchi, S. Kanaya, Y. Kudo, and T. Ikemura, "Statistical analysis of genomic information: long-range correlation in DNA sequences," *Genome Informatics*, vol. 11, pp. 315–3316, 2000.

[14] M. Li, "Fractal time series-a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.

[15] M. Li and W. Zhao, "Quantitatively investigating locally weak stationarity of modified multifractional Gaussian noise," *Physica A*, vol. 391, no. 24, pp. 6268–6278, 2012.

[16] F. D'Anselmi, M. Valerio, A. Cucina et al., "Metabolism and cell shape in cancer: a fractal analysis," *International Journal of Biochemistry and Cell Biology*, vol. 43, no. 7, pp. 1052–1058, 2011.

[17] I. Pantic, L. Harhaji-Trajkovic, A. Pantovic, N. T. Milosevic, and V. Trajkovic, "Changes in fractal dimension and lacunarity as early markers of UV-induced apoptosis," *Journal of Theoretical Biology*, vol. 303, no. 21, pp. 87–92, 2012.

[18] C. Vasilescu, D. E. Giza, P. Petrisor, R. Dobrescu, I. Popescu, and V. Herlea, "Morphometrical differences between resectable and non-resectable pancreatic cancer: a fractal analysis," *Hepatogastroentology*, vol. 59, no. 113, pp. 284–288, 2012.

[19] B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York, NY, USA, 1982.

[20] J. W. Baish and R. K. Jain, "Fractals and cancer," *Cancer Research*, vol. 60, no. 14, pp. 3683–3688, 2000.

[21] S. S. Cross, "Fractals in pathology," *Journal of Pathology*, vol. 182, no. 1, pp. 1–18, 1997.

[22] A. R. Backes and O. M. Bruno, "Segmentação de texturas por análise de complexidade," *Journal of Computer Science*, vol. 5, no. 1, pp. 87–95, 2006.

[23] M. Friedrich-Rust, M. F. Ong, E. Herrmann et al., "Real-time elastography for noninvasive assessment of liver fibrosis in chronic viral hepatitis," *American Journal of Roentgenology*, vol. 188, no. 3, pp. 758–764, 2007.

[24] A. Săftoui, D. I. Gheonea, and T. Ciurea, "Hue histogram analysis of real-time elastography images for noninvasive assessment of liver fibrosis," *American Journal of Roentgenology*, vol. 189, no. 4, pp. W232–W233, 2007.

[25] D. Dumont, R. H. Behler, T. C. Nichols, E. P. Merricks, and C. M. Gallippi, "ARFI imaging for noninvasive material characterization of atherosclerosis," *Ultrasound in Medicine and Biology*, vol. 32, no. 11, pp. 1703–1711, 2006.

[26] L. Zhai, M. L. Palmeri, R. R. Bouchard, R. W. Nightingale, and K. R. Nightingale, "An integrated indenter-ARFI imaging system for tissue stiffness quantification," *Ultrasonic Imaging*, vol. 30, no. 2, pp. 95–111, 2008.

[27] R. H. Behler, T. C. Nichols, H. Zhu, E. P. Merricks, and C. M. Gallippi, "ARFI imaging for noninvasive material characterization of atherosclerosis part II: toward in vivo characterization," *Ultrasound in Medicine and Biology*, vol. 35, no. 2, pp. 278–295, 2009.

[28] K. Nightingale, M. S. Soo, R. Nightingale, and G. Trahey, "Acoustic radiation force impulse imaging: in vivo demonstration of clinical feasibility," *Ultrasound in Medicine and Biology*, vol. 28, no. 2, pp. 227–235, 2002.

[29] M. Lupsor, R. Badea, H. Stefanescu et al., "Performance of a new elastographic method (ARFI technology) compared to unidimensional transient elastography in the noninvasive assessment of chronic hepatitis C. Preliminary results," *Journal of Gastrointestinal and Liver Diseases*, vol. 18, no. 3, pp. 303–310, 2009.

[30] B. J. Fahey, K. R. Nightingale, R. C. Nelson, M. L. Palmeri, and G. E. Trahey, "Acoustic radiation force impulse imaging of the abdomen: demonstration of feasibility and utility," *Ultrasound in Medicine and Biology*, vol. 31, no. 9, pp. 1185–1198, 2005.

[31] R. S. Goertz, K. Amann, R. Heide, T. Bernatik, M. F. Neurath, and D. Strobel, "An abdominal and thyroid status with acoustic radiation force impulse elastometry—a feasibility study: acoustic radiation force impulse elastometry of human organs," *European Journal of Radiology*, vol. 80, no. 3, pp. e226–e230, 2011.

[32] S. R. Rafaelsen, C. Vagn-Hansen, T. Sørensen, J. Lindebjerg, J. Pløen, and A. Jakobsen, "Ultrasound elastography in patients with rectal cancer treated with chemoradiation," *European Journal of Radiology*, 2013.

[33] G. Taverna, P. Magnoni, G. Giusti et al., "Impact of real-time elastography versus systematic prostate biopsy method on cancer detection rate in men with a serum prostate-specific antigen between 2.5 and 10 ng/mL," *ISRN Oncology*, vol. 2013, Article ID 584672, 5 pages, 2013.

[34] L. Rizzo, G. Nunnari, M. Berretta, and B. Cacopardo, "Acoustic radial force impulse as an effective tool for a prompt and reliable diagnosis of hepatocellular carcinoma—preliminary data," *European Review for Medical and Pharmacological Sciences*, vol. 16, no. 11, pp. 1596–1598, 2012.

[35] Y. F. Zhang, H. X. Xu, Y. He et al., "Virtual touch tissue quantification of acoustic radiation force impulse: a new ultrasound elastic imaging in the diagnosis of thyroid nodules," *PLoS One*, vol. 7, no. 11, Article ID e49094, 2012.

[36] M. Dighe, S. Luo, C. Cuevas, and Y. Kim, "Efficacy of thyroid ultrasound elastography in differential diagnosis of small thyroid nodules," *European Journal of Radiology*, 2013.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 208402, 9 pages
http://dx.doi.org/10.1155/2013/208402

Research Article

Nonlinear Radon Transform Using Zernike Moment for Shape Analysis

Ziping Ma,[1,2] Baosheng Kang,[1] Ke Lv,[3] and Mingzhu Zhao[4]

[1] *School of Information and Technology, Northwest University, Xi'an 710120, China*
[2] *School of Information and Computing Sciences, North University for Nationalities, Yinchuan 750021, China*
[3] *College of Computing & Communication Engineering, Graduate University of Chinese Academy of Sciences, Beijing 100049, China*
[4] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*

Correspondence should be addressed to Ziping Ma; zipingma@gmail.com

Received 18 January 2013; Accepted 22 March 2013

Academic Editor: Carlo Cattani

We extend the linear Radon transform to a nonlinear space and propose a method by applying the nonlinear Radon transform to Zernike moments to extract shape descriptors. These descriptors are obtained by computing Zernike moment on the radial and angular coordinates of the pattern image's nonlinear Radon matrix. Theoretical and experimental results validate the effectiveness and the robustness of the method. The experimental results show the performance of the proposed method in the case of nonlinear space equals or outperforms that in the case of linear Radon.

1. Introduction

Shape analysis methods have been broadly applied to biomedical signal processing, object recognition, image retrieval, target tracking, and so forth [1]. Moments methods [2, 3] can be referred to shape descriptors because of their good characterization in describing different shapes. The most important properties of shape descriptors achieved by different moments are invariance, including translation, rotation, scaling, and stretching, stability to noise, and completeness [4].

In the past twenty years, many attentions have been paid to the completeness property of the invariant descriptor set in pattern recognition and other similar application fields. These kinds of methods can be obtained by the following processes. Firstly, Fourier transform or Radon transform is employed to map the image into other space. Secondly, the different ideas can be conceived to construct invariant descriptors based on the information in new space. Sim et al. [5] gave a new method for texture image retrieval. They converted the images in Fourier domain and calculated modified Zernike moments to extract the texture descriptors. It is tested that the descriptor has higher accuracy comparing to Gabor, Radon, and wavelet based methods and requires low computational effort. However, it is not invariant to scale. Wang et al. [6] and Xiao et al. [7] introduced the Radon transform to Fourier-Mellin transform to achieve RST (rotation, scaling, and translation) invariance and RS invariance combined blur, respectively. In virtue of Xiao's idea, Zhu et al. [8] constructed RST invariants using Radon transforms and complex moments in digital watermarking. Similarly, the Zernike moments can be connected with Radon transform. Rouze et al. [9] described a method to design an approach by calculating the Zernike moments of an image from its Radon transform using a polynomial transform in the position coordinate and a Fourier transform in the angular coordinate. However, the proposed descriptors are only invariant to rotation. Meanwhile, in order to improve the precision of image retrieval and noise robustness, Hoang and Tabbone [10] proposed a new method similar to Xiao's descriptor to obtain RST invariance based on the Radon, Fourier, and Mellin transform.

Then, Radon transform is widely applied in many methods mainly because of its better properties in projection space

[11–15]. In the projective space, a rotation of the original image results in a translation in the angle variable, and a scaling of the original image leads to a scaling in the spatial variable together with an amplitude scaling [16, 17]. Based on these properties, a rotation and scaling invariant function is easy to construct and highly robust to noise.

Enlightened by the peers' research works, we extend Radon transform to nonlinear Radon transform and propose a new set of complete invariant descriptors by applying Zernike moments to the radial coordinate of the pattern's nonlinear Radon space of an image [18–22].

The remainder of this paper is organized as follows. In Section 2, we briefly review the definition of nonlinear Radon transform and Zernike moments, and propose a new method based on Zernike moment and nonlinear Radon transform. In Section 3, the comparative experiments of the proposed approach with Hu moment invariance, Chong's method is conducted in terms of image retrieval efficiency, different noise robustness. Section 4 concludes this paper.

2. Nonlinear Radon Transform and Zernike Moments

2.1. Nonlinear Radon Transform. The nonlinear Radon transform of an image function $f(x, y)$ is defined as [10]

$$P(r,\theta) = R(r,\theta)\{f(x,y)\} = \iint_{-\infty}^{\infty} f(x,y)\,\delta\left(r^{q_1} - T\left(\psi(x,y),\theta\right)\right)dx\,dy, \tag{1}$$

where $\psi(x, y) \in L^2(D)$, q_1 is a real instance, θ denotes the angle vector formed by the function $\psi(x, y)$, and $T(\psi(x, y), \theta)$ is a rotation function by $\psi(x, y)$ with an angel of θ and defined by

$$T\left(\psi(x,y),\theta\right) - r^{q_1} = 0. \tag{2}$$

The nonlinear Radon transform indicates curve integral of the image function $f(x, y)$ along different curves. The parameter q_1 can control the shape of curve. Different curves can be obtained by the values of parameter q_1 and function $\psi(x, y)$.

Especially when $\psi(x, y) = x$ and $q_1 = 1$, $T(\psi(x, y), \theta) = x\cos\theta + y\sin\theta$. This reveals that the linear Radon transform is the special case of nonlinear Radon transform. The results of different curves' Radon transform are shown in Table 1.

The nonlinear Radon transform has some properties that are beneficial for pattern recognition as outlined below.

(1) Periodicity: the nonlinear Radon transform of $f(x, y)$ is periodic in the variable θ with period 2π when $\psi(x, y)$ is an arbitrarily parametric inference,

$$P(r,\theta) = P(r,\theta \pm 2k\pi). \tag{3}$$

(2) Resistance: if $f_1(x, y)$ and $f_2(x, y)$ are two images with little difference when $\psi(x, y)$ is arbitrarily parametric inference, the corresponding nonlinear Radon transform of $f_1(x, y)$ and $f_2(x, y)$ are as followes:

$$\left|P_1(r,\theta) - P_2(r,\theta)\right| \le \iint_D \left|\left|f_1(r,\theta) - f_2(r,\theta)\right|\delta\left(r^{q_1} - T\left(\psi(x,y),\theta\right)\right)\right| dx\,dy. \tag{4}$$

(3) Translation: a translation of $f(x, y)$ by a vector $\overrightarrow{u} = (x_0, y_0)$ results in a shift in the variable r of $P(r, \theta)$ by a distance $d = x_0\cos\theta + y_0\sin\theta$ and equals to the length of the projection of $\overrightarrow{u}$ onto the line $x\cos\theta + y\sin\theta = r$,

$$P(r,\theta) = P\left(r - x_0\cos\theta - y_0\sin\theta,\theta\right). \tag{5}$$

(4) Rotation: a rotation of $f(x, y)$ by an angle θ_0 implies a shift in the variable θ of $P(r, \theta)$ by a distance θ_0 when $\psi(x, y)$ is arbitrarily parametric inference,

$$P(r,\theta) \longrightarrow P\left(r,\theta + \theta_0\right). \tag{6}$$

(5) Scaling: a scaling of $f(x, y)$ by a factor of a results in a scaling in the variable r and $1/a$ of amplitude of $P(r, \theta)$, respectively, when $\psi(x, y)$ represents ellipse or hyperbola curve,

$$f(ax,ay) \longrightarrow \frac{1}{a^2}P(ar,\theta). \tag{7}$$

2.2. Zernike Moment. The radial Zernike moments of order (p, q) of an image function $f(r, \theta)$, is defined as

$$Z_{pq} = \frac{(p+1)}{\pi}\int_0^{2\pi}\int_0^1 R_{pq}(r)\,e^{-q\tilde{i}\theta} f(r,\theta)\,r\,dr\,d\theta, \tag{8}$$

where the radial Zernike moment of order (p, q) is defined by the following equation:

$$R_{pq}(r) = \sum_{\substack{k=q \\ p-k=\text{even}}}^{p} B_{p|q|k}\,r^k. \tag{9}$$

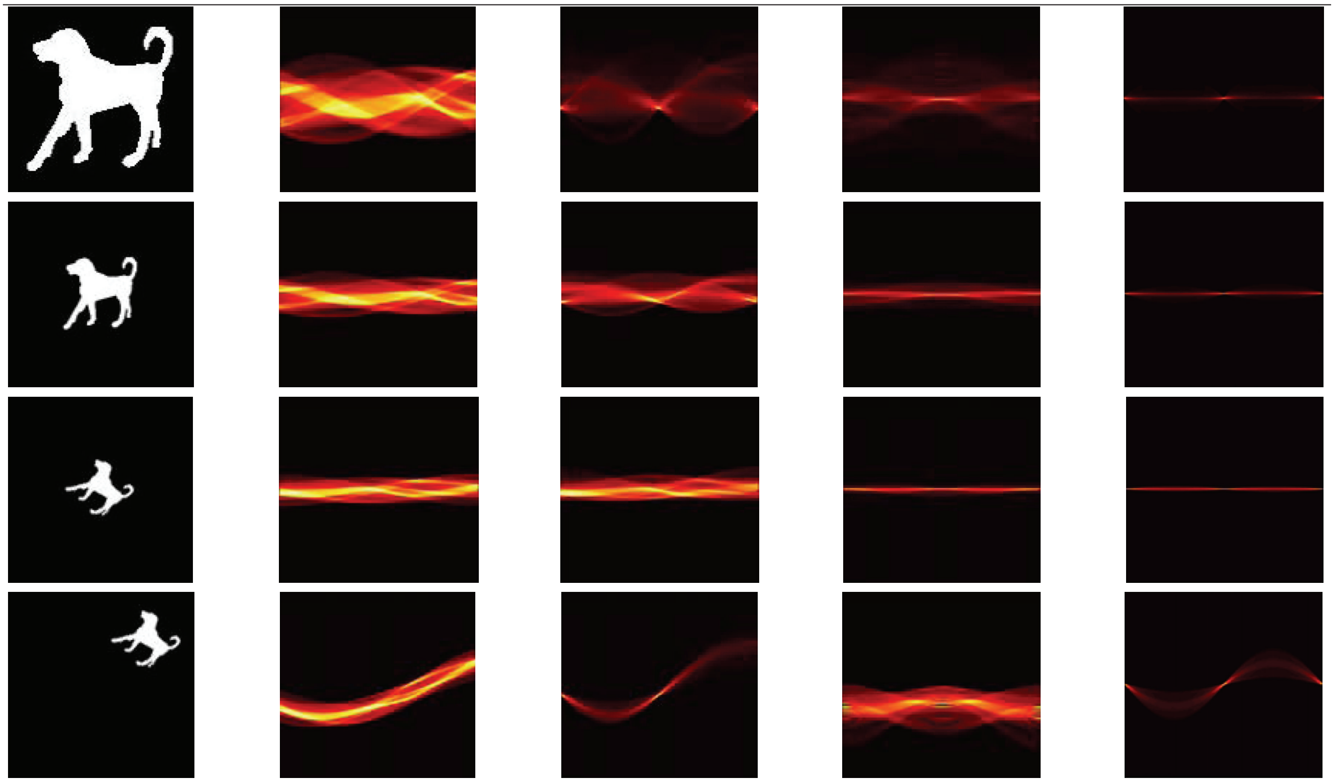
With

$$B_{p|q|k} = \begin{cases} \dfrac{(-1)^{((p-k)/2)}\left((p+k)/2\right)!}{\left((p-k)/2\right)!\left((q+k)/2\right)!\left((k-q)/2\right)!}, & p-k = \text{even} \\ 0, & p-k = \text{odd}. \end{cases} \tag{10}$$

2.3. NRZM Descriptor Based on Nonlinear Radon Transform and Zernike Moment. The Zernike moment is carried out to be computed after the projective matrix of nonlinear Radon transform is mapped to the polar coordinate (NRZM).

Table 1: The diagrams of results using different curves' Radon transform.

	Line Radon transform	Parabola Radon transform	Ellipse Radon transform	Hyperbola Radon transform



The computational process of our proposed method, NRZM, is illuminated in Figure 1.

Supposed $\tilde{f}(x,y)$ is the image $f(x,y)$ rotated by rotational angle β and scaled by scaling factor λ, and Radon transform of $\tilde{f}(x,y)$ is given by

$$\tilde{P}(r,\theta) = \lambda P\left(\frac{r}{\lambda}, \theta + \beta\right). \quad (11)$$

The Zernike moments of $\tilde{P}(r,\theta)$ is

$$\begin{aligned}\tilde{Z}_{pq} &= \frac{p+1}{\pi}\int_0^{2\pi}\int_0^1 \tilde{P}(r,\theta) R_{pq}(\lambda r) e^{(-\hat{i}q\theta)} r\,dr\,d\theta \\ &= \frac{p+1}{\pi}\int_0^{2\pi}\int_0^1 \lambda P\left(\frac{r}{\lambda}, \theta+\beta\right) R_{pq}(\lambda r) e^{(-\hat{i}q\theta)} r\,dr\,d\theta.\end{aligned} \quad (12)$$

The radial Zernike polynomials $R_{pq}(\lambda r)$ can be expressed as a series of $R_{pq}(r)$ as follows:

$$R_{pq}(\lambda r) = \sum_{k=q}^{p} R_{pk}(r) \sum_{i=q}^{k} \lambda^i B_{pqi} D_{pik}. \quad (13)$$

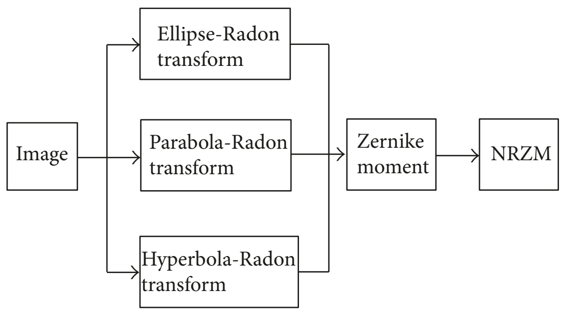


Figure 1: The computation process of NRZM.

The derivation process of (13) is given in the Appendix. According to (12), we have

$$\tilde{Z}_{pq} = \frac{p+1}{\pi} \times \int_0^{2\pi}\int_0^1 \lambda P\left(\frac{r}{\lambda}, \theta+\beta\right) \times \sum_{k=q}^{p} R_{pk}(r) \sum_{i=q}^{k} \lambda^i B_{pqi} D_{pik} e^{(-\hat{i}q\theta)} r\,dr\,d\theta. \quad (14)$$

Let $\tau = r/\lambda$, $\varphi = \theta + \beta$, (14) can be rewritten as

$$\begin{aligned}
\widetilde{Z}_{pq} &= \frac{p+1}{\pi} \\
&\quad \times \int_0^{2\pi} \int_0^1 \lambda P(\tau,\varphi) \sum_{k=q}^{p} R_{pk}(r) \\
&\qquad \times \sum_{i=q}^{k} \left(\lambda^i B_{pqi} D_{pik}\right) e^{(-\hat{i}q(\varphi-\beta))} \lambda^2 \tau d\tau\, d\varphi \\
&= \frac{p+1}{\pi} e^{\hat{i}q\beta} \\
&\quad \times \int_0^{2\pi} \int_0^1 P(\tau,\varphi) \\
&\qquad \times \sum_{k=q}^{p} R_{pk}(r) \sum_{i=q}^{k} \left(\lambda^{i+3} B_{pqi} D_{pik}\right) e^{-\hat{i}q\varphi} \tau d\tau\, d\varphi \\
&= \frac{p+1}{\pi} e^{\hat{i}q\beta} \\
&\quad \times \sum_{k=q}^{p} \sum_{i=q}^{k} \left(\lambda^{i+3} B_{pqi} D_{pik}\right) \\
&\qquad \times \int_0^{2\pi} \int_0^1 P(\tau,\varphi) R_{pk}(r) e^{-\hat{i}q\varphi} \tau d\tau\, d\varphi \\
&= e^{\hat{i}q\beta} \sum_{k=q}^{p} \sum_{i=q}^{k} \left(\lambda^{i+3} B_{pqi} D_{pik}\right) Z_{pk}.
\end{aligned} \tag{15}$$

Equation (15) shows that the radial Zernike moments of being rotated image can be expressed as a linear combination of the radial Zernike moments of original image. Based on this relationship, we can construct a set of rotation invariant I_{pq} which is described as follows:

$$I_{pq} = \exp\left(jqarg\left(Z_{11}\right)\right) \sum_{k=q}^{p} \left(\sum_{i=q}^{k} Z_{00}^{-((i+3)/3)} B_{pqi} D_{pik} \right) Z_{pk}. \tag{16}$$

Then, I_{pq} is invariant to rotation and translation.

3. Experimental Results and Discussions

This section is intended to test the performance of a complete family of similarity invariants introduced in the previous section for images retrieval by comparison, Chong's method presented in [12], Hu moment presented in [13]. In the experiments, the feature descriptors are calculated by

$$RZM = \left[I_f(1,0), I_f(1,1), \ldots, I_f(M,M)\right]. \tag{17}$$

Three subsections are included in this section. In the first subsection, we test the retrieval efficiency of proposed descriptors in shape 216 dataset. This dataset is composed of 18 shape categories with 12 samples per category, and each of every category cannot be obtained by RST transforming from any other shape from the same category. In the second subsection, we test robustness of proposed descriptors in different noisy dataset. In the third subsection, we verify the rotation invariance of the proposed method.

Table 2: The most suitable values of parameters.

The kind of curves	q_0	q_1
Ellipse	190/90	1
Hyperbola	350/100	2
Parabola	2000	2

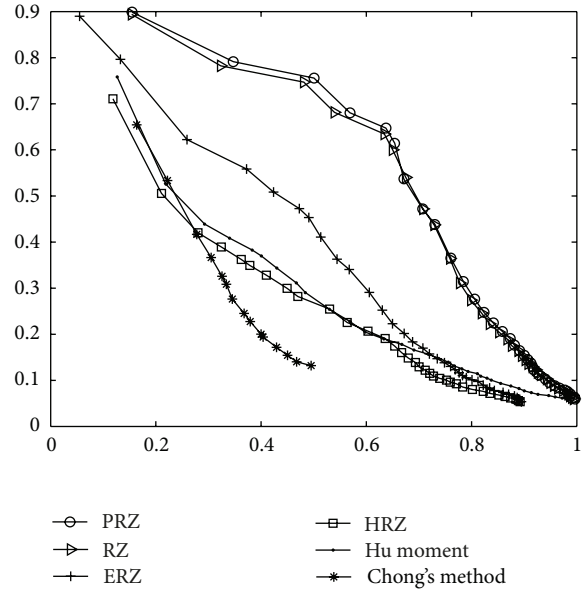


Figure 2: The precision-recall curve of shape 216.

3.1. Experiment 1. The kind of curves is changing with the controlled parameters varying. So, the retrieval efficiency is different with the controlled parameters. Many experiments are conducted to find the best parameters' values of every curve in nonlinear Radon transform, and finally the most suitable values of parameters are listed in Table 2. In the subsequent experiments, we analyze the retrieval efficiency of linear Radon transform, ellipse Radon transform, hyperbola Radon transform, and parabola Radon transform with Zernike moment, respectively, which is referred to as NZ, EPZ, HPZ, and PRZ, respectively.

In order to obtain the best retrieval efficiency of every curve Radon, the comparative precisions-recall curves in Shapes 216 are shown in Figure 2. It can be seen that the precision-recall curve of PRZ moves downward more slowly than those of others, which indicates that the retrieval efficient of PRZ is slightly higher than that of RZ while HRZ is weaker than PRZ and RZ.

The comparative number of relevant image upon every category is a better insight into the performance of proposed method as shown in Figure 3. It is easy to see that almost the

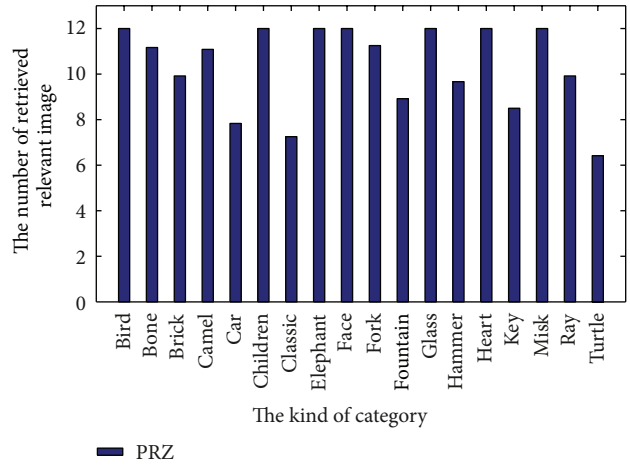


Figure 3: The retrieved number of every category in shape 216.

number of relevant image in every category is higher than 6, especially in bird, children, elephant, face, glass, hammer, heart, and misk.

3.2. Experiment 2. The robustness of the proposed descriptors is demonstrated using eight datasets added additive "salt & pepper" and "Gaussian" noise, respectively. The first seven datasets are generated from original shape 216 database, and each image is corrupted by "salt & pepper" noise with SNR varying from 16 to 4 dB with 2 dB decrements. The last one is generated from shape 216 added "Gaussian" noise with noise density = $0.01, \ldots, 0.2$.

The retrieval experiments are conducted again in the datasets mentioned above and the precision-recall curves of comparative descriptors are depicted in Figure 4. From Figures 4(a)–4(g), it can be observed that efficiency of the PRZ and RZ are similar. It also can be seen that the PRZ and RZ descriptors have better performances than other comparative methods in "salt and pepper" noisy datasets from SNR = 16 to 8, while Hu moment and Chong's descriptors have similarly the worse performance. However, when SNR = 6 and SNR = 4, the situation has changed. The deterioration appears in the PRZ and RZ because their precision-recall curves moves downward more rapidly than those of HPZ and EPZ, while they move downward more slowly than those of Chong's method and CMI. This demonstrates that PRZ and RZ descriptor are sensitive than other nonlinear methods' descriptors when the value of SNR is low of 8 though it has the stronger robustness than Chong's method and Hu moment. In short, the impact of noise on RZ, ERZ, HRZ, and PRZ curves sometimes were little similar or sometimes differ from one to another. It is also observed that

(1) as the values of SNR decrease, the curves of all the descriptors generally move downwards;

(2) Hu moment and Chong's descriptors are very sensitive to noise, and their performance has not changed much under different levels of noise;

(3) Hu moment method has more resistance to "salt & pepper" noise than Chong's descriptors;

(4) among the RZ, ERZ, PRZ, and HRZ, the resistance of PRZ is the strongest to "salt & pepper" noise and that of RZ is close to PRZ when the values of SNR are higher than 6;

(5) PRZ is always slightly more robust to "salt & pepper" noise than RZ except for SNR = 6 and SNR = 4;

(6) EPZ and HPZ descriptors are more robust to "salt & pepper" noise than PRZ and RZ when values of SNR are higher than 6.

However, the retrieval results shown in Figure 4(h) are essentially different from those in Figures 4(a)–4(g). It is clear that ERZ and HRZ are more robust to "Gaussian" noise than other methods because their precision-recall curves are absolutely on the top of others in the "Gaussian" noisy dataset. This indicates that "Gaussian" noise can result in poor performance in the case of linear transform. In these cases, the nonlinear Radon transform should be a top priority to be employed in the proposed method.

3.3. Experiment 3. The last test dataset is color objective dataset generated by choosing 7 sample images from Col and View subset. Each of the datasets is transformed by being rotated by 72 arbitrary angles (10–360) with 5 degree increment. As a result, the last dataset consists of 504 images, and the retrieval results are shown in Figure 5. From the figure, it can be concluded that the proposed descriptors are invariant to rotation, and the retrieval performance of PRZ is more efficient.

4. Conclusion

In this paper, we proposed a method to derive a set of rotation invariants using Radon transform and Zernike moments and extend linear Radon transform to nonlinear Radon transform.

Comparing to linear Radon transform, the proposed method can perform better or similar. However, the numerical experiments show that different curve Radon transforms

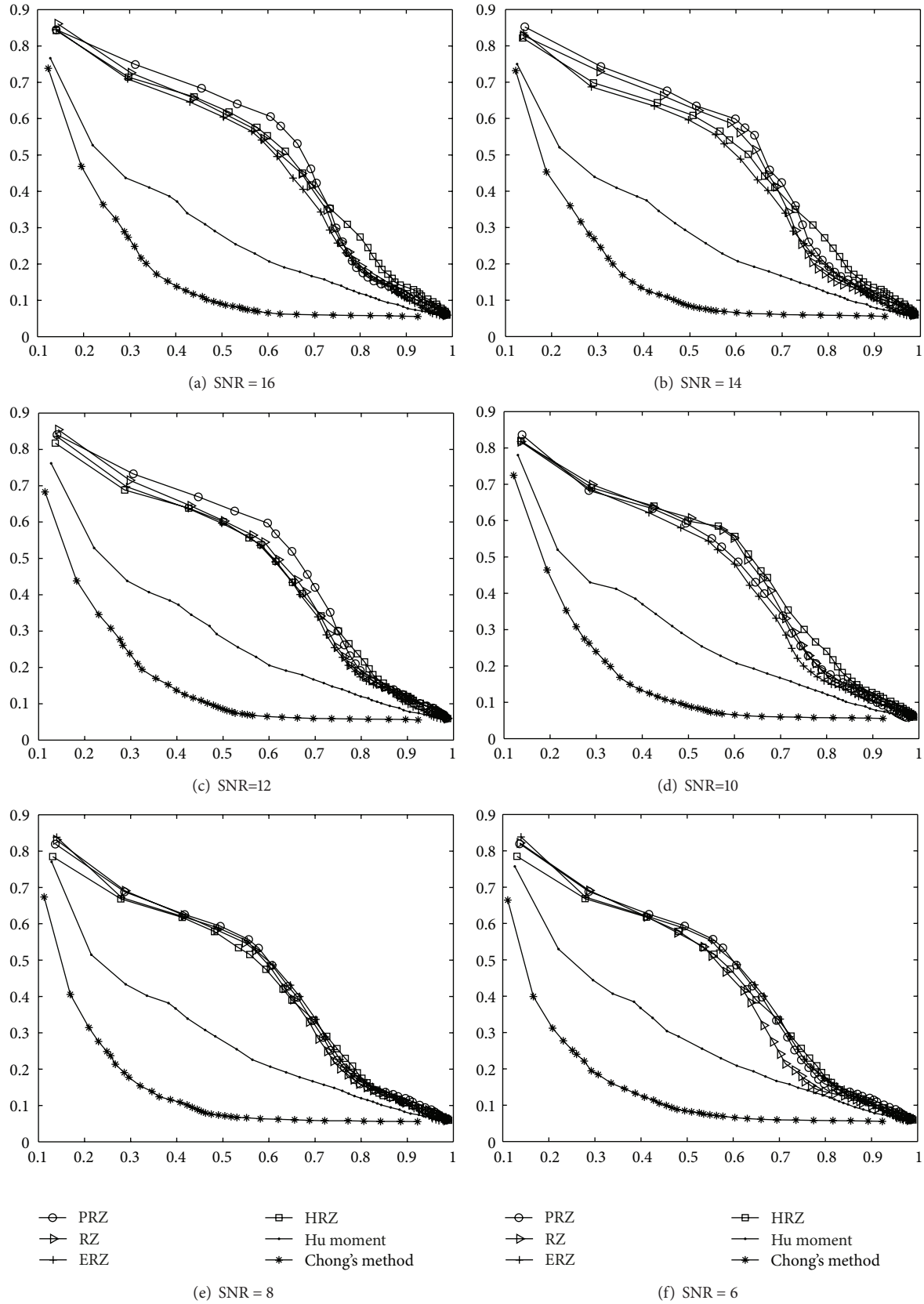
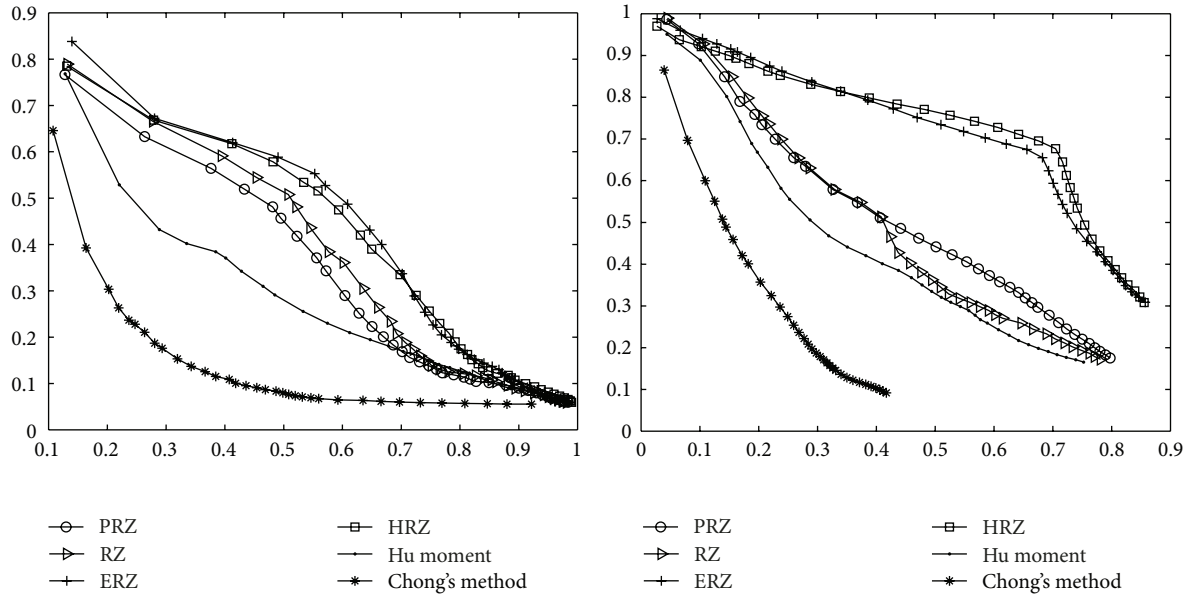


Figure 4: Continued.



(g) SNR =4 (h) Gaussian noisy dataset of shape 216

Figure 4: The precision upon recall curves of different descriptors on seven noisy datasets added "salt & pepper" and one "Gaussian" noisy dataset.

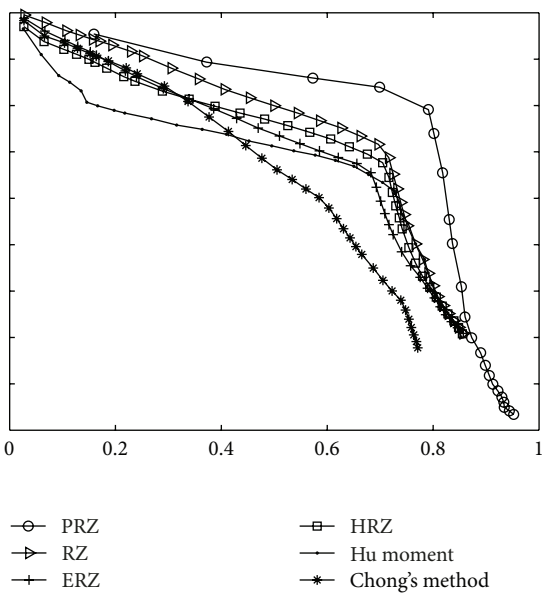


Figure 5: The precision-recall curves of different descriptors on rotated dataset.

and Zernike moment perform different. In the noiseless dataset, the retrieval efficiency of PRZ is higher than comparative methods. In the "salt & pepper" noise and the PRZ consistently performs better except SNR = 6 and SNR = 4. While when SNR = 6, SNR = 4, the EPZ and HPZ are most robust than RZ. And in "Gaussian" noise dataset, the proposed method in the cases of nonlinear Radon transform is more robust to "Gaussian" noise than that in the case of linear Radon transform. Moreover, the nonlinear Radon transform can be exploited to other application fields for engineer application and recognition for the sake of the good characteristic, especially their robustness.

Appendix

Proof of (13)

From (12), the radial Zernike polynomials can be expressed as a series of decreasing power of as follows:

$$\begin{pmatrix} R_{pq}(r) \\ R_{pq+1}(r) \\ \vdots \\ R_{pp}(r) \end{pmatrix} = \begin{pmatrix} B_{pqq} & B_{pqq+1} & \cdots & B_{pqp} \\ & B_{pq+1q+1} & \cdots & B_{pq+1p} \\ & & \ddots & \vdots \\ & & & B_{ppp} \end{pmatrix} \begin{pmatrix} r^q \\ r^{q+1} \\ \vdots \\ r^p \end{pmatrix}. \tag{A.1}$$

Since all the diagonal element B_{pii} are not zero, the matrix B is nonsingular, thus

$$\begin{pmatrix} r^q \\ r^{q+1} \\ \vdots \\ r^p \end{pmatrix} = \begin{pmatrix} B_{pqq} & B_{pqq+1} & \cdots & B_{pqp} \\ & B_{pq+1q+1} & \cdots & B_{pq+1p} \\ & & \ddots & \vdots \\ & & & B_{ppp} \end{pmatrix}^{-1} \times \begin{pmatrix} R_{pq}(r) \\ R_{pq+1}(r) \\ \vdots \\ R_{pp}(r) \end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} D_{pqq} & D_{pqq+1} & \cdots & D_{pqp} \\ & D_{pq+1q+1} & \cdots & D_{pq+1p} \\ & & \ddots & \vdots \\ & & & D_{ppp} \end{pmatrix} \\
&\quad \times \begin{pmatrix} R_{pq}(r) \\ R_{pq+1}(r) \\ \vdots \\ R_{pp}(r) \end{pmatrix}, \\
\begin{pmatrix} R_{pq}(\lambda r) \\ R_{pq+1}(\lambda r) \\ \vdots \\ R_{pp}(\lambda r) \end{pmatrix} &= \begin{pmatrix} B_{pqq} & B_{pqq+1} & \cdots & B_{pqp} \\ & B_{pq+1q+1} & \cdots & B_{pq+1p} \\ & & \ddots & \vdots \\ & & & B_{ppp} \end{pmatrix} \\
&\quad \times \begin{pmatrix} (\lambda r)^q \\ (\lambda r)^{q+1} \\ \vdots \\ (\lambda r)^p \end{pmatrix} \\
&= \begin{pmatrix} B_{pqq} & B_{pqq+1} & \cdots & B_{pqp} \\ & B_{pq+1q+1} & \cdots & B_{pq+1p} \\ & & \ddots & \vdots \\ & & & B_{ppp} \end{pmatrix} \\
&\quad \times \begin{pmatrix} \lambda^q & & & \\ & \lambda^{q+1} & & \\ & & \ddots & \\ & & & \lambda^p \end{pmatrix} \begin{pmatrix} r^q \\ r^{q+1} \\ \vdots \\ r^p \end{pmatrix} \\
&= \begin{pmatrix} B_{pqq} & B_{pqq+1} & \cdots & B_{pqp} \\ & B_{pq+1q+1} & \cdots & B_{pq+1p} \\ & & \ddots & \vdots \\ & & & B_{ppp} \end{pmatrix} \\
&\quad \times \begin{pmatrix} \lambda^q & & & \\ & \lambda^{q+1} & & \\ & & \ddots & \\ & & & \lambda^p \end{pmatrix} \\
&\quad \times \begin{pmatrix} D_{pqq} & D_{pqq+1} & \cdots & D_{pqp} \\ & D_{pq+1q+1} & \cdots & D_{pq+1p} \\ & & \ddots & \vdots \\ & & & D_{ppp} \end{pmatrix} \\
&\quad \times \begin{pmatrix} R_{pq}(r) \\ R_{pq+1}(r) \\ \vdots \\ R_{pp}(r) \end{pmatrix} \\
&= \sum_{k=q}^{p} R_{pk}(r) \sum_{i=q}^{k} r^i \cdot B_{pqi} \cdot D_{pik}.
\end{aligned} \tag{A.2}
$$

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant no. 61261043 and 61102008, College Scientific research project of Ningxia province (no. NGY2012147). The authors would like to thank the anonymous referees for their valuable comments and suggestions.

References

[1] Z. Teng, J. He, A. J. Degnan et al., "Critical mechanical conditions around neovessels in carotid atherosclerotic plaque may promote intraplaque hemorrhage," *Atherosclerosis*, vol. 223, no. 2, pp. 321–326, 2012.

[2] 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.

[3] J. Wood, "Invariant pattern recognition: a review," *Pattern Recognition*, vol. 29, no. 1, pp. 1–17, 1996.

[4] F. Ghorbel, S. Derrode, R. Mezhoud, T. Bannour, and S. Dhahbi, "Image reconstruction from a complete set of similarity invariants extracted from complex moments," *Pattern Recognition Letters*, vol. 27, no. 12, pp. 1361–1369, 2006.

[5] D. G. Sim, H. K. Kim, and R. H. Park, "Invariant texture retrieval using modified Zernike moments," *Image and Vision Computing*, vol. 22, no. 4, pp. 331–342, 2004.

[6] X. Wang, F. X. Guo, B. Xiao, and J. F. Ma, "Rotation invariant analysis and orientation estimation method for texture classification based on Radon transform and correlation analysis," *Journal of Visual Communication and Image Representation*, vol. 21, no. 1, pp. 29–32, 2010.

[7] B. Xiao, J. Ma, and J. T. Cui, "Combined blur, translation, scale and rotation invariant image recognition by Radon and pseudo-Fourier-Mellin transforms," *Pattern Recognition*, vol. 45, no. 1, pp. 314–321, 2012.

[8] H. Q. Zhu, M. Liu, and Y. Li, "The RST invariant digital image watermarking using Radon transforms and complex moments," *Digital Signal Processing*, vol. 20, no. 6, pp. 1612–1628, 2010.

[9] N. C. Rouze, V. C. Soon, and G. D. Hutchins, "On the connection between the Zernike moments and Radon transform of an image," *Pattern Recognition Letters*, vol. 27, no. 6, pp. 636–642, 2006.

[10] T. V. Hoang and S. Tabbone, "Invariant pattern recognition using the RFM descriptor," *Pattern Recognition*, vol. 45, no. 1, pp. 271–284, 2012.

[11] S. R. Deans, *The Radon Transform and Some of Its Applications*, Wiley, New York, NY, USA, 1983.

[12] H. P. Hiriyannaiah and K. R. Ramakrishnan, "Moments estimation in Radon space," *Pattern Recognition Letters*, vol. 15, no. 3, pp. 227–234, 1994.

[13] R. R. Galigekere, D. W. Holdsworth, M. N. S. Swamy, and A. Fenster, "Moment patterns in the Radon space," *Optical Engineering*, vol. 39, no. 4, pp. 1088–1097, 2000.

[14] F. Peyrin and R. Goutte, "Image invariant via the Radon transform," in *Proceedings of the IEEE International Conference on Image Processing and its Applications*, pp. 458–461, 1992.

[15] J. Flusser and T. Suk, "Degraded image analysis: an invariant approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 590–603, 1998.

[16] C. W. Chong, P. Raveendran, and R. Mukundan, "Translation and scale invariants of Legendre moments," *Pattern Recognition*, vol. 37, no. 1, pp. 119–129, 2004.

[17] X. Zhang, Y. Zhang, J. Zhang, X. Li, S. Chen, and D. Chen, "Unsupervised clustering for logo images using singular values region covariance matrices on Lie groups," *Optical Engineering*, vol. 51, no. 4, 8 pages, 2012.

[18] M. K. Hu, "Visual pattern recognition by moments invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[19] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Recognition of shapes by editing their shock graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 550–571, 2004.

[20] http://staff.science.uva.nl/~aloi/ .

[21] H. Zhu, M. Liu, H. Ji, and Y. Li, "Combined invariants to blur and rotation using Zernike moment descriptors," *Pattern Analysis and Applications*, vol. 13, no. 3, pp. 309–319, 2010.

[22] http://museumvictoria.com.au/bioinformatics/butter/images/bthumbliv.htm.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 453402, 8 pages
http://dx.doi.org/10.1155/2013/453402

Research Article

A Novel Automatic Detection System for ECG Arrhythmias Using Maximum Margin Clustering with Immune Evolutionary Algorithm

Bohui Zhu,[1,2] Yongsheng Ding,[1,2] and Kuangrong Hao[1,2]

[1] *College of Information Sciences and Technology, Donghua University, Shanghai 201620, China*
[2] *Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, China*

Correspondence should be addressed to Yongsheng Ding; ysding@dhu.edu.cn

Received 19 January 2013; Revised 1 April 2013; Accepted 2 April 2013

Academic Editor: Shengyong Chen

This paper presents a novel maximum margin clustering method with immune evolution (IEMMC) for automatic diagnosis of electrocardiogram (ECG) arrhythmias. This diagnostic system consists of signal processing, feature extraction, and the IEMMC algorithm for clustering of ECG arrhythmias. First, raw ECG signal is processed by an adaptive ECG filter based on wavelet transforms, and waveform of the ECG signal is detected; then, features are extracted from ECG signal to cluster different types of arrhythmias by the IEMMC algorithm. Three types of performance evaluation indicators are used to assess the effect of the IEMMC method for ECG arrhythmias, such as sensitivity, specificity, and accuracy. Compared with K-means and iterSVR algorithms, the IEMMC algorithm reflects better performance not only in clustering result but also in terms of global search ability and convergence ability, which proves its effectiveness for the detection of ECG arrhythmias.

1. Introduction

Electrocardiogram (ECG) is widely used in cardiology since it consists of effective, simple, noninvasive, low-cost procedures for the diagnosis of cardiovascular diseases (CVDs). Since the state of cardiac heart is generally reflected in the shape of ECG waveform and heart rate, ECG is considered to be a representative signal of cardiac physiology, useful in diagnosing cardiac disorders and detecting any arrhythmia [1, 2].

ECG arrhythmia can be defined as any of a group of conditions in which the electrical activity of the heart is irregular and can cause heartbeat to be slow or fast. It can take place in a healthy heart and be of minimal consequence, but they may also indicate a serious problem that leads to stroke or sudden cardiac death. As ECG signal being nonstationary signal, the arrhythmia may occur at random in the time-scale, which means, the arrhythmia symptoms may not show up all the time but would manifest at certain irregular intervals during the day. Therefore, for effective diagnostics, the variability of ECG signal may have to be observed over several hours. For this reason, together with the fact that the volume of the ECG data is enormous, the study is tedious and time consuming. Thus, automatic and computer-based detection and classification of arrhythmia is critical in clinical cardiology, especially for the treatment of patients in the intensive care unit [1].

In the recent years, several methods have been developed in the literatures for detection and classification of ECG arrhythmias. Artificial neural network (ANN) classification method is one of the main methods for ECG arrhythmia recognition. By integration of many data reduction and feature extraction techniques, such as principal component analysis (PCA), independent component analysis, fuzzy logic, and wavelet transform (WT), improved ANN techniques have been shown to be able to recognize and classify ECG arrhythmia accurately [3–7]. However, many ANN algorithms suffer from slow convergence to local and global minima and from random settings of initial values of weights [7]. Since support vector machine (SVM) classifiers

do not trap in local minima points and need less training input, various methods of SVM have been adopted for ECG signals classification and proved to be effective [8–11].

Although many ECG arrhythmia classification methods show good performance in the laboratory, there are only few techniques gaining popularity in practical applications. One of the main reasons is that most methods are supervised methods which require multiple samples manually labeled with the correct type of ECG signals in context. From these samples, a supervised system can learn to predict the correct sense of the similar ECG signal in a new context. However, these data sets are labor intensive, time consuming, and expensive to produce; thus, few data could be labeled and may be only for several ambiguous types. Therefore, using this technique to detect all kinds of arrhythmias is not optimal in the diagnosis of cardiovascular arrhythmia. Moreover, same state of cardiac heart presents different ECG waveforms for different individual characteristics because of the differences in their body, such as heart volume and coronary artery. Even for the same individual, the waveforms would present different shapes when the sample is involved in different activity states, such as walking, running, and sleeping. In order to address this problem, some methods containing unsupervised techniques are developed to analyze the ECG arrhythmia [4–6, 12–16], which do not need any labeled training sample and can find out unknown ECG arrhythmia. In these methods, the key point is the design of an ideal clustering method, as the accuracy of cluster analysis significantly affects the overall performance.

In this paper, we propose a novel immune evolution maximum margin clustering method (IEMMC) for ECG arrhythmias detection. Specifically, we decompose the ECG arrhythmias diagnosis procedure into three steps, including signal processing, feature extraction, and clustering. First, we apply a wavelet transform based adaptive filter to remove the noise and detect ECG waveform. Then, features are extracted to represent ECG signal. Finally, we employ maximum margin clustering (MMC) method to recognize ECG arrhythmias. Considering huge amount of ECG data and expensive computation of traditional MMC algorithm [17], we propose the IEMMC algorithm as the improvement of the existing MMC and make it more suitable for the detection of ECG abnormalities. Our key contribution is to utilize immune evolutionary algorithm to perform optimization directly on the nonconvex optimization problem formulated by original MMC problem and find the optimal solution which has maximum margin. Our IEMMC method avoids the requirement of solving a nonconvex integer problem and semidefinite programming (SDP) relaxations in the traditional MMC algorithm, which is computationally expensive and time consuming. Due to the outstanding global search ability and robustness of immune evolutionary algorithm, performance of the IEMMC algorithm could maintain at a high level even with a poor quality of random initialization, and the astringency of the IEMMC method is also superior to the existing approaches.

The rest of this paper is organized as follows. Section 2 describes our proposed ECG arrhythmias detection system, including signal preprocessing, feature extraction, and the IEMMC method for ECG arrhythmias. Then, the cluster performance is examined through simulation experiments in Section 3. Finally, the concluding remarks are given in Section 4.

Figure 1: The automatic detection system for ECG arrhythmias.

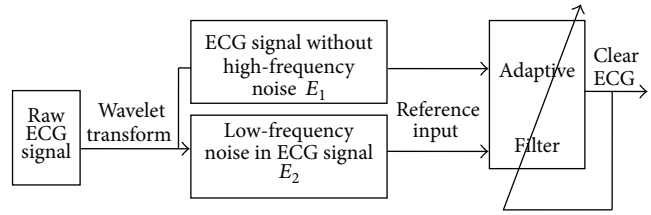


Figure 2: The adaptive ECG filter based on wavelet transforms.

2. A Novel Automatic Detection System for ECG Arrhythmias

The automatic detection system for ECG arrhythmias consists of three stages and is constructed as shown in Figure 1. The first stage is the preprocessing which includes filtering, baseline correction, and waveform detection. The second stage is the feature extraction which aims to find the best coefficients set to describe the ECG signal. The last stage is designed to cluster ECG periods using the IEMMC algorithm according to the previously extracted features in order to construct the arrhythmia classes.

2.1. Preprocessing

2.1.1. ECG Signal Filtering. ECG signals can be contaminated with several types of noise, such as motion artifact (MA), electromyogram noise (EMG), and baseline wandering (BW), which can affect the feature extraction algorithm. So, the ECG samples should be preprocessed before feature extraction and clustering. Due to the frequency spectrum overlapping between ECG signal and noise like motion artifact and baseline wandering which is less than 7 Hz, traditional wavelet decomposition and wavelet threshold method would make ECG waveform distorted, such as the distortion of P wave or T wave signal. For this situation, we apply a wavelet transform based adaptive filter which combines the advantages of wavelet transform and adaptive filtering techniques to preprocess the ECG signal. The construction of our ECG signal filter is demonstrated in Figure 2.

As Figure 2 shows, the procedures of the ECG signal filter can be summarized as the following four steps.

(1) According to the sampling frequency of ECG signal, the least wavelet decomposition level i could be

Table 1: Nine features of ECG signal.

RR_n (s)	RR'_n (s)	QRS_n (s)	PR_n (s)	QT_n (s)	ST_n (s)	R_n (mv)	P_n (mv)	T_n (mv)
0.8477	0.8692	0.0742	0.1663	0.2930	0.2188	1.8149	0.0570	0.6817
0.9023	0.8931	0.0742	0.1445	0.2891	0.2148	1.6339	0.0142	0.5926
0.8594	0.8916	0.0781	0.1406	0.2852	0.2070	2.3085	0.0579	0.6125
0.8281	0.8034	0.0742	0.1663	0.2931	0.2109	2.1007	0.0469	0.6247

determined by separating ECG signal from high-frequency noise. Then, the ECG signal with noise could be wavelet decomposed into i scales.

(2) After wavelet decomposition and removal of precise components containing high-frequency noise signal, we set the approximate components E_1 which contain ECG signal without high-frequency noise as the primary input signal of the adaptive filter.

(3) In line with spectrum relations between various waveform and low-frequency noise, such as baseline drift and motion artifact, the least wavelet decomposition level j which can separate ECG signal from low-frequency noise would be determined. By wavelet decomposition of E_1 into j scales, the left approximate components E_2 containing baseline drift, motion artifact, and other low-frequency interference would be taken as the reference input signal of the adaptive filter.

(4) Least mean squares (LMS) adaptive filtering is used to preprocess the primary input signal and get clear ECG signals.

2.1.2. Waveform Detection. The waveform detection of the ECG signal is the very basis of feature extraction. There are actually three separate algorithms, each of which is designated to detect certain waveform of ECG signal.

(1) *R Detection.* The detection of *QRS* complex takes a vital role in ECG waveform detection. In order to achieve QRS complex detection, *R* wave must be located at first. According to the fact that *R* wave boasts the largest slope, difference of ECG amplitude array is generated to make *R* peaks more noticeable. Then, a practically lower limit is employed to remove unrelated noisy peaks from the signal. In order to avoid interference of big *T* wave, the relative refractory period, which lasts 200 ms after *R* peak is detected, should be skipped. Meanwhile, every *RR* interval should be judged in case of escaped inspection of *R* peak.

(2) *QS Detection.* After finishing the positioning of *R* wave, *Q* and *S* peaks can be identified in accordance with the morphological characteristics. *Q* and *S* peaks occur around the *R* peak within 0.1 second. The turning point connecting baseline and falling edge is just the *Q* peak. Similarly, *S* peak could be found in the right side.

(3) *P and T Wave Detection.* In the light of waveform characteristics of the normal ECG signal, it is found that *P* wave, *QRS* wave, and *T* wave appear alternately. Besides, the gap between the peak of *P* wave and *QRS* is no more than 0.16 seconds. This suggests that the maximum voltage point within 0.16 seconds before the *Q* peak shall be *P* peak, while the maximum voltage point between *S* peak and the next *P* peak shall be the *T* peak.

2.2. Feature Extraction. Feature extraction is a process to determine the best coefficients which could describe the ECG waveform accurately. In order to extract the best features that represent the structure of the ECG signals, nine times domain coefficients belonging to two succeeding ECG periods are considered, as shown in Table 1. The first row in the table is the name of the features, while the rest show the value of each feature. All features are listed as follows:

(a) normalized *RR* interval between the acquired *R* wave and the preceding *R* wave (RR_n);

(b) normalized *RR* interval between the acquired *R* wave and the following *R* wave (RR'_n);

(c) normalized *QRS* interval of the acquired beat (QRS_n);

(d) normalized *PR* interval of the acquired beat (PR_n);

(e) normalized *QT* interval belonging to the acquired beat (QT_n);

(f) normalized *ST* interval of the acquired beat (ST_n);

(g) normalized *R* amplitude of the acquired beat (R_n);

(h) normalized *P* amplitude of the acquired beat (P_n);

(i) normalized *T* amplitude of the acquired beat (T_n).

QRS interval is calculated as the time interval between *Q* wave and *S* wave. *PR* interval is calculated as the time interval between the *P* peak and the *R* peak. *ST* interval is calculated as the time interval between *S* wave and *T* peak. *QT* interval is measured as the time interval between *T* wave and the onset time of the *Q* wave. From the medical point of view, the detection of arrhythmia depends on two or more ECG signal periods. The previous period of an ECG signal has many indicators of current arrhythmia. So, in our approach, two *QRS* periods' parameters RR_n and RR'_n are considered to be the features of ECG signal. *R* amplitude is measured as the distance between the peak of the *R* wave and the baseline. *P* amplitude and *T* amplitude are measured in the same way.

2.3. Clustering Method for ECG Arrhythmia

2.3.1. Maximum Margin Clustering. The MMC extends the theory of SVM to the unsupervised scenario, which aims to find a way to label the samples by running SVM implicitly with the maximum margin over all possible labels [18].

Mathematically, given a point set $\chi = \{x_1, \ldots, x_n\}$ and their labels $y = \{y_1, \ldots, y_n\} \in \{-1, +1\}^n$, SVM seeks a hyperplane $f(x) = w^T\phi(x) + b$ by solving the following optimization problem:

$$\begin{aligned} &\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{s.t.} \quad y_i\left(w^T\phi(x) + b\right) \geq 1 - \xi_i, \\ &\qquad \xi_i \geq 0, \ i = 1, \ldots, n, \end{aligned} \tag{1}$$

where $\phi(\cdot)$ is a nonlinear function that maps the data samples in a high dimensional feature space and makes the nonseparable problem in the original data space to be separable in the feature space. The ξ_i values are called slack variables, and $C > 0$ is a manually chosen constant.

Different from SVM, where the class labels are given and the only variables are the hyperplane parameters (w, b), MMC aims at finding not only the optimal hyperplane (w^*, b^*), but also the optimal labeling vector y [17]. Originally, this task was formulated in terms of the following optimization problem [18]:

$$\begin{aligned} &\min_{y\in\{-1,+1\}^n} \min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{s.t.} \quad y_i\left(w^T\phi(x) + b\right) \geq 1 - \xi_i, \\ &\qquad \xi_i \geq 0, \ i = 1, \ldots, n, \ C \geq 0. \end{aligned} \tag{2}$$

However, the previous optimization problem has a trivially "optimal" solution, which is to assign all data to the same class and obtain an unbounded margin. Moreover, another unwanted solution is to separate a single outlier or a very small group of samples from the rest of the data. To alleviate these trivial solutions, Xu et al. [18] imposed a class balance constraint on y,

$$-\ell \leq e^T y \leq \ell, \tag{3}$$

where $\ell \geq 0$ is a constant to control the class imbalance, which could bound the difference in class size and avoid assigning all patterns to the same class, and e is an all-one vector.

The MMC method often outperforms common clustering methods with respect to the accuracy [17, 18]. It can be expected that the detection of ECG arrhythmia by using the MMC algorithm will achieve a high level of accuracy. However, applying the approach requires solving a nonconvex integer problem, which is computationally expensive, and only small data sets can be handled by the MMC method so far. At present, various optimization techniques have been applied to handle this problem. Xu et al. [18] proposed to make several relaxations to the original MMC problem and reformulate it as a SDP problem, which can then be solved by standard SDP solvers such as SDPT3 and SeDuMi. Valizadegan and Jin [19] further proposed the generalized MMC algorithm which reduces the scale of the original SDP problem significantly. To make MMC method more practical, Zhang et al. [17] put forward a method which iteratively applied an SVM to improve an initial candidate obtained by a K-means preprocessing step. Recently, Zhao et al. [20] proposed a cutting plane MMC method based on constructing a sequence of intermediate tasks and each of the intermediate tasks, was solved using constrained concave-convex procedure. Although the recently proposed approaches have improved the efficiency of the MMC method, the application of these methods has not always been guaranteed. For example, as an iterative approach, the performance of iterSVR algorithm [17] which begins with assigning a set of initial labels is crucial for the quality of initialization. Random initialization will usually result in poor clustering.

2.3.2. Maximum Margin Clustering with Immune Evolution. The concept of SVMs can be considered to be a special case of regularization problems in the following form:

$$\inf_{f\in H} \frac{1}{n}\sum_{i=1}^{n} L\left(y_i, f\left(x_i\right)\right) + \lambda\|f\|_H^2, \tag{4}$$

where $\lambda > 0$ is a fixed real number, $L : Y \times \mathfrak{R} \to [0, \infty)$ is a loss function measuring the performance of the prediction function f on the training set, and $\|f\|_H^2$ is the squared norm in a reproducing kernel Hilbert space $H \subseteq \mathfrak{R}^x = \{f : X \to \mathfrak{R}\}$ induced by a kernel function. In the SVM approach (1), the hinge loss $L_h(y, f) = \max\{0, 1 - yf(x)\}$ with $y \in \{-1, +1\}$ is used. Instead of using the hinge loss, our approach penalizes overconfident predictions by using the square loss $L_s(y, f) = (y - f(x))^2$ leading to

$$\begin{aligned} &\min_{w,b,\eta} \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{n}\eta^2 \\ &\text{s.t.} \quad y_i\left(\left(w^T\phi\left(x_i\right)\right) + b\right) = 1 - \eta, \quad i = 1, \ldots, n. \end{aligned} \tag{5}$$

So, in our MMC algorithm, we aim at finding a solution for

$$\begin{aligned} &\min_{y\in\{-1,+1\}^n, w, b} J(y, w, b) = \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{n}\eta^2 \\ &\text{s.t.} \quad y_i\left(\left(w^T\phi\left(x_i\right)\right) + b\right) = 1 - \eta, \\ &\qquad i = 1, \ldots, n, \ -l \leq \sum_{i=1}^{n} y_i \leq l. \end{aligned} \tag{6}$$

In order to solve problem (6), the original non-convex problem is considered to be a special case of optimization problem, and immune evolutionary algorithm is proposed to find optimal solution. Recent studies have shown that the immune evolutionary algorithm possesses several attractive immune properties that allow evolutionary algorithms to avoid premature convergence and improve local search capability [21–25]. By utilizing powerful global search capability and fast convergence of the immune evolutionary algorithm,

IEMMC could avoid SDP relaxations and find optimal solution of the MMC method efficiently.

The Process of IEMMC Algorithm. The framework of our IEMMC algorithm is given by Algorithm 1.

Algorithm 1 (Maximum Margin Clustering with Immune Evolution).

Step 1. Generate a set of candidate solutions $P = \{y_1, \ldots, y_{m+r}\} \subseteq \{-1, +1\}^n$, composed of the subset of memory cells P_m added to the remaining P_r ($P = P_m + P_r$). P should fulfill the balance constraint (3) and $\|y_i - y_j\| > t_s$, t_s is the suppression threshold.

Step 2. Compute the affinity values $F(y)$ for each $y_j \in P$.

Step 3. Determine the N_c best individuals, P_c of the population P_r, based on an affinity measure. Perform clone selection on the population P_c to generate a temporary population of clones P_c^*.

Step 4. Determine the N_m best individuals, P_m of the remaining population $P_r - P_c$, based on an affinity measure. Apply mutation to the antibodies population P_m, where the hypermutation is proportional to affinity of the antibody. A maturated antibody population P_m^* is generated.

Step 5. Re-select the improved individuals from P_c^* and P_m^* to compose the memory set and the population P_r.

Step 6. Perform receptor editing, replace some low affinity antibodies of the population P_r by randomly created new antibodies, maintaining its diversity.

Step 7. If termination conditions are not satisfied, go to Step 2.

Step 8. Return the best individual y_i.

The starting point is generating a set of candidate solutions $P = \{y_1, \ldots, y_{m+r}\} \subseteq \{-1, +1\}^n$, composed of the subset of memory cells P_m added to the remaining P_r ($P = P_m + P_r$). Each of these individuals constitutes a possible solution for optimization problem (6). Throughout our IEMMC algorithm, we ensure that only valid individuals are created; that is, individuals y should fulfill the balance constraint (3). In Step 2, the affinity value $F(y)$ is computed for each of the initial individuals, where

$$F(y) = \exp\left(-\min J(y, w, b)\right). \tag{7}$$

Depending on the affinity values, the copies of the antibodies are generated, and clone selection is performed on superior individuals. In Step 4, mutation process is applied to the antibodies. If the affinity value of the new antibody is better than that of original value, new antibody is stored in the place of the original one; otherwise, old antibody is kept in population. After the mutation process, receptor editing is applied to the antibody population. In the receptor editing process, a percentage of antibodies in the antibody population are replaced by randomly created new antibodies. When the best individual satisfies termination condition, y_i would be returned.

Fitness Computation. For fixed solution y, the problem formulated in the function (6) could be solved by the standard SVM learning algorithm. So, we can compute (w, b) from the Karush-Kuhn-Tucker (KKT) conditions as usual to maximize margin between clusters. But this solution (w, b, y) is not the optimal clustering solutions for problem (6). Therefore, we continue to find a better bias b and cluster label y by fixing w and minimizing problem (6) which is reduced to

$$\begin{aligned} &\min_{y,b} \sum_{i=1}^{n} \left(w \cdot \phi(x_i) + b - y_i\right)^2 \\ &\text{s.t.} \quad y_i \in \{\pm 1\}, \\ &\quad i = 1, \ldots, n, \ -\ell \le e^T y \le \ell. \end{aligned} \tag{8}$$

Then, problem (8) can be solved without the use of any optimization solver by the following proposition. At first, we sort $w^T\phi(x_i)$ and use the set of midpoints between any two consecutive $w^T\phi(x_i)$ values as the candidates of b. From these candidates of b, the first $(n - l)/2$ and the last $(n - l)/2$ of the candidates should be removed for not satisfying the class balance constraint (3). For each remaining candidate, we determine $y = \text{sign}(w^T\varphi(x) + b)$ and compute the corresponding objective value in (8). Finally, we choose b and corresponding y that has the optimal objective. Since both w and b have been determined, fitness value $F(y)$ for the new individual y can be obtained by $F(y) = \exp(-\min J(y, w, b))$.

3. Experiment and Results

3.1. Experimental Data. Experimental data of ECG arrhythmias used in this study are taken from MIT-BIH ECG Arrhythmias Database [26]. All ECG data are classified into five classes according to standard of The Association for the Advancement of Medical Instrumentation (AAMI) [27], since this database urges all users to follow the AAMI recommendations. In this standard, abnormal ECG could be divided into following four types. Type S contains atrial premature (AP), nodal premature (NP), and supraventricular premature (SP). Type V contains premature ventricular contraction (PVC) and ventricular ectopic (VE). Type F contains fusion of ventricular and normal beat. Type Q contains paced beat, fusion of paced and normal beat, and unclassified beat. The other kinds of heartbeats are considered as N type, including normal beat, atrial escape (AE), nodal escape (NE), right bundle branch block (R), and left bundle branch block (L).

Totally 1682 ECG periods are selected from seven records of MIT/BIH database to test the correctness of the IEMMC algorithm. The distribution of records is shown in Table 2. The first row corresponds to the labels according to the AAMI standard. And the first column is the name of the records, whereas the others contain the number of heartbeats of each type.

Table 2: The number of sample records according to arrhythmia type.

MIT code	N	S	V	F	Q	Total
106	104	0	83	0	0	187
200	125	0	112	0	0	237
208	95	0	0	86	0	181
209	102	106	0	0	0	208
213	106	0	0	113	0	219
217	205	0	0	0	211	416
222	122	112	0		0	234
Total	859	218	195	199	211	1682

Table 3: The ECG arrhythmias clustering results using the IEMMC algorithm.

	Clustering result				
Arrhythmia type	N	S	F	V	Q
N	803	15	12	13	16
S	27	191	0	0	0
V	35	0	164	0	0
F	17	0	0	178	0
Q	28	0	0	0	183

3.2. Experimental Results. In this section, we demonstrate the superiority of the proposed IEMMC procedure for ECG arrhythmias detection, and the following three types of performance evaluation indicators are used to assess the effect of ECG arrhythmias clustering method:

$$\text{sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})},$$
$$\text{specificity} = \frac{\text{TN}}{(\text{FP} + \text{TN})}, \quad (9)$$
$$\text{accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FN} + \text{FP} + \text{TN})},$$

where true positive (TP) means the number of true arrhythmia that has been successfully detected; false positive (FP) is the number of true arrhythmia that has been missed; true negative (TN) means the number of corresponding nontarget arrhythmia that has been correctly detected; false negative (FN) is the count of nontarget arrhythmia that has been detected wrongly.

The simulation results are listed in Table 3, and the performance analysis of the clustering result is in Table 4. As shown in Tables 3 and 4, by using the IEMMC algorithm, the correctness of ECG arrhythmias is at a high level.

From the result, we can find that type N is the most regular and numerous heartbeats and easy to be separated from the other types; so, its result is better than other types. However, the performance of type F is lower than that in the previous case. Given that morphology of type F is often very similar to that of other types, it is very difficult to characterize type F.

In order to verify and measure the IEMMC algorithm's superiority, three methods are developed in parallel to compare with our algorithm, including standard K-means algorithm, iterSVR which is the first approach capable of dealing with large data sets [17], and SVM which has been proved to be a successful supervised learning method for ECG recognition and classification [8–11]. The performance of all clustering methods is shown in Figure 3. Two initialization schemes are developed for both iterSVR and IEMMC in the experiment: (1) random; (2) standard K-means clustering (KM). In the first scheme, initial candidate solutions of IEMMC and iterSVR are generated randomly. In the second scheme, iterSVR is initialized by standard K-means clustering. Only one of IEMMC candidate solutions is initialized by standard K-means clustering, and the rest solutions are generated at random. The class balance parameter of both IEMMC and iterSVR is always set as $L = 0.2 * n$. Also, 20% of the ECG data are extracted randomly to be the training data of the SVM classification. The radical basis function (RBF) kernel $k(x, x') = \exp(-\|x - x'\|/\sigma^2)$ is used for all the kernel methods in the experiment. As for the regularization parameter C, we choose the best value from a set of candidates (1, 10, 100, 500) for each data set. All algorithms are, respectively, repeated three times because of the inherent randomness. For each method and each data set, we report the result with its best value chosen from a set of candidates.

Table 4: The performance analysis result of the ECG arrhythmias clustering method.

Arrhythmia type	Sensitivity (%)	Specificity (%)	Accuracy (%)
N	97.9	92.7	95.4
S	83.0	98.0	95.8
F	82.4	97.5	95.6
V	82.8	98.7	96.6
Q	83.9	97.9	96.0
Total	90.3	97.4	95.9

From Figure 3, the IEMMC's performance is as similar as that of the SVM and better than those of all clustering methods. Also, we can find that the performance of iterSVR largely depends on the superiority of initialization. With random initialization, clustering result from iterSVR is even worse than that of K-means algorithm. Since the performance of K-means is also unsatisfactory, even initialized by K-means, iterSVR still cannot meet the expectation of the ECG arrhythmia diagnosis. However, inheriting the outstanding global optimization ability of immune evolutionary algorithm, the IEMMC algorithm can find the best clustering for objective function in a very short evolution period, even in the case of random initialization. Additionally, IEMMC algorithm not only excelled in performance but also in convergence. While iterSVR needs to iterate ten times to find solution, the IEMMC algorithm only needs to evolve four generations. Especially, the IEMMC algorithm could obtain the same optimal solution from different initializations in few generations of evolutions, due to the prominent convergence and global search ability. This excellent performance in the

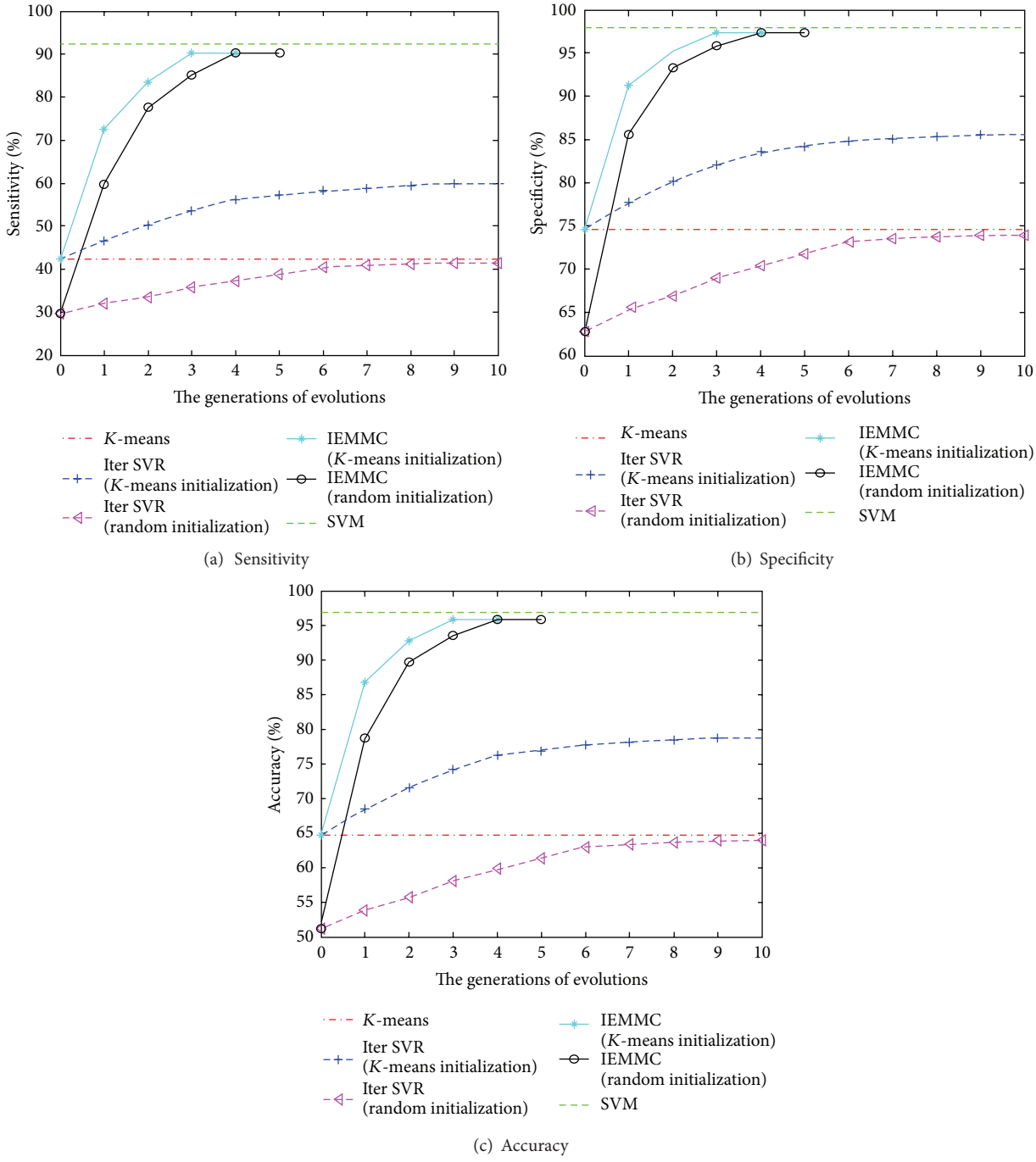


Figure 3: The performance comparison of different clustering methods.

experiment has proved that the IEMMC algorithm is very effective for the detection of ECG arrhythmia.

4. Conclusions

In this paper, a novel IEMMC algorithm is proposed to cluster the ECG signal and detect ECG arrhythmias, which iteratively updates the quality of candidates by means of immune evolutionary without employing any training process. The experimental analysis reveals that our approach yields better clustering performance than some competitive methods in most cases.

In the future, we will use some other biological principles based evolutionary algorithm to solve the MMC problem, like ant colony optimization and particle swarm optimizer, since they have been proved to have global optimizaton ability. Furthermore, comparison with immune evolutionary algorithm will be done to find out a more efficient ECG data clustering algorithm.

Acknowledgments

This work was supported in part by the Key Project of the National Nature Science Foundation of China (no. 61134009), Specialized Research Fund for Shanghai Leading Talents, Project of the Shanghai Committee of Science and Technology (nos. 11XD1400100 and 11JC1400200), and the Fundamental Research Funds for the Central Universities.

References

[1] U. R. Acharya, P. S. Bhat, S. S. Iyengar, A. Rao, and S. Dua, "Classification of heart rate data using artificial neural network and fuzzy equivalence relation," *Pattern Recognition*, vol. 36, no. 1, pp. 61–68, 2003.

[2] S. Osowski and T. H. Linh, "ECG beat recognition using fuzzy hybrid neural network," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1265–1271, 2001.

[3] S. N. Yu and K. T. Chou, "Integration of independent component analysis and neural networks for ECG beat classification," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2841–2846, 2008.

[4] R. Ceylan and Y. Özbay, "Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network," *Expert Systems with Applications*, vol. 33, no. 2, pp. 286–295, 2007.

[5] R. Ceylan, Y. Özbay, and B. Karlik, "A novel approach for classification of ECG arrhythmias: type-2 fuzzy clustering neural network," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6721–6726, 2009.

[6] Y. Özbay, R. Ceylan, and B. Karlik, "A fuzzy clustering neural network architecture for classification of ECG arrhythmias," *Computers in Biology and Medicine*, vol. 36, no. 4, pp. 376–388, 2006.

[7] A. De Gaetanoa, S. Panunzia, F. Rinaldia, A. Risia, and M. Sciandroneb, "A patient adaptable ECG beat classifier based on neural networks," *Applied Mathematics and Computation*, vol. 213, pp. 243–249, 2009.

[8] B. M. Asl, S. K. Setarehdan, and M. Mohebbi, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artificial Intelligence in Medicine*, vol. 44, no. 1, pp. 51–64, 2008.

[9] K. Polat, B. Akdemir, and S. Güneş, "Computer aided diagnosis of ECG data on the least square support vector machine," *Digital Signal Processing*, vol. 18, no. 1, pp. 25–32, 2008.

[10] K. Polat and S. Güneş, "Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine," *Applied Mathematics and Computation*, vol. 186, no. 1, pp. 898–906, 2007.

[11] M. Moavenian and H. Khorrami, "A qualitative comparison of artificial neural Networks and support vector machines in ECG arrhythmias classification," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3088–3093, 2010.

[12] M. Korürek and A. Nizam, "A new arrhythmia clustering technique based on ant colony optimization," *Journal of Biomedical Informatics*, vol. 41, no. 6, pp. 874–881, 2008.

[13] M. Korürek and A. Nizam, "Clustering MIT-BIH arrhythmias with ant colony optimization using time domain and PCA compressed wavelet coefficients," *Digital Signal Processing*, vol. 20, no. 4, pp. 1050–1060, 2010.

[14] G. Zheng and T. Yu, "Study of hybrid strategy for ambulatory ECG waveform clustering," *Journal of Software*, vol. 6, no. 7, pp. 1257–1264, 2011.

[15] F. Sufi, I. Khalil, and A. N. Mahmood, "A clustering based system for instant detection of cardiac abnormalities from compressed ECG," *Expert Systems with Applications*, vol. 38, no. 5, pp. 4705–4713, 2011.

[16] B. Doğan and M. Korürek, "A new ECG beat clustering method based on kernelized fuzzy c-means and hybrid ant colony optimization for continuous domains," *Applied Soft Computing*, vol. 12, pp. 3442–3451, 2012.

[17] K. Zhang, I. W. Tsang, and J. T. Kwok, "Maximum margin clustering made practical," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 583–596, 2009.

[18] L. Xu, J. Neufeld, B. Larson, and D. Schuurmans, "Maximum margin clustering," *Advances in Neural Information Processing Systems*, vol. 17, pp. 1537–1544, 2005.

[19] H. Valizadegan and R. Jin, "Generalized maximum margin clustering and unsupervised kernel learning," *Advances in Neural Information Processing Systems*, vol. 19, pp. 1417–1424, 2007.

[20] F. Wang, B. Zhao, and C. Zhang, "Linear time maximum margin clustering," *IEEE Transactions on Neural Networks*, vol. 21, no. 2, pp. 319–332, 2010.

[21] Y.-S. Ding, Z.-H. Hu, and W.-B. Zhang, "Multi-criteria decision making approach based on immune co-evolutionary algorithm with application to garment matching problem," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10377–10383, 2011.

[22] Y.-S. Ding, X.-J. Lu, K.-R. Hao, L.-F. Li, and Y. F. Hu, "Target coverage optimisation of wireless sensor networks using a multi-objective immune co-evolutionary algorithm," *International Journal of Systems Science*, vol. 42, no. 9, pp. 1531–1541, 2011.

[23] L.-J. Cheng, Y.-S. Ding, K.-R. Hao, and Y.-F. Hu, "An ensemble kernel classifier with immune clonal selection algorithm for automatic discriminant of primary open-angle glaucoma," *Neurocomputing*, vol. 83, pp. 1–11, 2012.

[24] J. T. Tsai, W. H. Ho, T. K. Liu, and J. H. Chou, "Improved immune algorithm for global numerical optimization and job-shop scheduling problems," *Applied Mathematics and Computation*, vol. 194, no. 2, pp. 406–424, 2007.

[25] J. Gao and J. Wang, "A hybrid quantum-inspired immune algorithm for multiobjective optimization," *Applied Mathematics and Computation*, vol. 217, no. 9, pp. 4754–4770, 2011.

[26] "MIT-BIH arrhythmia database," http://physionet.org/physiobank/database/mitdb/.

[27] *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*, Association for the Advancement of Medical Instrumentation, 1998.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 628036, 11 pages
http://dx.doi.org/10.1155/2013/628036

Research Article

Structural Complexity of DNA Sequence

Cheng-Yuan Liou, Shen-Han Tseng, Wei-Chen Cheng, and Huai-Ying Tsai

Department of Computer Science and Information Engineering, National Taiwan University, Taipei 10617, Taiwan

Correspondence should be addressed to Cheng-Yuan Liou; cyliou@csie.ntu.edu.tw

Received 10 January 2013; Accepted 3 March 2013

Academic Editor: Carlo Cattani

In modern bioinformatics, finding an efficient way to allocate sequence fragments with biological functions is an important issue. This paper presents a structural approach based on context-free grammars extracted from original DNA or protein sequences. This approach is radically different from all those statistical methods. Furthermore, this approach is compared with a topological entropy-based method for consistency and difference of the complexity results.

1. Introduction

DNA sequence analysis becomes important part in modern molecular biology. DNA sequence is composed of four nucleotide bases—adenine (abbreviated A), cytosine (C), guanine (G), and thymine (T) in any order. With four different nucleotides, 2 nucleotides could only code for maximum of 4^2 amino acids, but 3 nucleotides could code for a maximum 4^3 amino acids. George Gamow was the first person to postulate that every three bases can translate to a single amino acid, called a codon. Marshall Nirenberg and Heinrich J. Matthaei were the first to elucidate the nature of a genetic code. A short DNA sequence can contain less genetic information, while lots of bases may contain much more genetic information, and any two nucleotides switch place may change the meaning of genetic messages.

Sequence arrangement can produce many different results, but only few codons exist in living bodies. Some sequences do not contain any information which is known as junk DNA. Finding an efficient way to analyze a sequence fragment corresponding to genetic functions is also a challenging problem.

In recent papers, methods broadly fall into two categories, sequence complexity [1, 2] and structural pattern analysis [3–8]. Koslicki [1] presented a method for computing sequence complexities. He redefined topological entropy function so that the complexity value will not converge toward zero for much longer sequences. With separate sequence into several segments, it can determine the segments where are exons or introns, and meaningful or meaningless. Hao et al. [7] given a graphical representation of DNA sequence, according to this paper, we can find some rare occurred subsequences. R. Zhang and C. T. Zhang [4] used four-nucleotide-related function drawing 3D curves graph to analyze the number of four-nucleotide occurrence probabilities. Liou et al. [9] had given a new idea in modeling complexity for music rhythms; this paper translated text messages into computable values, so computers can score for music rhythms.

In this paper, we propose a new method for calculating sequences different from other traditional methods. It holds not only statistical values but also structural information. We replace four nucleotides with tree structure presented in [9] and use mathematical tools to calculate complexity values of the sequences. So we can compare two sequences with values and determine dissimilarity between these two sequences. In biomedical section, we can use this technique to find the effective drugs for new virus with priority.

2. DNA Sequence Represented with Tree Structure

Our method uses Lindenmayer system [10–12] property among calculated complexities from tree structure [9]; it is a different way of computing complexities of sequences. At first, we introduce *DNA tree* and convert DNA sequence to

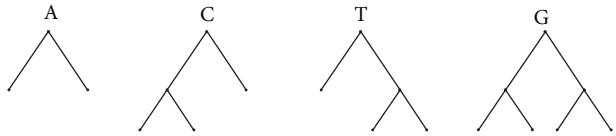


Figure 1: Nucleotide bases corresponding trees.

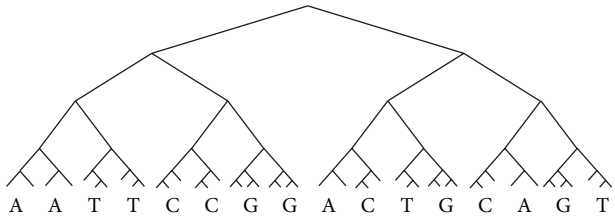


Figure 2: DNA sequence represented with tree structure.

tree structure. A DNA tree is a binary tree of which each subtree is also a DNA tree. Every tree node is either a terminal node or a node with two childrens (branches or descendants).

Lindenmayer system is a powerful rewriting system used to model the growth processes of plant development. We will introduce it in Section 2.2 in detail. Lindenmayer system uses some initial and rewriting rules to construct beautiful graphs. Since it can construct a tree from rewriting rules, it also can extract rewriting rules from a tree. In this section, we will use tools to generate the rules from tree.

We use 4 fixed *tree representations* for nucleotide bases A, T, C, and G (see Figure 1). When we apply this method to amino acid sequence, we can construct more tree representation for amino acids, respectively.

When we transfer a sequence to DNA tree, we will replace every word to tree elements step by step, and two consecutive trees can combine to a bigger tree. Following the previous steps, a DNA sequence will be transfer to a DNA tree (see Figure 2).

2.1. Bracketed Strings for a DNA Sequence. For computing complexity of our DNA tree, we need some rules for converting tree to another structure. We use a stack similarly structure to represent the hierarchy of DNA tree, called *bracketed string*. DNA tree can transfer to a unique bracketed string by the following symbols, and it can transfer back to the original tree:

(i) F: the current location of tree nodes; it can be replaced by any word or be omitted;

(ii) +: the following string will express the right subtree;

(iii) −: the following string will express the left subtree;

(iv) [: this symbol is pairing with]; "[···]" denotes a subtree where "···"; indicates all the bracketed strings of its subtree;

(v)]: see [description.

Following the previous symbols, Figure 3 shows that nucleotide base A and T represented tree can transfer to $[F[-F][+F]]$ and $[F[-F][+F[-F][+F]]]$, respectively.

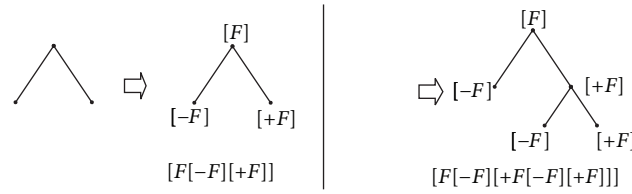


Figure 3: Bracketed strings representation for two trees.

And Figure 4 is the bracketed string of Figure 2. We can see that when the tree grows, string seems to be more redundant. Since we focus here only on DNA trees, we can simplify the bracketed string representations. First, our trees have only two subtrees. Second, the "F" notation for the tree is trivial. With these two characteristics, we may omit the "F" notation from the bracketed string and use only four symbols, {[,], −, +}, to represent trees. In our cases, "[···]" denotes a subtree where "···" indicates all the bracketed strings of its subtrees. "−" indicated the next "[···]" notation for a tree is a left subtree of current node, and "+" is a right subtree vice versa. Figure 5 is the simplified string of bracketed string shown in Figure 4.

2.2. DNA Sequence Represented with L-System. When we obtain DNA tree and bracketed string representation, we need rewriting rules for analyzing tree structure. There are some types of rewriting mechanism such as Chomsky grammar and Lindenmayer system (*L-system* for short). The largest difference between two string rewriting mechanisms lies in the technique used to apply productions. Chomsky grammar is suitable for applying productions sequentially, while L-system is for parallel. In our structure, applying L-system to our representations is better than Chomsky grammar.

The L-system was introduced by the biologist Lindenmayer in 1968 [13]. The central concept of the L-system is rewriting. In general, rewriting is a technique used to define complex objects by successively replacing parts of a simple initial object, using a set of rewriting rules or productions. In the next section, we will present how we use L-system to our DNA tree. The L-system is defined as follows.

Definition 1. L-system grammars are very similar to the Chomsky grammar, defined as a tuple [14]:

$$G = (V, \omega, P), \tag{1}$$

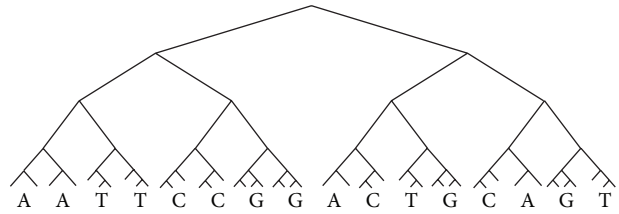
where

(i) $V = \{s_1, s_2, \ldots, s_n\}$ is an alphabet,

(ii) ω (start, axiom, or initiator) is a string of symbols from V defining the initial state of the system,

(iii) P is defined by a production map $P : V \rightarrow V^*$ with $s \rightarrow P(s)$ for each s in V. The identity production $s \rightarrow s$ is assumed. These symbols are called constants or terminals.

2.3. Rewriting Rules for DNA Sequences. As discussed earlier, we want to generate the rules from DNA trees. In this section,



[[−F[−F[−F[−F[−F][+F]][+F[−F][+F]]][+F[−F[−F][+F[−F][+F]]][+F[−F][+F[−F][+F]]]]][+F[−F[−F[−F[−F][+F]]
[+F]][+F[−F[−F][+F]][+F]]][+F[−F[−F[−F][+F]][+F[−F][+F]]][+F[−F[−F][+F]][+F[−F][+F]]]]]][+F[−F[−F[−F
[−F][+F]][+F[−F[−F][+F]][+F]]][+F[−F[−F][+F[−F][+F]]][+F[−F[−F][+F]][+F[−F][+F]]]]][+F[−F[−F[−F[−F]
[+F]][+F]][+F[−F][+F]]][+F[−F[−F[−F][+F]][+F[−F][+F]]][+F[−F][+F[−F][+F]]]]]]]

Figure 4: Bracketed strings representation for Figure 2.

[−[−[−[−[−+]+[−+]]+[−[−+[−+]]+[−+[−+]]]]+[−[−[−[−+]+]+[−[−+]+]]+[−[−[−+]+[−+]]+[−[−+]+[−+]]]]]
+[−[−[−[−+]+[−[−+]+]]+[−[−+[−+]]+[−[−+]+[−+]]]]+[−[−[−[−+]+]+[−+]]+[−[−[−+]+[−+]]+[−+[−+]]]]]]

Figure 5: More simply bracketed strings representation for Figure 2.

we will explain how we apply rewriting rules to those trees. We can apply distinct variables to each node. Since the technique described previously always generates two subtrees for each node, for every nonterminal node, they always can be explained in the following format:

$$P \longrightarrow LR, \tag{2}$$

where P denotes the current node, L denotes its left subtree, and R denotes its right subtree, respectively. We give an example shown in Figure 6; left tree has three nodes and only root is nonterminal node, it can be rewritten as $P \rightarrow LR$. Right tree has five nodes, root P with left subtree L and right subtree R. Left subtree is terminal, but right is not. R has two terminal subtrees R_L and R_R, so this tree can be rewritten as $P \rightarrow LR$ and $R \rightarrow R_L R_R$.

2.4. Rewriting Rules for Bracketed Strings. Similarly, we can also use rewriting rules to generate bracketed strings. In rewriting rules for DNA trees shown in Section 2.3, we write $P \rightarrow LR$ for a tree with left and right subtrees. Note that we call L and R as the nonterminals. In this section, terminal nodes will be separated from trees, and we use "null" to represent a terminal. Such tree will have a corresponding bracketed string as follows: $[[-F\cdots][+F\cdots]]$. "$[-F\cdots]$" represents the left subtree, while "$[+F\cdots]$" represents the right subtree. Therefore, we can replace the rewriting rules with

$$\begin{aligned} P &\longrightarrow [-FL]\,[+FR]\,, \\ F &\longrightarrow \cdots, \\ R &\longrightarrow \cdots, \end{aligned} \tag{3}$$

where "$\cdots$" is the rewriting rule for the bracketed string of each subtree. For the sake of readability, we replace the words such as "R_{R_L}" and "R_{R_R}". In Figure 7, we show the rewriting rules for the bracketed string of the tree in Figure 3.

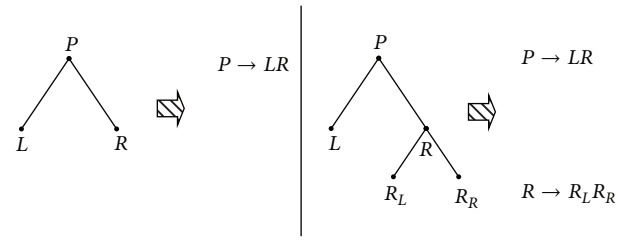


Figure 6: Example of rewriting rules for trees.

As we can see, there are "nulls" in the rules. Those "nulls" do not have significant effects to our algorithm, so we simply ignore the nulls. Now, Figure 3 can apply new rewriting rules without trivial nulls as Figure 8.

When tree grows up, the rewriting rules may generate identical rules. Assume that we have the following rules:

$$\begin{aligned} P &\longrightarrow [-FT_L]\,[+FT_R]\,, \\ T_L &\longrightarrow [-F]\,[+F]\,, \\ T_R &\longrightarrow [-F]\left[+FT_{R_R}\right], \\ T_{R_R} &\longrightarrow [-F]\left[+FT_{R_{R_R}}\right], \\ T_{R_{R_R}} &\longrightarrow [-F]\,. \end{aligned} \tag{4}$$

These rules can generate exactly one bracketed string and, thus, exactly one DNA tree. All these rules form a rule set that represents a unique DNA tree. When we look at $T_R \rightarrow [-F][+FT_{R_R}]$ and $T_{R_R} \rightarrow [-F][+FT_{R_{R_R}}]$, they have the same structure since they both have a right subtree and do not have a left subtree. The only difference is that one of the subtrees is T_{R_R} and that the other is $T_{R_{R_R}}$. We will define two terms to

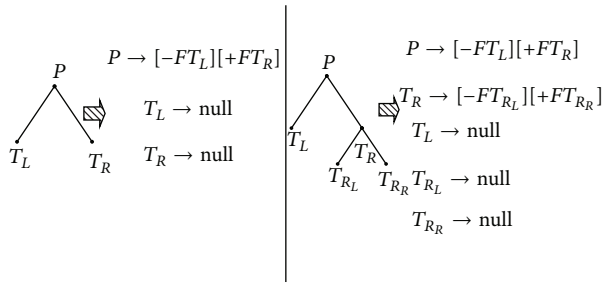


Figure 7: Rewriting rules for the bracketed string of trees.

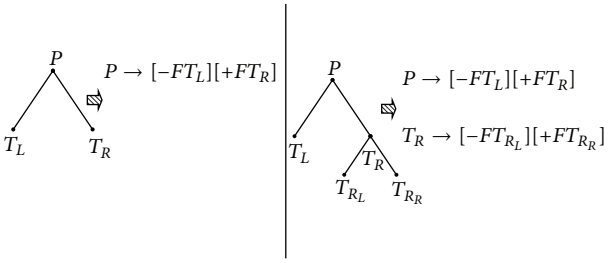


Figure 8: Rewriting rules for the bracketed string without nulls of trees.

express the similarity between two rewriting rules, and these terms can simplify complexity analysis.

2.5. Homomorphism and Isomorphism of Rewriting Rules. At the end of the previous section, we discussed that $T_R \to [-F][+FT_{R_R}]$ and $T_{R_R} \to [-F][+FT_{R_{R_R}}]$ are almost the same. How can we summarize or organize an effective feature to them? Liou et al. [9] gave two definitions to classify similar rewriting rules described before as follows.

Definition 2. Homomorphism in rewriting rules. We define that rewriting rule R_1 and rewriting rule R_2 are homomorphic to each other if and only if they have the same structure.

In detail, rewriting rule R_1 and rewriting rule R_2 in DNA trees both have subtrees in corresponding positions or both not. Ignoring all nonterminals, if rule R_1 and rule R_2 generate the same bracketed string, then they are homomorphic by definition.

Definition 3. Isomorphism on level X in rewriting rules. Rewriting rule R_1 and rewriting rule R_2 are isomorphic on depth X if they are homomorphic and their nonterminals are relatively isomorphic on depth $X - 1$. Isomorphic on level 0 indicates homomorphism.

Applying to the bracketed string, we ignore all nonterminals in (4) as follows:

$$P \longrightarrow [-FT_L][+FT_R] \longrightarrow [-F][+F],$$
$$T_L \longrightarrow [-F][+F] \longrightarrow [-F][+F],$$
$$T_R \longrightarrow [-F]\left[+FT_{R_R}\right] \longrightarrow [-F][+F],$$
$$T_{R_R} \longrightarrow [-F]\left[+FT_{R_{R_R}}\right] \longrightarrow [-F][+F],$$
$$T_{R_{R_R}} \longrightarrow [-F] \longrightarrow [-F]. \tag{5}$$

We find that P, T_L, T_R, and T_{R_R} are homomorphic to each other; they generate the same bracketed string, $[-F][+F]$. But $T_{R_{R_R}}$ is not homomorphic to any of the other rules; its bracketed string is $[-F]$.

Let us recall DNA tree example in Figure 2; we will use this figure as an example to clarify these definitions. Now we marked some nodes shown in Figure 9; there are tree rooted at A, B, C, and D, respectively, tree A, tree B, tree C, and tree D. Tree A is isomorphic to tree C on depth 0 to 3, but they are not isomorphic on depth 4. Tree B is isomorphic to tree C on depth from 0 to 2, but they are not isomorphic on depth 3. D is not isomorphic to any other trees, nor is it homomorphic to any other trees.

After we define the similarity between rules by homomorphism and isomorphism, we can classify all the rules into different subsets, and every subset has the same similarity relation. Now we list all the rewriting rules of Figure 2 into Table 1 but ignore terminal rules such as "$\to$ null" and transfer rule's name to class name (or class number). For example, we can give terminal rewriting rule a class, "$C_3 \to$ null", and a rule link to two terminals; we can give them "$C_2 \to C_3C_3$"; here C_3 is the terminal class. After performing classification, we obtain not only a new rewriting rule set but also a context-free grammar, which can be converted to automata.

In Table 1, rules such as $T_{R_{L_{L_L}}} \to [-F][+F]$, and $T_{R_{R_{R_{L_L}}}} \to [-F][+F]$ and $T_{R_{L_{R_{L_R}}}} \to [-F][+F]$ are isomorphic on depth 1 and assigned to Class 4. There are twenty such rules before classification, so we write "$(20)C_4 \to [-F][+F]$". Similar rules such as $P \to [-FT_L][+FT_R]$, $T_{R_{L_{L_L}}} \to [-F][+F]$, and $T_{R_{R_{R_R}}} \to [-F][+FT_{R_{R_{R_{R_R}}}}]$ are isomorphic on depth 0, and there are 47 such rules. They are all assigned to Class 1 by following a similar classification procedure. The classification of the all rules is listed in Table 2. Note that this section also presents a new way to convert a context-sensitive grammar to a context-free one.

3. DNA Sequence Complexity

When we transfer the DNA sequence to the rewriting rules, and classify all those rules we attempt to explore the redundancy in the tree that will be the base for building the cognitive map [15]. We compute the complexity of the tree which those classified rules represent. We know that a classified rewriting rule set is also a context-free grammar, so there are some methods for computing complexity of rewriting rule as follows.

Definition 4. Topological entropy of a context-free grammar. The topological entropy K_0 of (context-free grammar) CFG can be evaluated by means of the following three procedures [16, 17].

Table 1: Rewriting rules for the DNA tree in Figure 2.

$$P \rightarrow [-FT_L][+FT_R]$$
$$T_L \rightarrow \left[-FT_{L_L}\right]\left[+FT_{L_R}\right]$$
$$T_{L_L} \rightarrow \left[-FT_{L_{L_L}}\right]\left[+FT_{L_{L_R}}\right]$$
$$T_{L_{L_L}} \rightarrow [-F][+F]$$
$$T_{L_{L_R}} \rightarrow [-F][+F]$$
$$T_{L_{L_R}} \rightarrow \left[-FT_{L_{L_{R_L}}}\right]\left[+FT_{L_{L_{R_R}}}\right]$$
$$T_{L_{L_{R_L}}} \rightarrow [-F]\left[+FT_{L_{L_{R_{L_R}}}}\right]$$
$$T_{L_{L_{R_{L_R}}}} \rightarrow [-F][+F]$$
$$T_{L_{L_{R_R}}} \rightarrow [-F]\left[+FT_{L_{L_{R_{R_R}}}}\right]$$
$$T_{L_{L_{R_{R_R}}}} \rightarrow [-F][+F]$$
$$T_{L_R} \rightarrow \left[-FT_{L_{R_L}}\right]\left[+FT_{L_{R_R}}\right]$$
$$T_{L_{R_L}} \rightarrow \left[-FT_{L_{R_{L_L}}}\right]\left[+FT_{L_{R_{L_R}}}\right]$$
$$T_{L_{R_{L_L}}} \rightarrow \left[-FT_{L_{R_{L_{L_L}}}}\right][+F]$$
$$T_{L_{R_{L_{L_L}}}} \rightarrow [-F][+F]$$
$$T_{L_{R_{L_R}}} \rightarrow \left[-FT_{L_{R_{L_{R_L}}}}\right][+F]$$
$$T_{L_{R_{L_{R_L}}}} \rightarrow [-F][+F]$$
$$T_{L_{R_R}} \rightarrow \left[-FT_{L_{R_{R_L}}}\right]\left[+FT_{L_{R_{R_R}}}\right]$$
$$T_{L_{R_{R_L}}} \rightarrow \left[-FT_{L_{R_{R_{L_L}}}}\right]\left[+FT_{L_{R_{R_{L_R}}}}\right]$$
$$T_{L_{R_{R_{L_L}}}} \rightarrow [-F][+F]$$
$$T_{L_{R_{R_{L_R}}}} \rightarrow [-F][+F]$$
$$T_{L_{R_{R_R}}} \rightarrow \left[-FT_{L_{R_{R_{R_L}}}}\right]\left[+FT_{L_{R_{R_{R_R}}}}\right]$$
$$T_{L_{R_{R_{R_L}}}} \rightarrow [-F][+F]$$
$$T_{L_{R_{R_{R_R}}}} \rightarrow [-F][+F]$$
$$T_R \rightarrow \left[-FT_{R_L}\right]\left[+FT_{R_R}\right]$$
$$T_{R_L} \rightarrow \left[-FT_{R_{L_L}}\right]\left[+FT_{R_{L_R}}\right]$$
$$T_{R_{L_L}} \rightarrow \left[-FT_{R_{L_{L_L}}}\right]\left[+FT_{R_{L_{L_R}}}\right]$$
$$T_{R_{L_{L_L}}} \rightarrow [-F][+F]$$
$$T_{R_{L_{L_R}}} \rightarrow \left[-FT_{R_{L_{L_{R_L}}}}\right][+F]$$
$$T_{R_{L_{L_{R_L}}}} \rightarrow [-F][+F]$$
$$T_{R_{L_R}} \rightarrow \left[-FT_{R_{L_{R_L}}}\right]\left[+FT_{R_{L_{R_R}}}\right]$$
$$T_{R_{L_{R_L}}} \rightarrow [-F]\left[+FT_{R_{L_{R_{L_R}}}}\right]$$
$$T_{R_{L_{R_{L_R}}}} \rightarrow [-F][+F]$$
$$T_{R_{L_{R_R}}} \rightarrow \left[-FT_{R_{L_{R_{R_L}}}}\right]\left[+FT_{R_{L_{R_{R_R}}}}\right]$$
$$T_{R_{L_{R_{R_L}}}} \rightarrow [-F][+F]$$
$$T_{R_{L_{R_{R_R}}}} \rightarrow [-F][+F]$$
$$T_{R_R} \rightarrow \left[-FT_{R_{R_L}}\right]\left[+FT_{R_{R_R}}\right]$$
$$T_{R_{R_L}} \rightarrow \left[-FT_{R_{R_{L_L}}}\right]\left[+FT_{R_{R_{L_R}}}\right]$$
$$T_{R_{R_{L_L}}} \rightarrow \left[-FT_{R_{R_{L_{L_L}}}}\right][+F]$$
$$T_{R_{R_{L_{L_L}}}} \rightarrow [-F][+F]$$
$$T_{R_{R_{L_R}}} \rightarrow [-F][+F]$$
$$T_{R_{R_R}} \rightarrow \left[-FT_{R_{R_{R_L}}}\right]\left[+FT_{R_{R_{R_R}}}\right]$$
$$T_{R_{R_{R_L}}} \rightarrow \left[-FT_{R_{R_{R_{L_L}}}}\right]\left[+FT_{R_{R_{R_{L_R}}}}\right]$$
$$T_{R_{R_{R_{L_L}}}} \rightarrow [-F][+F]$$
$$T_{R_{R_{R_{L_R}}}} \rightarrow [-F][+F]$$
$$T_{R_{R_{R_R}}} \rightarrow [-F]\left[+FT_{R_{R_{R_{R_R}}}}\right]$$
$$T_{R_{R_{R_{R_R}}}} \rightarrow [-F][+F]$$

Table 1: Continued.

(1) For each variable V_i with productions (in Greibach form),

$$V_i \longrightarrow t_{i_1}U_{i_1}, t_{i_2}U_{i_2}, \ldots, t_{i_{k_i}}U_{i_{k_i}}, \tag{6}$$

where $\{t_{i_1}, t_{i_2}, \ldots, t_{i_{k_i}},\}$ are terminals and $\{U_{i_1}, U_{i_2}, \ldots, U_{i_{k_i}},\}$ are nonterminals. The formal algebraic expression for each variable is

$$V_i = \sum_{j=1}^{k_i} t_{i_j}U_{i_j}. \tag{7}$$

(2) By replacing every terminal t_{i_j} with an auxiliary variable z, one obtains the generating function

$$V_i(z) = \sum_{n=1}^{\infty} N_i(n)\, z^n, \tag{8}$$

where $N_i(n)$ is the number of words of length n descending from V_i.

(3) Let $N(n)$ be the largest one of $N_i(n)$, $N(n) = \max\{N_i(n), \text{for all } i\}$. The previous series converges

Table 2: Classification based on the similarity of rewriting rules.

Classification of rules	Isomorphic Depth #0	Isomorphic Depth #1	Isomorphic Depth #2	Isomorphic Depth #3
Class #1	(19) $C_1 \rightarrow C_1C_1$ (4) $C_1 \rightarrow C_1C_2$ (4) $C_1 \rightarrow C_2C_1$ (20) $C_1 \rightarrow C_2C_2$	(8) $C_1 \rightarrow C_1C_1$ (1) $C_1 \rightarrow C_1C_3$ (1) $C_1 \rightarrow C_2C_2$ (1) $C_1 \rightarrow C_2C_4$ (1) $C_1 \rightarrow C_3C_1$ (1) $C_1 \rightarrow C_3C_3$ (1) $C_1 \rightarrow C_4C_2$ (5) $C_1 \rightarrow C_4C_4$	(3) $C_1 \rightarrow C_1C_1$ (1) $C_1 \rightarrow C_4C_2$ (1) $C_1 \rightarrow C_7C_5$ (1) $C_1 \rightarrow C_8C_8$ (1) $C_1 \rightarrow C_3C_1$ (1) $C_1 \rightarrow C_8C_6$	(1) $C_1 \rightarrow C_1C_1$ (1) $C_1 \rightarrow C_4C_3$ (1) $C_1 \rightarrow C_5C_2$
Class #2	(48) $C_2 \rightarrow$ null	(4) $C_2 \rightarrow C_4C_5$	(1) $C_2 \rightarrow C_8C_{10}$	(1) $C_2 \rightarrow C_8C_6$
Class #3		(4) $C_3 \rightarrow C_5C_4$	(1) $C_3 \rightarrow C_9C_9$	(1) $C_3 \rightarrow C_9C_7$
Class #4		(20) $C_4 \rightarrow C_5C_5$	(1) $C_4 \rightarrow C_9C_{11}$	(1) $C_4 \rightarrow C_{12}C_{10}$
Class #5		(48) $C_5 \rightarrow$ null	(1) $C_5 \rightarrow C_{10}C_8$	(1) $C_5 \rightarrow C_{13}C_{11}$
Class #6			(1) $C_6 \rightarrow C_{10}C_{10}$	(1) $C_6 \rightarrow C_{13}C_{13}$
Class #7			(1) $C_7 \rightarrow C_{11}C_9$	(1) $C_7 \rightarrow C_{13}C_{15}$
Class #8			(5) $C_8 \rightarrow C_{11}C_{11}$	(1) $C_8 \rightarrow C_{14}C_{14}$
Class #9			(4) $C_9 \rightarrow C_{11}C_{12}$	(1) $C_9 \rightarrow C_{14}C_{16}$
Class #10			(4) $C_{10} \rightarrow C_{12}C_{11}$	(1) $C_{10} \rightarrow C_{15}C_{13}$
Class #11			(20) $C_{11} \rightarrow C_{12}C_{12}$	(1) $C_{11} \rightarrow C_{15}C_{15}$
Class #12			(48) $C_{12} \rightarrow$ null	(1) $C_{12} \rightarrow C_{16}C_{14}$
Class #13				(5) $C_{13} \rightarrow C_{16}C_{16}$
Class #14				(4) $C_{14} \rightarrow C_{16}C_{17}$
Class #15				(4) $C_{15} \rightarrow C_{17}C_{16}$
Class #16				(20) $C_{16} \rightarrow C_{17}C_{17}$
Class #17				(48) $C_{17} \rightarrow$ null

when $z < R = e^{-K_0}$. The topological entropy is given by the radius of convergence R as

$$K_0 = -\ln R. \tag{9}$$

Our productions have some difference from the aforementioned definitions. First, our productions are written in Chomsky-reduced form instead of Greibach form. Second, DNA is finite sequence; it generates finite tree, but the previous formulas are applied on infinite sequences. For convenience in the DNA tree case, we rewrite the definition as follows [9].

Definition 5. Topological entropy of context free grammar for DNA tree.

(1) Assume that there are n classes of rules and that each class C_i contains n_i rules. Let $V_i \in \{C_1, C_2, \ldots, C_n\}$, $U_{ij} \in \{R_{ij}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, n_i\}$, and $a_{ijk} \in \{x : x = 1, 2, \ldots, n\}$, where each U_{ij} has the following form:

$$\begin{aligned} U_{i1} &\longrightarrow V_{a_{i11}} V_{a_{i12}}, \\ U_{i2} &\longrightarrow V_{a_{i21}} V_{a_{i22}}, \\ &\cdots \longrightarrow \cdots, \\ U_{in_i} &\longrightarrow V_{a_{in_i1}} V_{a_{in_i2}}. \end{aligned} \tag{10}$$

(2) The generating function of $V_i, V_i(z)$ has a new form as follows:

$$V_i(z) = \frac{\sum_{p=1}^{n_i} n_{ip} z V_{a_{ip1}}(z) V_{a_{ip2}}(z)}{\sum_{q=1}^{n_i} n_{iq}}. \tag{11}$$

If V_i does not have any nonterminal variables, we set $V_i(z) = 1$.

(3) After formulating the generating function $V_i(z)$, we intend to find the largest value of $z, z^{\max}$, at which $V_1(z^{\max})$ converges. Note that we use V_1 to denote the

rule for the root node of the DNA tree. After obtaining the largest value, $z^{\max}$, of $V_1(z)$, we set $R = z^{\max}$, the radius of convergence of $V_1(z)$. We define the complexity of the DNA tree as

$$K_0 = -\ln R. \tag{12}$$

Now we can do some examples of computation procedure for the complexity. According to our definition, the given values for the class parameters are listed in Table 3. There are five classes, so we obtain the formulas for $V_5(z'), V_4(z'), V_3(z'), V_2(z')$, and $V_1(z')$ successively. They are

$$\begin{aligned}
V_5\left(z'\right) &= 1 \,(\text{by definition}),\\
V_4\left(z'\right) &= \frac{\sum_{p=1}^{n_4} n_{4p} z' V_{a_{4p1}}\left(z'\right) V_{a_{4p2}}\left(z'\right)}{\sum_{q=1}^{n_i} n_{iq}}\\
&= \frac{z' \times \left(20 \times V_5\left(z'\right) \times V_5\left(z'\right)\right)}{20} = z',\\
V_3\left(z'\right) &= \frac{\sum_{p=1}^{n_3} n_{3p} z' V_{a_{3p1}}\left(z'\right) V_{a_{3p2}}\left(z'\right)}{\sum_{q=1}^{n_i} n_{iq}}\\
&= \frac{z' \times \left(4 \times V_5\left(z'\right) \times V_4\left(z'\right)\right)}{4} = z'^2,\\
V_2\left(z'\right) &= \frac{\sum_{p=1}^{n_2} n_{2p} z' V_{a_{2p1}}\left(z'\right) V_{a_{2p2}}\left(z'\right)}{\sum_{q=1}^{n_i} n_{iq}}\\
&= \frac{z' \times \left(4 \times V_4\left(z'\right) \times V_5\left(z'\right)\right)}{4} = z'^2,\\
V_1\left(z'\right) &= \frac{\sum_{p=1}^{n_1} n_{1p} z' V_{a_{1p1}}\left(z'\right) V_{a_{1p2}}\left(z'\right)}{\sum_{q=1}^{n_i} n_{iq}}\\
&= \frac{8z' \times V_1\left(z'\right)^2 + 2\left(z'\right)^3 \times V_1\left(z'\right)}{19}\\
&\quad + \frac{\left(2\left(z'\right)^5 + 2\left(z'\right)^4 + 5\left(z'\right)^3\right)}{19}.
\end{aligned} \tag{13}$$

Rearranging the previous equation for $V_1(z')$, we obtain a quadratic for $V_1(z')$:

$$\begin{aligned}
&\frac{8}{19}\left(z'\right) \times V_1\left(z'\right) + \left(1 - \frac{2}{19}\left(z'\right)^3\right) \times V_1\left(z'\right)\\
&\quad + \frac{1}{19}\left(2\left(z'\right)^5 + 2\left(z'\right)^4 + 5\left(z'\right)^3\right) = 0.
\end{aligned} \tag{14}$$

Solving $V_1(z')$, we obtain the formula

$$V_1\left(z'\right) = \left(\frac{\left(z'\right)^2}{4} - \frac{19}{8z'}\right) \pm \frac{19}{8z'}\sqrt{B^2 - A}, \tag{15}$$

Table 3: The values for the class parameters of Table 2.

	Classification of rules	Isomorphic depth #1
($n = 5$)	Class #1 ($n_1 = 8$)	n_{11} $n_{111}n_{112}$ (8) $C_1 \rightarrow C_1C_1$ n_{12} $n_{121}n_{122}$ (1) $C_1 \rightarrow C_1C_3$ n_{13} $n_{131}n_{132}$ (1) $C_1 \rightarrow C_2C_2$ n_{14} $n_{141}n_{142}$ (1) $C_1 \rightarrow C_2C_4$ n_{15} $n_{151}n_{152}$ (1) $C_1 \rightarrow C_3C_1$ n_{16} $n_{161}n_{162}$ (4) $C_1 \rightarrow C_3C_3$ n_{17} $n_{171}n_{172}$ (1) $C_1 \rightarrow C_4C_2$ n_{18} $n_{181}n_{182}$ (5) $C_1 \rightarrow C_4C_4$
	Class #2 ($n_2 = 1$)	n_{21} $n_{211}n_{212}$ (4) $C_2 \rightarrow C_4C_5$
	Class #3 ($n_3 = 1$)	n_{31} $n_{311}n_{312}$ (4) $C_3 \rightarrow C_5C_4$
	Class #4 ($n_4 = 1$)	n_{41} $n_{411}n_{412}$ (20) $C_4 \rightarrow C_5C_5$
	Class #5 ($n_5 = 1$)	n_{51} $n_{511}n_{512}$ (48) $C_5 \rightarrow$ null

Table 4: Test data with topological entropy method and our method.

Type	Name	Koslicki method	Our method
DNA	E. coli[a]	Available	Available
	EV71[b]	Available	Available
	H1N1[c]	Available	Available
	H5N1[d]	Available	Available
	SARS[e]	Available	Available
Amino acid	Abrin	Too short	Available
	Ricin	Too short	Available
	BSE[f]	Too short	Available
	CJD[g]	Too short	Available

[a] *Escherichia coli* O157:H7.
[b] Enterovirus 71.
[c] Influenza A virus subtype H1N1.
[d] Influenza A virus subtype H5N1.
[e] Severe acute respiratory syndrome.
[f] Bovine spongiform encephalopathy.
[g] Creutzfeldt-Jakob disease.

where

$$\begin{aligned}
A &= \frac{32}{361}\left(2\left(z'\right)^6 + 2\left(z'\right)^5 + 5\left(z'\right)^4\right),\\
B &= 1 - \frac{2}{19}\left(z'\right)^3.
\end{aligned} \tag{16}$$

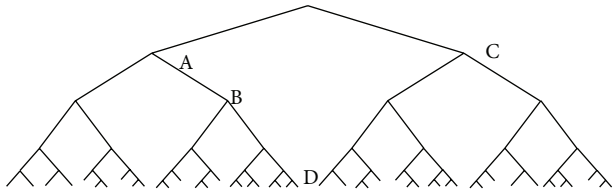


Figure 9: Example of homomorphism and isomorphism.

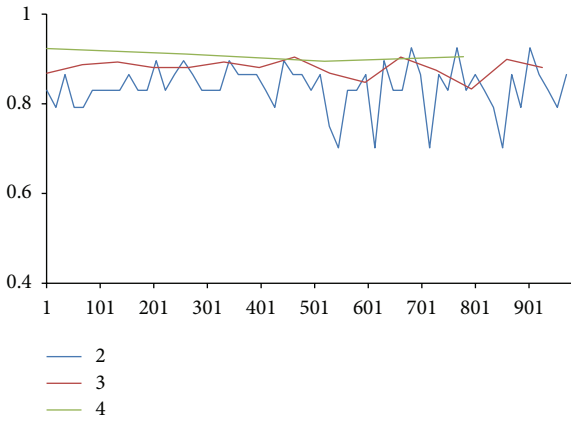


Figure 10: Koslicki method (topological entropy method, TE for short) example.

Finally, the radius of convergence, R, and complexity, $K_0 = -\ln R$, can be obtained from this formula. But, computing the $z^{\max}$ directly is difficult, so we use iterations and region tests to approximate the complexity; details are as follows.

(1) Rewrite the generating function as

$$V_i^m\left(z'\right) = \frac{\sum_{p=1}^{n_i} n_{ip} z' V_{a_{ip1}}^{m-1}\left(z'\right) V_{a_{ip2}}^{m-1}\left(z'\right)}{\sum_{q=1}^{n_i} n_{iq}}, \quad (17)$$

$$V_i^0\left(z'\right) = 1.$$

(2) The value from $V_i^0(z')$ to $V_i^m(z')$. When $V_i^{m-1}(z') = V_i^m(z')$ for all rules, we say that $V_i^m(z')$ reach the convergence, but z' is not the $z^{\max}$ we want. Here, we set $m = 1000$ for each iteration.

(3) Now we can test whether $V_i(z')$ is convergent or divergent at a number z'. We use binary search to test every real number between 0 and 1; in every test, when $V_i(z')$ converges, we set bigger z' next time, but when $V_i(z')$ diverges, we set smaller z' next time. Running more iterations will obtain more precise radius.

4. Results

In 2011, Koslicki [1] gave an efficient way to compute the topological entropy of DNA sequence. He used fixed length depending on subword size to compute topological entropy of sequence. For example, in Figure 10 (all DNA and amino acid data can be found in NCBI website, http://www.ncbi.nlm.nih.gov/), the sequence length is 1027 characters, and there are three subword sizes 2, 3, and 4 with blue, red, and green lines, respectively. For larger subword size, much larger fragment is required for complexity computation. The required fragment size grows exponentially, while the length of sequence is not dependent on the growth rate of subword size, so it is not a good method for us overall.

We present a new method called structural complexity in previous sections, and there are several benefits from using our method instead of Koslicki method, described as follows.

(1) Our results are very different from those obtained by the topological entropy method; see the colored lines in Figures 11~14. These figures showed that our method is much sensitive to certain arrangements of the elements in the sequence.

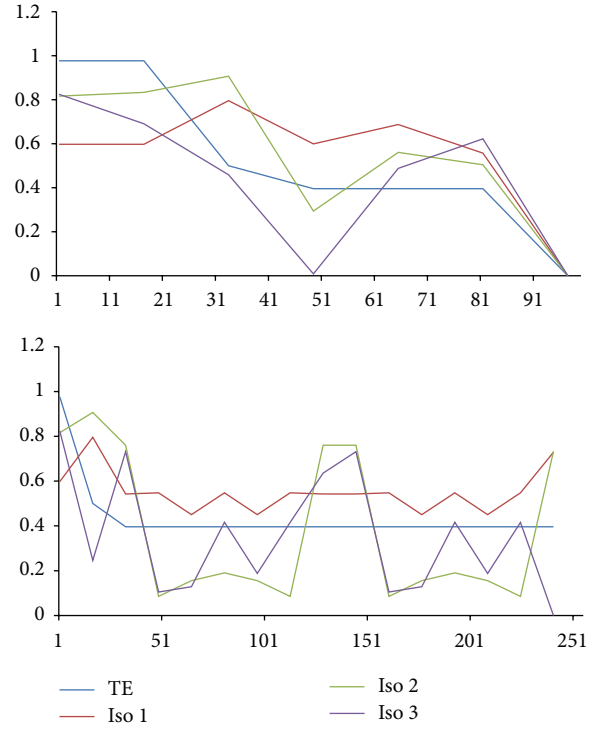


Figure 11: Our method compared with TE using test sequences.

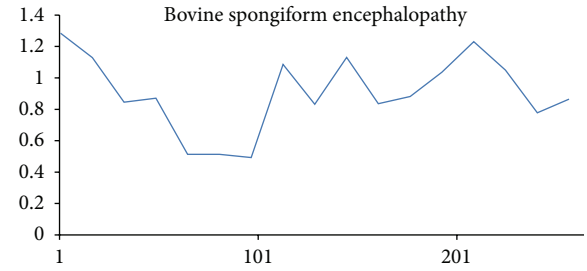


Figure 12: An amino acid sequence example, Bovine spongiform encephalopathy.

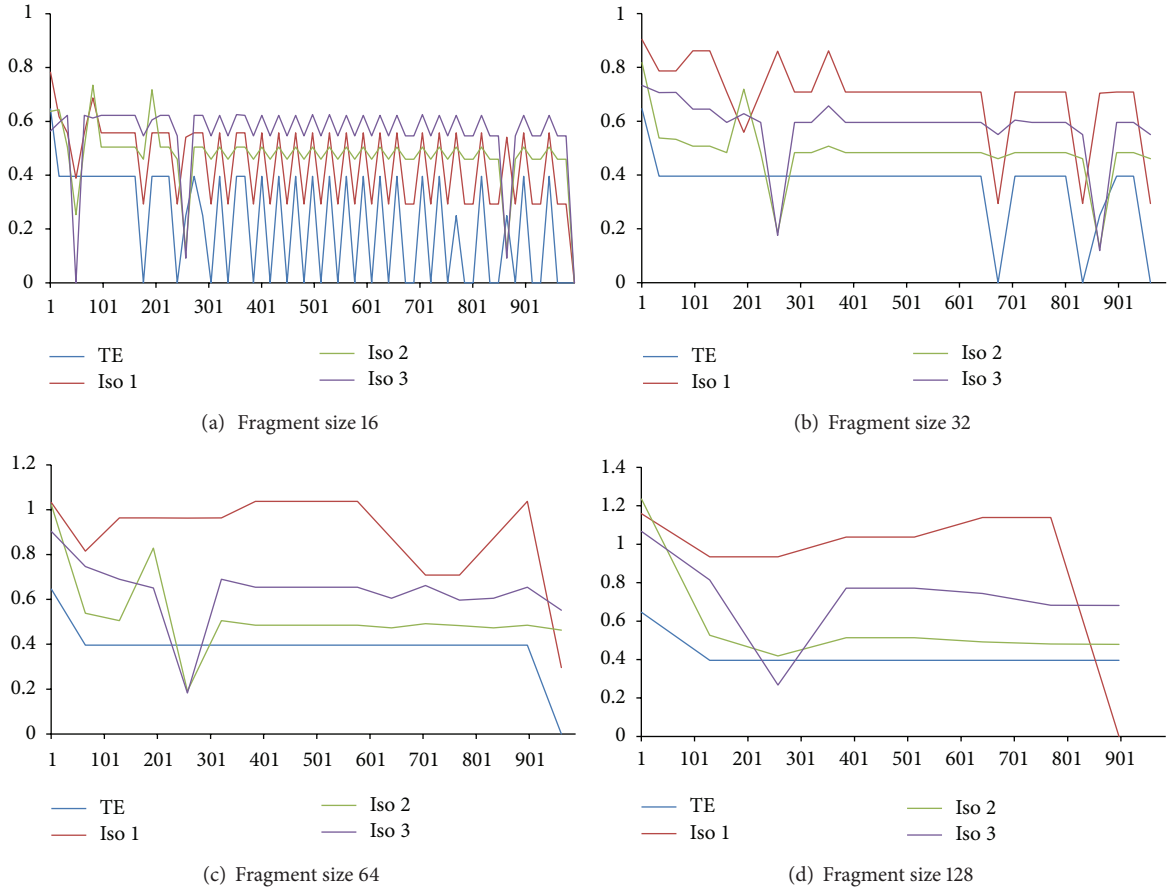


Figure 13: Compare with different methods.

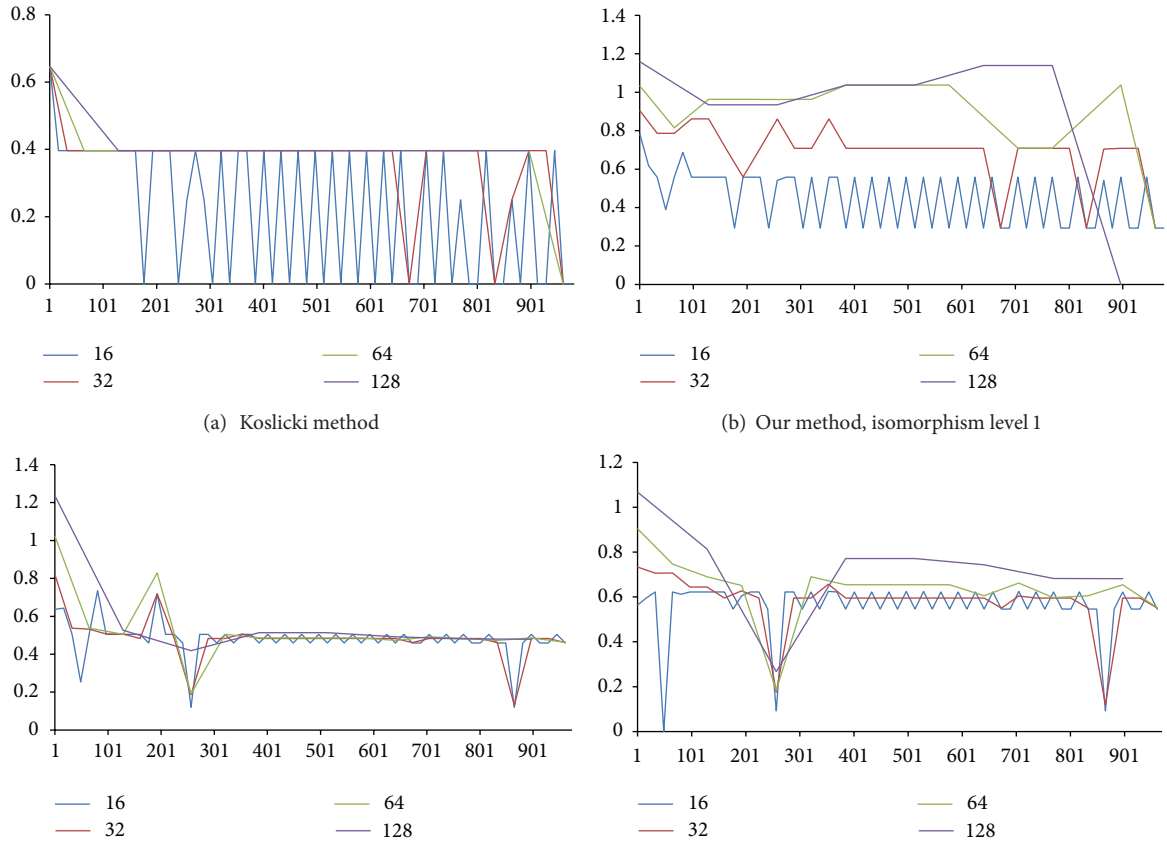
(2) Two different characters that exchange position will change value since Koslicki method just calculates the statistical values without structural information. Result was shown in Figure 11 bottom chart; the test sequence repeats the same subword several times. For blue line, all complexity values from topological entropy are equal within the region of repeated subwords. For red line, complexity values depend on the structure of subword. When the fragment of sequence is different from each other, our method will evaluate to different values.

(3) Our method can also calculate amino acid sequences. The Koslicki method depends on alphabet size and subword size, for example, in the basic length 2 substring calculation; since standard amino acid types have up to 20, it requires a minimum length of $20^2 + 2-1$ to calculate, but the amino acid strings are usually very short. Sometimes, Koslicki method cannot compute the amino acid sequence efficiently. Figure 12 shows that complexity of amino acid sequence can also be calculated by our method.

We also did experiments with lots of data, including fixed fragment size and fixed method on test sequences (see Figures 13 and 14). Here, we redefine the Koslicki method; the fragment size is no longer dependent on subword size. Instead, fixed length fragment like our method is applied. This change allows us to compare the data easier, and not restricted to the exponentially growing fragment size anymore. In Figure 13, we found that for larger fragment, the complexity curve will become smoothly because fragments for each data point contain more information. And we note that there is a common local peak value of those figures; the *simple sequence region* is big enough that our fragment size still contains the same simple sequence.

When we compare with the same method shown in Figure 14, we found the same situation more obviously. Thus, if we have many complexity values with different sizes, we have the opportunity to restore the portion of the DNA.

4.1. Application to Virus Sequences Database and Other Sequences. Now we can apply our technique to Chinese word sequences. Togawa et al. [18] gave a complexity of Chinese words, but his study was based on the number of strokes, which is different from our method. Here we use Big5 encoding for our system. Since the number of Chinese words is larger than 10000, we cannot directly use words as alphabet, so we need some conversion. We read a Chinese word into four hexadecimal letters so that we can replace the sequence with tree representation and compute the complexity.



(a) Koslicki method

(b) Our method, isomorphism level 1

(c) Our method, isomorphism level 2

(d) Our method, isomorphism level 3

Figure 14: Compare with different fragment sizes.

When it comes to biomedical section, we can create virus comparison database. Once a new virus or prion has been found, it will be easy to select corresponding drugs at the first time, according to cross comparison with each other by complexity in the database. We focus on most important viruses in recent years, such as *Escherichia coli* O157:H7 (*E. coli* o157), Enterovirus 71 (EV71), Influenza A virus subtype H1N1 (H1N1), Influenza A virus subtype H5N1 (H5N1), and severe acute respiratory syndrome (SARS). In recent years, these viruses have a significant impact and threat on the human world. We test these viruses and prions listed in Table 4. Here we can see that all prion regions cannot be analyzed by Koslicki method, but we can do it.

Finally, if any object can be written as a sequence, and there exists tree representation with alphabet of sequence, we can compute the complexity of the object.

5. Summary

In this paper, we give a method for computing complexity of DNA sequences. The traditional method focused on the statistical data or simply explored the structural complexity without value. In our method, we transform the DNA sequence to DNA tree with tree representations at first.

Then we transform the tree to context-free grammar format, so that it can be classified. Finally, we use redefined generating function and find the complexity values. We give a not only statistical but also structural complexity for DNA sequences, and this technique can be used in many important applications.

Acknowledgment

This work was supported by the National Science Council under project NSC 100-2221-E-002-234-MY3.

References

[1] D. Koslicki, "Topological entropy of DNA sequences," *Bioinformatics*, vol. 27, no. 8, Article ID btr077, pp. 1061–1067, 2011.

[2] C. Cattani, G. Pierro, and G. Altieri, "Entropy and multifractality for the myeloma multiple tet 2 gene," *Mathematical Problems in Engineering*, vol. 2012, Article ID 193761, 14 pages, 2012.

[3] S. Manna and C. Y. Liou, "Reverse engineering approach in molecular evolution: simulation and case study with enzyme proteins," in *Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP '06)*, pp. 529–533, 2006.

[4] R. Zhang and C. T. Zhang, "Z curves, an intutive tool for visualizing and analyzing the DNA sequences," *Journal of*

Biomolecular Structure and Dynamics, vol. 11, no. 4, pp. 767–782, 1994.

[5] P. Tiňo, "Spatial representation of symbolic sequences through iterative function systems," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 29, no. 4, pp. 386–393, 1999.

[6] 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.

[7] B. L. Hao, H. C. Lee, and S. Y. Zhang, "Fractals related to long DNA sequences and complete genomes," *Chaos, solitons and fractals*, vol. 11, no. 6, pp. 825–836, 2000.

[8] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.

[9] C. Y. Liou, T. H. Wu, and C. Y. Lee, "Modeling complexity in musical rhythm," *Complexity*, vol. 15, no. 4, pp. 19–30, 2010.

[10] P. Prusinkiewicz, "Score generation with lsystems," in *Proceedings of the International Computer Music Conference*, pp. 455–457, 1986.

[11] P. Prusinkiewicz and A. Lindenmayer, *The Algorithmic Beauty of Plants*, Springer, New York, NY, USA, 1996.

[12] P. Worth and S. Stepney, "Growing music: musical interpretations of L-systems," in *Applications of Evolutionary Computing*, vol. 3449 of *Lecture Notes in Computer Science*, pp. 545–550, Springer, Berlin, Germany, 2005.

[13] A. Lindenmayer, "Mathematical models for cellular interactions in development II. Simple and branching filaments with two-sided inputs," *Journal of Theoretical Biology*, vol. 18, no. 3, pp. 300–315, 1968.

[14] "Wikipedia: L-system—Wikipedia, the free encyclopedia," 2012.

[15] H. Barlow, "Unsupervised learning," *Neural Computation*, vol. 1, no. 3, pp. 295–311, 1989.

[16] R. Badii and A. Politi, *Complexity: Hierarchical Structures and Scaling in Physics*, vol. 6, Cambridge University Press, Cambridge, UK, 1999.

[17] W. Kuich, "On the entropy of context-free languages," *Information and Control*, vol. 16, no. 2, pp. 173–200, 1970.

[18] T. Togawa, K. Otsuka, S. Hiki, and H. Kitaoka, "Complexity of chinese characters," *Forma*, vol. 15, pp. 409–414, 2001.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 902143, 8 pages
http://dx.doi.org/10.1155/2013/902143

Research Article

Improving Spatial Adaptivity of Nonlocal Means in Low-Dosed CT Imaging Using Pointwise Fractal Dimension

Xiuqing Zheng,[1] Zhiwu Liao,[2] Shaoxiang Hu,[3] Ming Li,[4] and Jiliu Zhou[1]

[1] *College of Computer Science, Sichuan University, No. 29 Jiuyanqiao Wangjiang Road, Chengdu 610064, Sichuan, China*
[2] *School of Computer Science, Sichuan Normal University, No. 1819 Section 2 of Chenglong Road, Chengdu 610101, Sichuan, China*
[3] *School of Automation Engineering, University of Electronic Science and Technology of China, No. 2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu 611731, Sichuan, China*
[4] *School of Information Science and Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China*

Correspondence should be addressed to Zhiwu Liao; liaozhiwu@163.com

Received 25 January 2013; Accepted 6 March 2013

Academic Editor: Shengyong Chen

NLMs is a state-of-art image denoising method; however, it sometimes oversmoothes anatomical features in low-dose CT (LDCT) imaging. In this paper, we propose a simple way to improve the spatial adaptivity (SA) of NLMs using pointwise fractal dimension (PWFD). Unlike existing fractal image dimensions that are computed on the whole images or blocks of images, the new PWFD, named pointwise box-counting dimension (PWBCD), is computed for each image pixel. PWBCD uses a fixed size local window centered at the considered image pixel to fit the different local structures of images. Then based on PWBCD, a new method that uses PWBCD to improve SA of NLMs directly is proposed. That is, PWBCD is combined with the weight of the difference between local comparison windows for NLMs. Smoothing results for test images and real sinograms show that PWBCD-NLMs with well-chosen parameters can preserve anatomical features better while suppressing the noises efficiently. In addition, PWBCD-NLMs also has better performance both in visual quality and peak signal to noise ratio (PSNR) than NLMs in LDCT imaging.

1. Introduction

Radiation exposure and associated risk of cancer for patients from CT examination have been increasing concerns in recent years. Thus minimizing the radiation exposure to patients has been one of the major efforts in modern clinical X-ray CT radiology [1–8]. However, the presentation of serious noise and many artifacts degrades the quality of low-dose CT images dramatically and decreases the accuracy of diagnosis dose. Although many strategies have been proposed to reduce their noise and artifacts [9–14], filtering noise from clinical scans is still a challenging task, since these scans contain artifacts and consist of many structures with different shape, size, and contrast, which should be preserved for making correct diagnosis.

Recently nonlocal means (NLMs) is proposed for improving the performance of classical adaptive denoising methods [15–17] and shows good performance even in low-dose CT (LDCT) imaging [18–20].

There are two novel ideas for NLMs. One is that the similar points should be found by comparing the difference between their local neighborhoods instead of by comparing their gray levels directly. Since gray levels of LDCT will be polluted seriously by noises and artifacts, finding similar points by local neighborhoods instead of by gray levels directly will help NLMs find correct similar points. The other

important idea for NLMs is that the similar points should be searched in large windows to guarantee the reliability of estimation.

Following the previous discussion, the NLMs denoising should be performed in two windows: one is comparison patch and the other is searching window. The sizes of these two windows and the standard deviation σ_r of the Gaussian kernel, which is used for computing the distance between two neighborhoods, should be determined according to the standard deviation of noises [15–17], and these three parameters are identical in an image.

Some researchers find that identical sizes of two windows and identical Gaussian kernel σ_r in an image are not the best choice for image denoising [21–25]. The straightest motivation is that the parameters should be modified according to the different local structures of images. For example, the parameters near an edge should be different from parameters in a large smooth region.

An important work to improve the performance of NLMs is quasi-local means (QLMs) proposed by us [21, 22]. We argue that nonlocal searching windows are not necessary for most of image pixels. In fact, for points in smooth regions, which are the majority of image pixels, local searching windows are big enough, while for points near singularities, only the minority of image pixels, nonlocal search windows are necessary. Thus the method is named quasi-local whereit is local for most of image pixels and nonlocal only for pixels near singularities. The searching windows for quasi-local means (QLMs) are variable for different local structures, and QLMs can get better singularity preservation in image denoising than classical NLMs.

Other important works about improving spatial adaptivity of NLMs are proposed very recently [23–25]. The starting point for these works is that the image pixels are parted into different groups using supervised learning or semisupervised learning and clustering. However, the learning and clustering will waste a lot of computation time and resource, which will hamper them to be applied in medical imaging. Thus we must propose a new method for improving the spatial adaptivity with a simple way.

In this paper we propose a simple and powerful method to improve spatial adaptivity for NLMs in LDCT imaging using pointwise fractal dimension (PWFD) where PWFD is computed pixel by pixel in a fixed-size window centered at the considering pixel. According to the new definition of PWFD, different local structures will be with different local fractal dimensions, for example, pixels near edge regions will be with relatively big PWFDs, while PWFDs of pixels in smooth regions will be zeros. Thus PWFD can provide local structure information for image denoising. After defined PWFD, which can fit different local structures of images well, we design a new weight function by combining the new PWFD difference between two considering pixels with the weight of original NLMs measured by gray level difference between two comparison windows. Thus using this new weight function, the proposed method will not only preserve the gray level adaptivity of NLMs but also improve the SA of NLMs.

The arrangement of this paper is as follows: In Section 2, the backgrounds are introduced, then the new proposed method is presented in Section 3, the experiment results are shown and discussed in Section 4, and the final part is the conclusions and acknowledgment.

2. Backgrounds

In this section, we will introduce related backgrounds of the proposed method.

2.1. Noise Models. Based on repeated phantom experiments, low-mA (or low-dose) CT calibrated projection data after logarithm transform were found to follow approximately a Gaussian distribution with an analytical formula between the sample mean and sample variance; that is, the noise is a signal-dependent Gaussian distribution [11].

The photon noise is due to the limited number of photons collected by the detector. 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 denotes 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

$$r_i = -\ln\left[\frac{N(i, \alpha)}{N_0(i, \alpha)}\right], \tag{1}$$

where $N_0(i, \alpha)$ is a constant and $N(i, \alpha)$ is Poisson distribution with mean N.

Thus we have

$$N(i, \alpha) = N_0(i, \alpha)\exp(-r_i). \tag{2}$$

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 [11]. We have

$$\sigma_i^2(\mu_i) = f_i \exp\left(\frac{\mu_i}{\gamma}\right), \tag{3}$$

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 [11].

2.2. Nonlocal Means (NLMs). Given a discrete noisy image y, the estimated value $(\hat{y}_i)$, for a pixel i, is computed as a weighted nonlocal average:

$$\hat{y}_i = \frac{1}{C(i)} \sum_{j \in B(i,r)} y_j \omega(i, j), \tag{4}$$

where $B(i,r)$ indicates a neighborhood centered at i and size $(2r+1)\times(2r+1)$, called searching window, and $C(i) = \sum_{j\in B(i,r)}\omega(i,j)$. The family of weights $\{\omega(i,j)\}$ depend on the similarity between the pixels i and j and satisfy $0 \le \omega(i,j) \le 1$ and $\sum_{j\in B(i,r)}\omega(i,j) = 1$.

The similarity between two pixels i and j, $d^2(i,j)$ depends on the similarity of the intensity gray level vectors $B(i,f)$ and $B(j,f)$, where $B(k,f)$ denotes a square window with fixed size $(2f+1)\times(2f+1)$ and centered at a pixel k, named comparison patch:

$$d^2(i,j) = \frac{1}{(2f+1)^2}\sum_{k\in B(0,f)}\left(y_{i+k} - y_{j+k}\right)^2, \tag{5}$$

and the weights $\omega(i,j)$ are computed as

$$\omega(i,j) = e^{-\max(d^2-2\sigma_N^2,0)/h^2}, \tag{6}$$

where σ_N denotes the standard deviation of the noise and h is the filtering parameter set depending on the value σ_N.

2.3. Box-Counting Dimension. Box-counting dimension, also known as Minkowski dimension or Minkowski-Bouligand dimension, is a way of determining the fractal dimension of a set S in a Euclidean space R^n or more generally in a metric space (X,d). To calculate this dimension for a fractal S, putting this fractal on an evenlyspaced grid and count how many boxes are required to cover the set. The box-counting dimension is calculated by seeing how this number changes as we make the grid finer by applying a box-counting algorithm.

Suppose that $N(\varepsilon)$ is the number of boxes of side length ε required to cover the set. Then the box-counting dimension is defined as

$$\dim(S) = \lim_{\varepsilon\to 0}\frac{\log N(\varepsilon)}{\log(1/\varepsilon)}. \tag{7}$$

Given an $N\times N$ image whose gray level is G, then the image is part into the $\varepsilon\times\varepsilon$ grids, which are related to $\varepsilon\times\varepsilon\times\varepsilon$ cube grids. If for the jth grid, the greatest gray level is in the ιth box and the smallest is in the κth box, then the box number for covering the grid is

$$n_\varepsilon = \iota - \kappa + 1. \tag{8}$$

Therefore the box number for covering the whole image is

$$N_\varepsilon = \sum_j n_\varepsilon(j). \tag{9}$$

Selecting different scale ε, we can get related N_ε. Thus we have a group of pairs $(\varepsilon, N_\varepsilon)$. The group can be fit with a line using least-squares fitting, the slope of the line is the box-counting dimension.

3. The New Method

In this section, we will present our new proposed algorithm in detail. The motivation for the proposed method is that SA of NLMs should be improved in a simpler way. The new PWFD is introduced firstly to adapt complex image local structures, and then the new weight functions based on PWFD are discussed. At the end of this section, the procedures of the proposed method are shown.

3.1. Pointwise Box-Counting Dimension. In image processing, the fractal dimension usually is used for characterizing roughness and self-similarity of images. However, most of works only focus on how to compute fractal dimensions for images or blocks of images [26–30]. Since fractal dimension can characterize roughness and self-similarity of images, it also can be used for characterizing the local structures of images by generalizing it to PWFD, which is computed pixel by pixel using a fixed-size window centered in the considered pixel. Thus, each pixel in an image has a PWFD and it equals the fractal dimension of the fixed-size window centered in the considered pixel.

Following the previous discussion, the pointwise box-counting dimension (PWBCD) starts from replacing each pixel i to a fixed-size window $r\times r$ centered at i. It is obvious that PWFD can be generalized to all definitions of fractal dimensions. However, in order to make our explanation more clearly, we only extend the new definition to PWBCD.

According to the new PWFD, PWBCD should be computed for each pixel in the image. For each pixel i, the PWBCD is computed in a fixed-size $r\times r$ window centered at i.

The $r\times r$ window is parted into the $\varepsilon\times\varepsilon$ grids, which are related to $\varepsilon\times\varepsilon\times\varepsilon$ cube grids. If for the jth grid, the greatest gray level is in the ιth box and the smallest is in the κth box, then the box number for covering the grid is

$$n_\varepsilon(i) = \iota - \kappa + 1. \tag{10}$$

Therefore the box number for covering the whole $r\times r$ window is

$$N_\varepsilon(i) = \sum_j n_\varepsilon(j). \tag{11}$$

Selecting different scale ε, we can get related $N_\varepsilon(i)$. Thus we have a group of pairs $(\varepsilon, N_\varepsilon(i))$. The group can be fit with a line using least-squares fitting; the slope $k(i)$ of the line is the box-counting dimension.

Note that each pixel in an image has a PWBCD value. Thus we can test the rationality for PWBCD by showing PWBCD values using an image. In these PWBCD images, high PWBCD values are shown as white points, while low PWBCD values are shown as gray or black points. If PWBCD images are similar to the original images with big PWBCD values near singularities and small PWBCD values in smooth regions, the rationality is testified.

Figure 1 shows PWBCD images for three images: an test image composed by some blocks with different gray levels, a LDCT image, and 512×512 barbara. The white points signify the pixels with big fractal dimensions, while black points signify the pixels with small fractal dimensions. Here, $r = 32$ and $\varepsilon = 2, 4, 8, 16, 32$. Note that the white parts correspond the texture parts of barbara and soft tissues of the

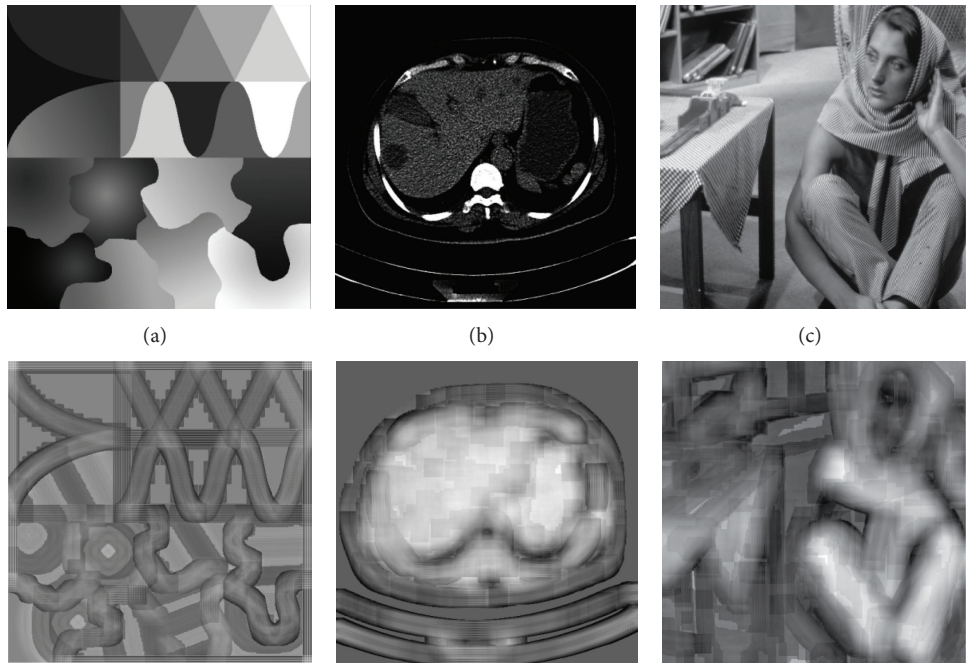


Figure 1: Images and their pointwise box-counting dimension images: the first row shows images while the second row shows their pointwise box-counting dimension images. Here $r = 32$ and $\varepsilon = 2, 4, 8, 16, 32$.

second image in the first row. Moreover, the PWBCD images are very similar to the original images which demonstrate that the PWBCD can be used for characterizing the local structure of images.

3.2. The New Weight Function. After defining the PWBCD, we must find an efficient and powerful way to use the PWBCD in NLMs directly. Just as discussed in the previous subsection, PWBCD can characterize the local structures for images well. Thus PWBCD should be used to weight the points in the searching patch. That is, (6) should be changed as

$$\omega(i, j) = e^{-\max(d^2 - 2\sigma_N^2, 0)/h_1^2 - (k(i) - k(j))^2/h_2^2}, \quad (12)$$

where $k(\cdot)$ is FDBCD value for the considering pixel and is computed according to the method proposed in Section 3.1, σ_N denotes the standard deviation of the noise, h_1, h_2 are the filtering parameters. $d^2(i, j)$ is the similarity between two pixels i and j depending on the similarity of the intensity gray level vectors $B(i, f)$ and $B(j, f)$, where $B(k, f)$ denotes a square window with fixed size $(2f+1) \times (2f+1)$ and centered at a pixel k:

$$d^2(i, j) = \frac{1}{(2f+1)^2} \sum_{k \in B(0, f)} \left(y_{i+k} - y_{j+k}\right)^2. \quad (13)$$

Given a discrete noisy image y, the estimated value $(\widehat{y}_i)$, for a pixel i is computed as a weighted nonlocal average:

$$\widehat{y}_i = \frac{1}{C(i)} \sum_{j \in B(i, r)} y_j \omega(i, j), \quad (14)$$

where $B(i, r)$ indicates a neighborhood centered at i and size $(2r + 1) \times (2r + 1)$, called searching window, and $C(i) = \sum_{j \in B(i, r)} \omega(i, j)$. Note that the family of weights $\{\omega(i, j)\}$ depend on the similarity between the pixels i and j and satisfy $0 \le \omega(i, j) \le 1$ and $\sum_{j \in B(i, r)} \omega(i, j) = 1$.

3.3. The Steps of the New Method. The steps of PWBCD-NLMs are as follows.

(1) Compute pointwise box-counting dimension for each ofthe pixels.
For each of the pixels, given $r = 2^n$, $n \in Z$ and $\varepsilon = 2, 4, \ldots, r$, compute PWBCD according to Section 3.1, and get a matrix K with the same size as the image.

(2) Compute weights. determine parameters: σ_N, h_1, h_2, the size of comparison window cr, and the size of the searching patch sr.
Compute the difference between two comparison windows, d^2, using (13).
Compute the weights $\omega(i, j)$ using (12).

(3) Estimate real gray levels: estimate real levels $\widehat{y}(i)$ using (14).

4. Experiments and Discussion

The main objective for smoothing LDCT images is to delete the noise while preserving anatomy features for the images.

In order to show the performance of PWBCD-NLMs, a 2-dimensional 512×512 test phantom is shown in Figure 1(a).

(a) Noisy image with $f_i = 2.5$, $T = 2e + 4$

(b) Reconstructed image from (a) using NLMs

(c) Reconstructed image from (a) using PWBCD-NLMs

(d) Noisy image with $f_i = 4.0$, $T = 2e + 4$

(e) Reconstructed image from (d) using NLMs

(f) Reconstructed image from (d) using PWBCD-NLMs

Figure 2: Noisy test images and reconstructed images.

The number of bins per view is 888 with 984 views evenly spanned on a circular orbit of 360°. The detector arrays are on an arc concentric to the X-ray source with a distance of 949.075 mm. The distance from the rotation center to the X-ray source is 541 mm. The detector cell spacing is 1.0239 mm.

The LDCT projection data (sinogram) is simulated by adding Gaussian-dependent noise (GDN) whose analytic form between its mean and variance has been shown in (3) with $f_i = 2.5$, 3.5, 4.0 and $T = 2e + 4$. The projection data is reconstructed by standard Filtered Back Projection (FBP). Since both the original projection data and sinogram have been provided, the evaluation is based on peak signal to noise ration (PSNR) between the ideal reconstructed image and reconstructed image.

The PWBCDs for images are computed according to Section 3.1, and the parameters are $r = 32$ and $\varepsilon = 2, 4, 8, 16, 32$. The new proposed method is compared with NLMs, and their common parameters includes the standard deviation of noise $\sigma_N = 15$; the size of comparison window is 7×7 ($cr = 7$), while the size of searching patch is 21×21 ($sr = 21$). The other parameter for NLMswhick is the Gaussian kernel for weights defined on (13) is $h = 12$ and the parameters for the new method are the sizes of Gaussian kernel for two weights defined on (12): $h_1 = 15$ for the weights of difference between comparison window and $h_2 = 10$ for the weights between two PWBCDs. All parameters are chosen by hand with many experiments, which has the best performance.

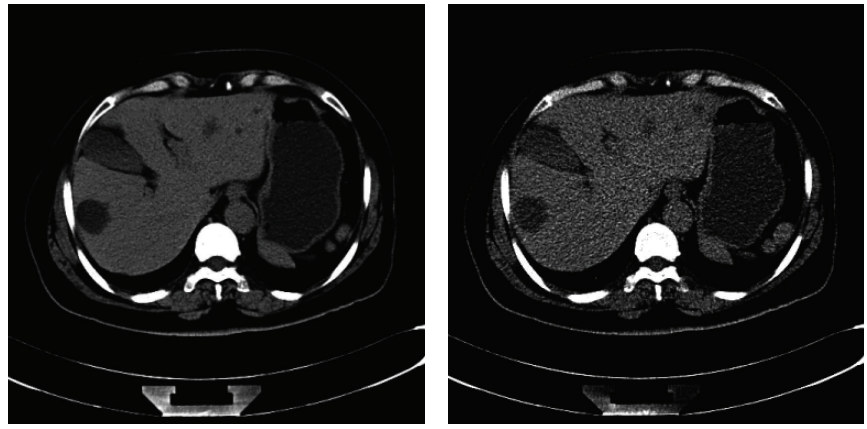
Table 1 summarized PSNR between the ideal reconstructed image and filtered reconstructed image. The PWBCD-NLMs has better performance in different noise levels in the term of PSNR than NLMs.

Table 1: PSNR for the test image.

Noise parameters	PSNR of the noisy image	PSNR of NLMs	PSNR of PWBCD-NLMs
$f_i = 2.5, T = 2e + 4$	23.29	34.19	**34.95**
$f_i = 3.5, T = 2e + 4$	21.88	33.79	**34.59**
$f_i = 4, T = 2e + 4$	21.30	33.45	**34.16**

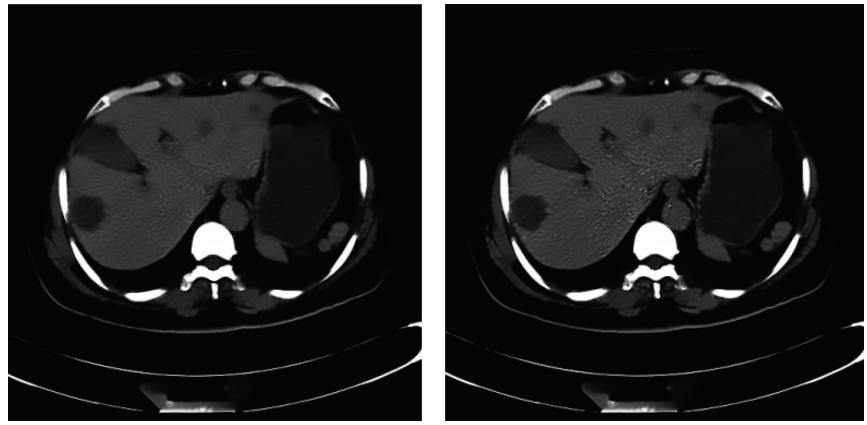
Figure 2 shows noisy test images and their reconstructed images using NLMs and the proposed method. Although the reconstructed images are very similar to each other, the reconstructed images using the new method also show better performance in edge preservation especially in weak and curve edge preserving than the NLMs. Since PWBCD-NLMs provides a more flexible way for handling different local image structures, it has much good performance in denoising while preserving structures.

One abdominal CT images of a 62-year-old woman were scanned from a 16 multidetector row CT unit (Somatom Sensation 16; Siemens Medical Solutions) using 120 kVp and 5 mm slice thickness. Other remaining scanning parameters are gantry rotation time, 0.5 second; detector configuration (number of detector rows section thickness), 16×1.5 mm; table feed per gantry rotation, 24 mm; pitch, 1 : 1; and reconstruction method, Filtered Back Projection (FBP) algorithm with the soft-tissue convolution kernel "B30f". Different CT doses were controlled by using two different fixed tube



(a) Original SDCT image with tube current time product 150 mAs

(b) Original LDCT image with tube current time product 60 mAs



(c) Reconstructed image from (b) using NLMs

(d) Reconstructed image from (b) using PWBCD-NLMs

Figure 3: (b) Real LDCT reconstructed image, (a) related SDCT reconstructed images and (c)-(d) reconstructed images from LDCT sinogram using NLMs and the new method.

currents 60 mAs for LDCT and 150 mAs (60 mA or 300 mAs) for SDCT, resp.). The CT dose index volumes (CTDIvol) for LDCT images and SDCT images are in positive linear correlation to the tube current and are calculated to be approximately ranged between 15.32 mGy and 3.16 mGy [18].

On sinogram space, the PWBCDs for images are computed according to Section 3.1 and the parameters are $r = 32$ and $\varepsilon = 2, 4, 8, 16, 32$. The new proposed method is compared with NLMs and their common parameters includes the standard deviation of noise $\sigma_N = 15$; the size of comparison window is 7×7 ($cr = 7$), while the size of searching patch is 21×21 ($sr = 21$). The other parameter for NLMswhich is the Gaussian kernel for weights defined on (13) is $h = 12$ and the parameters for the new method are the sizes of Gaussian kernel for two weights defined on (12): $h_1 = 15$ for the weights of difference between comparison window and $h_2 = 10$ for the weights between two PWBCDs.

Comparing the original SDCT images and LDCT images in Figure 3, we found that the LDCT images were severely degraded by nonstationary noise and streak artifacts. In Figure 3(d), for the proposed approach, experiments obtain more smooth images. Both in Figures 3(c) and 3(d), we can observe better noise/artifacts suppression and edge preservation than the LDCT image. Especially, compared to their corresponding original SDCT images, the fine features representing the hepatic cyst were well restored by using the proposed method. We can observe that the noise grains and artifacts were significantly reduced for the NLMs and PWBCD-NLMs processed LDCT images with suitable parameters both in Figures 3(c) and 3(d). The fine anatomical/pathological features can be well preserved compared to the original SDCT images (Figure 3(a)) under standard dose conditions.

5. Conclusions

In this paper, we propose a new PWBCD-NLMs method for LDCT imaging based on pointwise boxing-counting dimension and its new weight function. Since PWBCD can characterize the local structures of image well and also can be combined with NLMs easily, it provides a more flexible way

to balance the noise reduction and anatomical details preservation. Smoothing results for phantoms and real sinograms show that PWBCD-NLMs with suitable parameters has good performance in visual quality and PSNR.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (no 60873102), Major State Basic Research Development Program (no. 2010CB732501), and Open Foundation of Visual Computing and Virtual Reality Key Laboratory Of Sichuan Province (no. J2010N03). Ming Li also acknowledges the supports by the NSFC under the Project Grant nos. 61272402, 61070214 and 60873264 and the 973 plan under the Project Grant no. 2011CB302800.

References

[1] D. J. Brenner and E. J. Hall, "Computed tomography-an increasing source of radiation exposure," *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.

[2] J. Hansen and A. G. Jurik, "Survival and radiation risk in patients obtaining more than six CT examinations during one year," *Acta Oncologica*, vol. 48, no. 2, pp. 302–307, 2009.

[3] H. J. Brisse, J. Brenot, N. Pierrat et al., "The relevance of image quality indices for dose optimization in abdominal multidetector row CT in children: experimental assessment with pediatric phantoms," *Physics in Medicine and Biology*, vol. 54, no. 7, pp. 1871–1892, 2009.

[4] L. Yu, "Radiation dose reduction in computed tomography: techniques and future perspective," *Imaging in Medicine*, vol. 1, no. 1, pp. 65–84, 2009.

[5] J. Weidemann, G. Stamm, M. Galanski, and M. Keberle, "Comparison of the image quality of various fixed and dose modulated protocols for soft tissue neck CT on a GE Lightspeed scanner," *European Journal of Radiology*, vol. 69, no. 3, pp. 473–477, 2009.

[6] W. Qi, J. Li, and X. Du, "Method for automatic tube current selection for obtaining a consistent image quality and dose optimization in a cardiac multidetector CT," *Korean Journal of Radiology*, vol. 10, no. 6, pp. 568–574, 2009.

[7] A. Kuettner, B. Gehann, J. Spolnik et al., "Strategies for dose-optimized imaging in pediatric cardiac dual source CT," *RoFo*, vol. 181, no. 4, pp. 339–348, 2009.

[8] P. Kropil, R. S. Lanzman, C. Walther et al., "Dose reduction and image quality in MDCT of the upper abdomen: potential of an adaptive post-processing filter," *RoFo*, vol. 182, no. 3, pp. 248–253, 2009.

[9] H. B. Lu, X. Li, L. Li et al., "Adaptive noise reduction toward low-dose computed tomography," in *Proceedings of the Medical Imaging 2003: Physics of Medical Imaging, parts 1 and 2*, vol. 5030, pp. 759–766, February 2003.

[10] J. C. Giraldo, Z. S. Kelm, L. S. Guimaraes et al., "Comparative study of two image space noise reduction methods for computed tomography: bilateral filter and nonlocal means," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 3529–3532, 2009.

[11] H. B. 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. 1–4, pp. 1662–1666, November 2002.

[12] P. J. La Rivière, "Penalized-likelihood sinogram smoothing for low-dose CT," *Medical Physics*, vol. 32, no. 6, pp. 1676–1683, 2005.

[13] S. Hu, Z. Liao, and W. Chen, "Reducing noises and artifacts simultaneously of low-dosed X-ray computed tomography using bilateral filter weighted by Gaussian filtered sinogram," *Mathematical Problems in Engineering*, vol. 2012, Article ID 138581, 14 pages, 2012.

[14] S. Hu, Z. Liao, and W. Chen, "Sinogram restoration for low-dosed X-ray computed tomography using fractional-order Perona-Malik diffusion," *Mathematical Problems in Engineering*, vol. 2012, Article ID 391050, 13 pages, 2012.

[15] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[16] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 60–65, June 2005.

[17] A. Buades, B. Coll, and J. M. Morel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 123–139, 2008.

[18] C. Yang, C. Wufan, Y. Xindao et al., "Improving low-dose abdominal CT images by weighted intensity averaging over large-scale neighborhoods," *European Journal of Radiology*, vol. 80, no. 2, pp. e42–e49, 2011.

[19] Y. Chen, Z. Yang, W. Chen et al., "Thoracic low-dose CT image processing using an artifact suppressed largescale nonlocal means," *Physics in Medicine and Biology*, vol. 57, no. 9, pp. 2667–2688, 2012.

[20] Y. Chen, D. Gao, C. Nie et al., "Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior," *Computerized Medical Imaging and Graphics*, vol. 33, no. 7, pp. 495–500, 2009.

[21] 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.

[22] Z. Liao, S. Hu, M. Li, and W. Chen, "Noise estimation for single-slice sinogram of low-dose X-ray computed tomography using homogenous patch," *Mathematical Problems in Engineering*, vol. 2012, Article ID 696212, 16 pages, 2012.

[23] T. Thaipanich, B. T. Oh, P.-H. Wu, and C.-J. Kuo, "Adaptive nonlocal means algorithm for image denoising," in *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE '10)*, 2010.

[24] T. Thaipanich and C.-C. J. Kuo, "An adaptive nonlocal means scheme for medical image denoising," in *Proceedings of the SPIE Medical Imaging 2010: Image Processing*, vol. 7623, March 2010.

[25] R. Yan, L. Shao, S. D. Cvetkovic, and J. Klijn, "Improved nonlocal means based on pre-classification and invariant block matching," *Journal of Display Technology*, vol. 8, no. 4, pp. 212–218, 2012.

[26] A. K. Bisoi and J. Mishra, "On calculation of fractal dimension of images," *Pattern Recognition Letters*, vol. 22, no. 6-7, pp. 631–637, 2001.

[27] R. Creutzberg and E. Ivanov, "Computing fractal dimension of image segments," in *Proceedings of the 3rd International Conference of Computer Analysis of Images and Patterns (CAIP '89)*, 1989.

[28] M. Ghazel, G. H. Freeman, and E. R. Vrscay, "Fractal image denoising," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1560–1578, 2003.

[29] M. Ghazel, G. H. Freeman, and E. R. Vrscay, "Fractal-wavelet image denoising revisited," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2669–2675, 2006.

[30] B. Pesquet-Popescu and J. L. Vehel, "Stochastic fractal models for image processing," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 48–62, 2002.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 162105, 6 pages
http://dx.doi.org/10.1155/2013/162105

Research Article

Three-Dimensional Identification of Microorganisms Using a Digital Holographic Microscope

Ning Wu,[1] Xiang Wu,[2] and Tiancai Liang[3]

[1] *Shenzhen Key Lab of Wind Power and Smart Grid, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 518055, China*
[2] *School of Mechanical and Electrical Engineering, Harbin Institute of Technology, 92 West Dazhi Street, Nan Gang District, Harbin 150001, China*
[3] *GRG Banking Equipment Co., Ltd., 9 Kelin Road, Science Town, Guangzhou 510663, China*

Correspondence should be addressed to Xiang Wu; xiangwu@hit.edu.cn

Received 4 February 2013; Accepted 6 March 2013

Academic Editor: Shengyong Chen

This paper reports a method for three-dimensional (3D) analysis of shift-invariant pattern recognition and applies to holographic images digitally reconstructed from holographic microscopes. It is shown that the sequential application of a 2D filter to the plane-by-plane reconstruction of an optical field is exactly equivalent to the application of a more general filter with a 3D impulse response. We show that any 3D filters with arbitrary impulse response can be implemented in this way. This type of processing is applied to the two-class problem of distinguishing different types of bacteria. It is shown that the proposed technique can be easily implemented using a modified microscope to develop a powerful and cost-effective system with great potential for biological screening.

1. Introduction

In the past, high-resolution imaging of three-dimensional (3D) objects, or matter suspended in a volume of fluid, has mainly been accomplished using confocal microscopes [1]. In recent years, however, attention has returned to wide-field optical microscopy using coherent illumination and holographic recording techniques that exploit advances in digital imaging and image processing to compute 3D images. In contrast, with confocal imaging, coherent microscopy provides 3D information from a single recording that can be processed to obtain imaging modes analogous to dark field, phase or interference contrast as required [2–7]. In comparison with incoherent microscopes, a coherent instrument provides an image that can be focused at a later stage and can be considered as a microscope with an extended depth of field. For screening purposes, the increased depth of field is significant, particularly at high magnifications and high numerical aperture. For example a conventional, high magnification microscope has a depth of field of only a few microns whereas a comparable coherent instrument can have a depth of field of a few millimetres or so. This means that around 1000 times the volume of fluid can be screened from the information contained in a single digital recording [8].

The potential of coherent microscopes for automated biological screening is clearly dependent on the development of robust image or pattern recognition algorithms [9]. In essence, the application of pattern recognition techniques to coherent images is similar to that applied to their incoherent counterpart. The task can be defined as that of highlighting objects of interest (e.g., harmful bacteria) from other clutter (e.g., cell tissue and benign bacteria). This process should be accomplished regardless of position and orientation of the objects of interest within the image. It can be accomplished using variations on correlation processing. Linear correlation processing has been criticized in the past for its lack of rotation invariance and its inability to generalize in the manner of neural network classifiers; however, a cascade of correlators, separated by nonlinear (decision) layers, has considerably enhanced performance [5, 10]. Furthermore, we have shown that this is the architecture a neural network

classifier assumes if it is trained to provide a shift-invariant output [11, 12].

The application of linear correlation processing to the complex images recorded by a digital phase shifting interferometer has recently been demonstrated by Javidi and Tajahuerce [13]. Pattern recognition techniques implemented using a holographic microscope for the detection of microscale objects has also been considered by Dubois et al. [5, 14] In these works, the 3D sample field was reconstructed plane by plane and image classification was performed by the application of a 2D correlation filter to each of the reconstructed planes. It is noted, however, that although 2D correlation can be applied independently to different image planes it does not take into account the true nature of 3D optical fields, nor that the information in any two planes of these fields is, in fact, highly correlated [15].

In this paper, we considered, from first principles, 3D shift-invariant pattern recognition applied to optical fields reconstructed from digital holographic recordings. It will be shown that the sequential application of a 2D filter to plane-by-plane reconstructions is exactly equivalent to the application of a 3D filter to the full 3D reconstruction of the optical field. However, a linear filter designed based on the plane of focus will not necessarily work for planes out of focus, and therefore a 3D nonlinear filtering scheme is introduced into the optical propagation field. The 3D nonlinear filter is a system implemented with a general impulse response and followed by a nonlinear threshold. We will prove with experiment that a 3D nonlinear filtering structure can significantly improve the classification performance in 3D pattern recognition. In the experiment, we will apply the 3D nonlinear filter to 3D images of two types of bacteria recorded from a holographic microscope, and the enhanced classification performance will be shown.

2. Theory

Firstly, we define the 3D cross-correlation of complex functions $u(\mathbf{r})$ and $h(\mathbf{r})$ as

$$R(\mathbf{r}) = \int_{-\infty}^{+\infty} u(\mathbf{x})\, h(\mathbf{x}-\mathbf{r})\, d\mathbf{x}, \tag{1}$$

where $\mathbf{r}$ is a position vector and $d\mathbf{x}$ conventionally denotes the scalar quantity (dx, dy, dz). Assume that $H(\mathbf{k})$ and $U(\mathbf{k})$ are the Fourier transforms of $h(\mathbf{r})$ and $u(\mathbf{r})$, respectively; according to the convolution theorem, $R(\mathbf{r})$ can also be written:

$$R(\mathbf{r}) = \int_{-\infty}^{+\infty} U(\mathbf{k})\, H^*(\mathbf{k})\, e^{j2\pi\mathbf{k}\cdot\mathbf{r}} d\mathbf{k}, \tag{2}$$

where the superscript $*$ denotes complex conjugation. For pattern recognition purposes, (1) and (2) are equivalent ways to describe the process of correlation filtering defined in space domain and frequency domain, respectively.

It is clear from (1) and (2) that in general 3D correlation filtering requires 3D integration (in either the space or frequency domains). However, this is not the case when correlation filtering is applied to monochromatic optical fields propagating forward, typically the holographic reconstruction of optical fields by digital or optical means. In essence this is because $U(\mathbf{k})$ is nonzero only within an area of a 2D surface, and consequently $u(\mathbf{r})$ is highly correlated.

According to scalar diffraction theory, the complex amplitude $u(\mathbf{r})$ representing a monochromatic optical field propagation in a uniform dielectric must obey the Helmholtz equation [16] such that

$$\nabla^2 u(\mathbf{r}) + 4\pi^2 k^2 u(\mathbf{r}) = 0, \tag{3}$$

where k is a constant. Neglecting evanescent waves that occur close to boundaries and other obstructions, it is well known that the solutions to this equation are plane waves of the form:

$$u(\mathbf{r}) = A \exp\left(j2\pi\mathbf{k}\cdot\mathbf{r}\right), \tag{4}$$

where A is a complex constant. In these equations, k and $\mathbf{k}$ are the wave number and wave vector, respectively and are defined here such that $k = |\mathbf{k}| = 1/\lambda$, where λ is wavelength. In consequence, any monochromatic optical field propagating a uniform dielectric is described completely by the superposition of plane waves such that

$$u(\mathbf{r}) = \int_{-\infty}^{+\infty} U(\mathbf{k}) \exp\left(j2\pi\mathbf{k}\cdot\mathbf{r}\right) d\mathbf{k}, \tag{5}$$

where $U(\mathbf{k})$ is the spectral density, and $U(\mathbf{k})$ is the Fourier transform of $u(\mathbf{r})$ such that

$$U(\mathbf{k}) = \int_{-\infty}^{+\infty} u(\mathbf{r}) \exp\left(-j2\pi\mathbf{k}\cdot\mathbf{r}\right) d\mathbf{k}. \tag{6}$$

It is noted that because $u(\mathbf{r})$ consists of plane waves of single wavelength, the values of $U(\mathbf{k})$ only exist on an infinitely thin, spherical shell with a radius $k = |\mathbf{k}| = 1/\lambda$. In consequence, if a general 3D correlation filter with transfer function $H(\mathbf{k})$ is applied to a monochromatic optical field, $U(\mathbf{k})$, then in frequency domain the product $U(\mathbf{k})H^*(\mathbf{k})$ is also nonzero only on the spherical shell and consequently will obey the Helmholtz equation. If we expand (5), we have

$$\begin{aligned} &u\left(r_x, r_y, r_z\right) \\ &\quad = \iiint_{\infty} U\left(k_x, k_y, k_z\right) \exp\left(j2\pi\left(k_x r_x + k_y r_y + k_z r_z\right)\right) \delta \\ &\qquad \times \left(k_z \pm \sqrt{\frac{1}{\lambda^2} - k_x^2 - k_y^2}\right) dk_x dk_y dk_z \\ &\quad = \iint_{\infty} U\left(k_x, k_y, \pm\sqrt{\frac{1}{\lambda^2} - k_x^2 - k_y^2}\right) \\ &\qquad \times \exp\left(j2\pi\left(k_x r_x + k_y r_y \right.\right. \\ &\qquad\qquad \left.\left. \mp r_z \sqrt{\frac{1}{\lambda^2} - k_x^2 - k_y^2}\right)\right) dk_x dk_y. \end{aligned} \tag{7}$$

The square root in these equations represents light propagating through the xy plane in the positive and negative

z-directions, respectively. Since most holographic recordings record the flux in only one direction, we will consider only the positive root. According to (7), we can define $U_z(k_x, k_y)$ as the 2D projection of the spectrum onto the plane, $k_z = 0$, such that

$$U_z\left(k_x, k_y\right) = U\left(k_x, k_y, \sqrt{\frac{1}{\lambda^2} - k_x^2 - k_y^2}\right). \tag{8}$$

If $u_Z(r_x, r_y)$ represents the optical field in the plane $r_z = Z$, we have

$$\begin{aligned} u_Z\left(r_x, r_y\right) &= \iint_{\infty} U_Z\left(k_x, k_y\right) \exp\left(j2\pi Z\sqrt{\frac{1}{\lambda^2} - k_x^2 - k_y^2}\right) \\ &\quad \times \exp\left(j2\pi\left(k_x r_x + k_y r_y\right)\right) dk_x dk_y. \end{aligned} \tag{9}$$

In addition, taking the Fourier transform, we have

$$\begin{aligned} U_Z\left(k_x, k_y\right) &= \exp\left(-j2\pi Z\sqrt{\frac{1}{\lambda^2} - k_x^2 - k_y^2}\right) \\ &\quad \times \iint_{\infty} u_Z\left(r_x, r_y\right) \exp\left(-j2\pi\left(k_x r_x + k_y r_y\right)\right) dr_x dr_y. \end{aligned} \tag{10}$$

Equation (10) allows the spectrum to be calculated from the knowledge of the optical field propagating through a single plane. Equation (9) allows the field in any parallel plane to be calculated.

If we consider the application of a general 3D filter to the reconstruction of a propagating monochromatic field, we remember that the product $U(\mathbf{k})H^*(\mathbf{k})$ only exists on the surface of a sphere. Consequently, according to the derivation from (7) to (9), we have

$$\begin{aligned} R_Z\left(r_x, r_y\right) &= \int_{-\infty}^{+\infty} U_Z\left(k_x, k_y\right) H_z^*\left(k_x, k_y\right) \\ &\quad \times \exp\left(j2\pi Z\sqrt{\frac{1}{\lambda^2} - k_x^2 - k_y^2}\right) \\ &\quad \times \exp\left(j2\pi\left(r_x k_x + r_x k_y\right)\right) dk_x dk_y, \end{aligned} \tag{11}$$

where $R_Z(r_x, r_y)$ is the 3D correlation output in the plane $r_z = Z$, and

$$H_Z\left(k_x, k_y\right) = H\left(k_x, k_y, \sqrt{\frac{1}{\lambda^2} - k_x^2 - k_y^2}\right). \tag{12}$$

Finally, we note that in the space domain the correlation is

$$R_Z\left(r_x, r_y\right) = \int_{-\infty}^{+\infty} u_Z(u, v)\, h_Z\left(u - r_x, v - r_y\right) du\, dv, \tag{13}$$

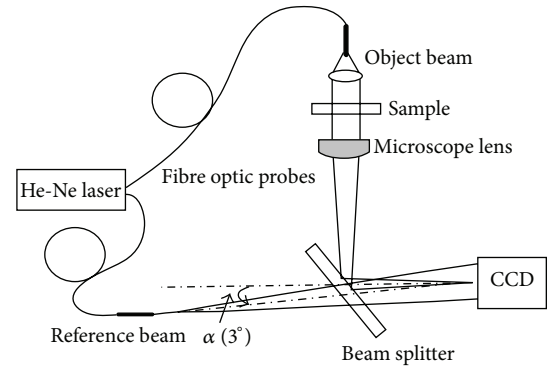


Figure 1: Holographic microscope with a coherent laser source.

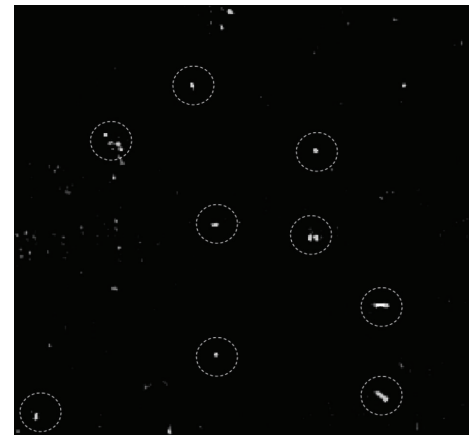


Figure 2: Holographic image with a field of view of 72 × 72 μm (absolute value shown).

where

$$\begin{aligned} h_Z\left(r_x, r_y\right) &= \int_{-\infty}^{+\infty} H_Z\left(k_x, k_y\right) \\ &\quad \times \exp\left(-j2\pi\left(r_x k_x + r_x k_y\right)\right) dk_x dk_y. \end{aligned} \tag{14}$$

Equation (13) shows that a single plane ($r_z = Z$) of the 3D correlation of a propagating optical field, $u(\mathbf{r})$, with a general impulse response function, $h(\mathbf{r})$, can be calculated as a 2D correlation of the field in that plane, $u_Z(r_x, r_y)$ with an impulse function, $h_Z(r_x, r_y)$, that is defined by (14).

In the recent literature, 2D correlation filtering has been applied to complex images reconstructed from a digital holographic microscope [14]. Practically, a digital holographic microscope measures the complex amplitude in the plane of focus, and the complex amplitude images in the parallel planes are calculated based on optical propagation theory. It is noted that a linear filter that is designed to perform well in one plane of focus will not necessarily perform well in another, and therefore a nonlinear filtering process is required.

When the 3D complex amplitude distribution of samples is reconstructed from the digital holographic recording, correlation filters can be applied for pattern recognition.

In the field of statistical pattern recognition, it is common to describe a digitized image of any dimension by the ordered variables in a vector [17], and we adopt this notation here. In this way, the discrete form of a complex 3D image can be written in vector notation by lexicographically scanning the 3D image array. Thus, an n-dimensional vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ represents a 3D image with n volume elements. We define a correlation operator $\widehat{H}$, with a filter kernel (or impulse response), $\mathbf{h} = [h_1, h_2, \ldots, h_n]^T$, is defined as,

$$\widehat{H}\mathbf{x} = \sum_{i=1}^{n} h^*_{i-n+1} x_i, \tag{15}$$

where the superscript "$*$" denotes the complex conjugate, and the subscript is taken to be modulo n such that

$$h_{n+a} = h_a. \tag{16}$$

A nonlinear threshold operator $\widehat{T}$ can be defined in the same way to operate on the individual components of a vector such that

$$\widehat{T}\mathbf{x} = \left[ax_1^3 + bx_1^2 + cx_1 + d, ax_2^3 + bx_2^2 + cx_2 + d, \ldots, ax_n^3 + bx_n^2 + cx_n + d\right]^T. \tag{17}$$

In general, image data from a hologram is a complex-amplitude field; however, we consider only the intensity distribution and define a modulus operator $\widehat{M}$ that operates on the the output such that

$$\widehat{M}\mathbf{x} = \left[|x_1|^2, |x_2|^2, \ldots, |x_n|^2\right]. \tag{18}$$

In this way, a 3D nonlinear filter, $\widehat{F}$, can be expressed as

$$\widehat{F} = \widehat{M}\widehat{T}_i\widehat{H}_i, \tag{19}$$

where the subscript to each operator denotes the layer in which a given operator is applied.

Without loss of generality, we design the 3D nonlinear filter to generate a delta function for the objects to be recognized, and zero outputs for the patterns to be rejected. For this purpose, we define a matrix set, $\mathbf{S}$, of m reference images such that $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_m]$ and the corresponding output matrix, $\mathbf{R}$, is given by

$$\mathbf{R} = \widehat{F}\mathbf{S}. \tag{20}$$

For the optimization of the 3D nonlinear filter, a matrix O with all the desired outputs intensity images is defined. In general, the desired outputs for in-class images will be a zero-valued vector with the first element set to be unit magnitude, and for an out-of-class image the desired output is zero. In order to train the filter with the desired performance, the error function below is requested to be minimized:

$$E = \sum_{i=1,\, j=1}^{n,m} \left(R_{ij} - O_{ij}\right)^2 + n\sum_{j=1}^{m} \left(R_{1j} - O_{1j}\right)^2, \tag{21}$$

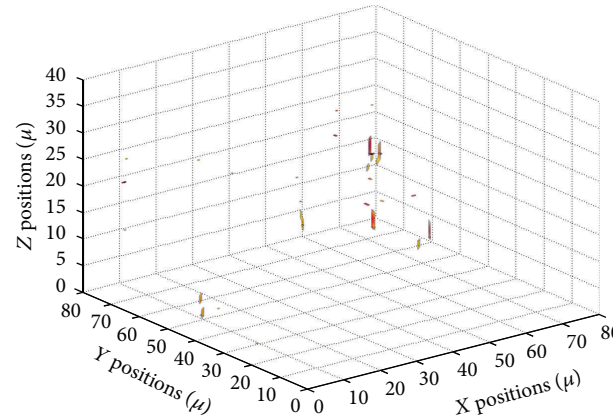


Figure 3: 3D image of the optical field reconstructed from Figure 2.

where R_{ij} and O_{ij} represent the ith pixel of the jth training and output image, respectively. The first term in this expression is the variance of the actual output from the desired output. The second term represents the signal peaks (that for simplicity are defined to be the first term in the output vector) and is given extra weight to ensure that they have the desired unit magnitude. Because (21) is a nonlinear function with a large number of variables, it is not possible to find an analytical solution. Hence, an iterative method is used in the minimization process. In this case, a simulated annealing algorithm was implemented in the optimization because it is more likely to reach a global minimum [18].

In the practical implementations of the 3D nonlinear filter described in this paper, we require a filter to identify the presence of fairly small objects in a relatively large field. In these cases, a relatively small filter kernel is used, and the kernel is zero-padded to the same size as the input image. In the test of this paper, the training images are selected to be $32 \times 32 \times 16$ elements, and we use 16×16 elements transfer function (2D). The filter output, the filter kernel, and the desired output images are all zero-padded to a resolution of $32 \times 32 \times 16$ elements. In this way, edge effects in pattern recognition for large images can be avoided.

3. Experiment

The objective of the work described in this section was to demonstrate 3D rotationally invariance pattern recognition based on digital holographic microscopy for the classification of two species of live bacteria, *E. coli* and *Pantoea*.

The digital holographic microscope setup used for this study is illustrated in Figure 1. In this arrangement, a He-Ne laser (633 nm) is used as coherent light source and is divided by a beam splitter and launched into a pair of optical fibres of equal length. One fibre supplies the light that forms the object beam for the holographic recording and is collimated. The microscope is used in a transmission mode and has an objective lens with 100x magnification and an oil immersion objective with an equivalent numerical aperture of NA = 1.25. The object plane is imaged onto a CCD array placed

(a) (b)

Figure 4: Typical bacteria (a) *E. coli* and (b) *Pantoea* in different rotated orientations.

approximately 200 mm from the objective. It is noted that because the microscope is holographic, the object of interest need not be located in the object plane.

The fibre that supplies the reference beam has an open termination that is arranged to diverge from a point in the rear focal plane of the microscope objective. In this way, the interference of the light from the reference beam and the light scattered is recorded at the CCD. Phase curvature introduced by the imaging process [19] is precisely matched by the reference curvature, and straight interference fringes are observed in the image plane in the absence of any scattering objects. From the analysis in Section 2, we can see that the interference pattern recorded by the CCD can be demodulated to give the complex amplitude describing the propagating field in the object plane. For reasons of processing, efficiency care was taken to adjust the magnification of the microscope to match the CCD resolution such that an optimally sampled (Nyquist) reconstruction is produced.

The holographic microscope is implemented with a flow cell that defines an experimental volume. The nutrient fluid with two species of living bacteria, *E. coli* and *Pantoea,* is syringed into the flow cell through a pipe. Figure 2 shows an image taken from the microscope corresponding to the absolute value of the complex amplitude in the object plane. In this image, the bacteria understood to be *E. coli* are highlighted with circles; some out-of-focus bacteria are invisible on this plane. Figure 3 shows a 3D image of the field in Figure 2 reconstructed using the method demonstrated in the above section.

In this study, a 3D nonlinear filter was trained to highlight live *E. coli* bacteria floating in the flow cell, while the *Pantoea* bacteria will be ignored. However, the reference set preparation is one of the most challenging problems for the identification of the living cells because each of the live bacteria varies in size and shape and appears at random orientation. To recognise the bacteria regardless of their shapes and orientations, adequate representative distortions of bacteria images must be provided for the 3D nonlinear filter as reference images.

The bacteria images registered as training set can be obtained by directly cropping the cell images from the 3D reconstructed field, or by simulating from the recorded images. For example, a selected bacteria image can be rotated to generate several orientation versions. Figure 4(a) shows eight absolute value images of a typical rod-shaped *E. coli* rotated in steps of 45 degrees. *Pantoea* bacteria have a similar rod shape, but slightly different in size from *E. coli*. Figure 4(b) shows one of the selected *Pantoea* in eight different rotated versions.

To demonstrate the performance of the 3D nonlinear filter, we train the system to detect *E. coli* bacteria with 42 images, including 25 *E. coli* and 17 *Pantoea* images, and the filer is tested with the complex amplitude image in Figure 2. Figure 5 shows the 3D image of the 3D filter output. Figure 6 reports the projection of the output volume onto a plane. It can be seen that most of the *E. coli* bacteria had been highlighted by correlation peaks and the *Pantoea* had been ignored. However, a small portion of the *E. coli* cannot be detected; this is because the training set with limited number of reference images does not represent all the distortions and orientations of the bacteria. It is expected that classification rate can be improved if more reference images are included in the training set.

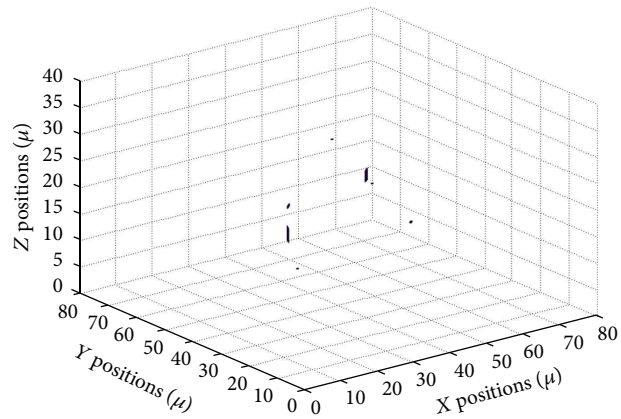


Figure 5: 3D output for the 3D nonlinear filter trained to recognize *E. coli* (absolute amplitude value shown).

Figure 6: The projection of the output volume (absolute amplitude value shown).

4. Conclusion

This paper describes 3D pattern recognition with a 3D nonlinear filter applied to monochromatic optical fields that can be recorded and reconstructed by holographic microscopes. The 3D extension and formulation of the nonlinear filter concept has been introduced. We have shown with experimental data that the 3D nonlinear filtering system provides additional capability as a means to perform 3D pattern recognition in a shift and rotationally invariant means. We demonstrate this in practice by applying the 3D nonlinear filter to a holographic recording of the light scattered from two kinds of living bacteria suspended in water. The experimental data demonstrated that the 3D nonlinear filter has good shift and rotationally invariant property in 3D space.

Acknowledgment

Financial support from The Research Fund for the Doctoral Program of Higher Education (No. 20122302120072) to initiate this research is gratefully acknowledged.

References

[1] M. Minsky, "Memoir on inventing the confocal scanning microscope," *Scanning*, vol. 10, no. 4, pp. 128–138, 1988.

[2] U. Schnars and W. P. O. Jüptner, "Digital recording and numerical reconstruction of holograms," *Measurement Science and Technology*, vol. 13, no. 9, pp. R85–R101, 2002.

[3] T. Zhang and I. Yamaguchi, "Three-dimensional microscopy with phase-shifting digital holography," *Optics Letters*, vol. 23, no. 15, pp. 1221–1223, 1998.

[4] E. Cuche, P. Marquet, and C. Depeursinge, "Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms," *Applied Optics*, vol. 38, no. 34, pp. 6994–7001, 1999.

[5] F. Dubois, L. Joannes, and J. C. Legros, "Improved three-dimensional imaging with a digital holography microscope with a source of partial spatial coherence," *Applied Optics*, vol. 38, no. 34, pp. 7085–7094, 1999.

[6] 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, 3 pages, 2006.

[7] Z. Teng, A. J. Degnan, U. Sadat et al., "Characterization of healing following atherosclerotic carotid plaque rupture in acutely symptomatic patients: an exploratory study using in vivo cardiovascular magnetic resonance," *Journal of Cardiovascular Magnetic Resonance*, vol. 13, no. 1, article 64, 2011.

[8] L. Lin, S. Chen, Y. Shao, and Z. Gu, "Plane-based sampling for ray casting algorithm in sequential medical images," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 874517, 5 pages, 2013.

[9] Q. Guan and B. Du, "Bayes clustering and structural support vector machines for segmentation of carotid artery plaques in multi-contrast MRI," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 549102, 6 pages, 2012.

[10] F. Dubois, "Nonlinear cascaded correlation processes to improve the performances of automatic spatial-frequency-selective filters in pattern recognition," *Applied Optics*, vol. 35, no. 23, pp. 4589–4597, 1996.

[11] S. Reed and J. Coupland, "Statistical performance of cascaded linear shift-invariant processing," *Applied Optics*, vol. 39, no. 32, pp. 5949–5955, 2000.

[12] N. Wu, R. D. Alcock, N. A. Halliwell, and J. M. Coupland, "Rotationally invariant pattern recognition by use of linear and nonlinear cascaded filters," *Applied Optics*, vol. 44, no. 20, pp. 4315–4322, 2005.

[13] B. Javidi and E. Tajahuerce, "Three-dimensional object recognition by use of digital holography," *Optics Letters*, vol. 25, no. 9, pp. 610–612, 2000.

[14] F. Dubois, C. Minetti, O. Monnom, C. Yourassowsky, J. C. Legros, and P. Kischel, "Pattern recognition with a digital holographic microscope working in partially coherent illumination," *Applied Optics*, vol. 41, no. 20, pp. 4108–4119, 2002.

[15] S. Chen and M. Zhao, "Recent advances in morphological cell image analysis," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 101536, 10 pages, 2012.

[16] A. Sommerfeld, *Partial Differential Equations in Physics*, Academic Press, New York, NY, USA, 1949.

[17] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, NY, USA, 1972.

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

[19] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, NY, USA, 1968.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 248380, 10 pages
http://dx.doi.org/10.1155/2013/248380

Research Article

Thresholded Two-Phase Test Sample Representation for Outlier Rejection in Biological Recognition

Xiang Wu[1] and Ning Wu[2]

[1] *Harbin Institute of Technology, 92 West Dazhi Street, Nan Gang District, Harbin 150001, China*
[2] *Shenzhen Key Lab of Wind Power and Smart Grid, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 518055, China*

Correspondence should be addressed to Ning Wu; aning.wu@gmail.com

Received 22 January 2013; Accepted 9 February 2013

Academic Editor: Carlo Cattani

The two-phase test sample representation (TPTSR) was proposed as a useful classifier for face recognition. However, the TPTSR method is not able to reject the impostor, so it should be modified for real-world applications. This paper introduces a thresholded TPTSR (T-TPTSR) method for complex object recognition with outliers, and two criteria for assessing the performance of outlier rejection and member classification are defined. The performance of the T-TPTSR method is compared with the modified global representation, PCA and LDA methods, respectively. The results show that the T-TPTSR method achieves the best performance among them according to the two criteria.

1. Introduction

Object recognition has become a hot topic in the field of computer vision and pattern recognition in recent years, and many approaches have been proposed for face image classification with a given database. One type of the methods is to reduce the dimensionality of sample by extracting the feature vector with linear transformation methods, such as the principal component analysis (PCA) [1–3] and the linear discriminant analysis (LDA) [4, 5]. In the PCA method, the training samples and the testing samples are transformed from the original sample space into a space with the maximum variance of all the samples, while the LDA method converts the samples to a feature space where the distances of the centers of different classes are maximized. In these two transformation methods, both the training samples and the testing samples have their corresponding representations in the new feature space, and the classification is carried out based on the distance between the representations related to the training set and the testing set.

Another type of transformation-based method was proposed to focus on local information of the training samples. Instead of using the whole training set, this type of method only uses part of the samples, since the performance of the classifier is usually limited within some local areas. By concentrating on the local distribution of training data, the design and testing of the classifier can be much more efficient than the global methods [6]. Typical examples of local LDA methods include the method for multimodal data projection [7, 8] and the approach to use the local dependencies of samples for classification [9]. It is also found that the local PCA is more efficient than the global PCA in feature extraction [10] or sample clustering [11].

In recent years, the sparse representation theory has been applied to pattern recognition problems and has drawn a lot of attentions [12–21]. The sparse representation method also uses only part of the training data for classification by linearly representing a testing sample with the training set, and part of the linear combination coefficients is set to zero. The classification criterion of the sparse representation method is based on the biggest contribution from the sample classes during the linear representation.

In a recent study, a two-phase test sample representation (TPTSR) method was proposed for face recognition [22]. In this method, classification process is divided into two steps: the first step selects M-nearest neighbors of the testing sample

from the training set by using linear representation method and the second step processes the selected M samples further by using them to linearly represent the testing sample. The classification result is based on the linear contribution of the classes among the M-nearest neighbors in the second phase of the TPTSR. By selecting M-closest neighbors from the training set for further processing, the TPTSR method identifies a local area that may contain the target class sample, reducing the risk of misclassification because of a similar nontarget sample.

Even the TPTSR method has been proven to be very useful in face classification; however, for face recognition applications with outliers the classification emphasis is different and the performance measurement criterion is also new. In face recognition problems with outliers, like security registration systems, only a small and particular group of members is required to be classified and compared with a large population of irrelevant people or intruders. In the application of identifying wanted criminals at airports, train station and other public places, the classifier is also required to identify a minor number of target members from a large number of irrelevant passengers. In previous studies, the approaches for pattern classification with outliers include two main methods, one is to train the classifier with only the member samples, and the other is to take into account a small number of outliers as a separate class in the training set [23]. However, neither of the methods can guarantee a low false alarm rate while maintaining a reasonable recognition rate for members.

In this paper, we further develop the TPTSR method by applying a threshold in the classification process for outlier rejection and member classification, and it is referred to as thresholded TPTSR (T-TPTSR) method. In the T-TPTSR, the distance between the testing sample and the weighted contribution of the target class in the second-phase linear representation is measured and compared with a threshold, by which an outlier will be identified. In this study, we also propose two different criteria for assessing the performance of classifier for outlier rejection as well as member classification, and, based on these criteria, we test the thresholded global representation (T-GR) method, thresholded PCA (T-PCA) method, and thresholded LDA (T-LDA) method, respectively. The test results show that the T-TPTSR achieves better performance in rejecting the outliers while maintaining outstanding classification rate for members.

In Sections 2 and 3 of this paper, we will introduce the theory of the T-TPTSR, T-GR, T-PCA, and T-LDA, respectively. Section 4 presents our experimental results with different face image databases, and finally a conclusion will be drawn in Section 5.

2. Thresholded Two-Phase Test Sample Representation (T-TPTSR)

In this section, the TTPTSR method will be introduced with a threshold applied to the second-phase output in the classification process.

2.1. First Phase of the T-TPTSR with M-Nearest Neighbor Selection. The first phase of the T-TPTSR is to select M-nearest neighbors from all the training samples for further processing in the second phase, narrowing the sample space down to a local area for the target class [22]. The M-nearest neighbors are selected by calculating the weighted distances of the testing sample from each of the training samples. Firstly, let us assume that there are L classes and n training images, $x_1, x_2, \ldots, x_n$, and if some of these images are from the jth class ($j = 1, 2, \ldots, L$), then j is their class label. It is also assumed that a test image y can be written in the form of linear combination of all the training samples, such as

$$y = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n, \tag{1}$$

where a_i ($i = 1, 2, \ldots, n$) is the coefficient for each training image x_n. Equation (1) can also be written in the form of vector operation, such as

$$y = XA, \tag{2}$$

where $A = [a_1 \cdots a_n]^T$, $X = [x_1 \cdots x_n]^T . x_1 \cdots x_n$, and y are all column vectors. If X is a singular square matrix, (2) can be solved by using $A = (X^T X + \mu I)^{-1} X^T y$, or it can be solved by using $A = X^{-1} y$, where μ is a small positive constant and I is the identity matrix. In our experiment with the T-TPTSR method, μ in the solution is set to be 0.01.

By solving (2), we can represent the testing image using the linear combination of the training set as shown in (1), which means that the testing image is essentially an approximation of the weighted summation of all the training images, and the weighted image $a_i x_i$ is a part of the approximation. In order to measure the distance between the training image x_i and the testing image y, a distance metric is defined as followed:

$$e_i = \left\| y - a_i x_i \right\|^2, \tag{3}$$

where e_i is called the distance function, and it gives the difference between the testing sample y and the training sample x_i. It is clear that a smaller value of e_i means that the ith training sample is closer to the testing sample, and it is more probable to be the member of the target class. These M-nearest neighbors are chosen to be processed further in the second phase of the T-TPTSR where the final decision will be made within a much smaller sample space. We assume that the M-nearest neighbors selected are denoted as $x_1 \cdots x_M$, and the corresponding class labels are $C = \{c_1 \cdots c_M\}$, where $c_i \in \{1, 2, \ldots, L\}$. In the second phase of the T-TPTSR, if a sample x_p's class label does not belong to C, then this class will not be considered as a target class, and only a class from C will be regarded as a potential target class.

2.2. Second Phase of the T-TPTSR for Outlier Rejection. In the second phase of the T-TPTSR method, the M-nearest neighbors selected from the first phase are further calculated to obtain a final decision for the recognition task. We represent the testing sample with the linear combination of the training samples again, but only with the M-nearest

neighbors selected from the first phase. If the M-nearest neighbors selected are denoted as $x_1 \cdots x_M$, and their linear combination for the approximation of the testing image y is assumed to be satisfied, such as

$$y = b_1 x_1 + \cdots + b_M x_M, \tag{4}$$

where b_i $(i = 1, 2, \ldots, M)$ are the coefficients. In vector operation form, (4) can be written as

$$y = \widetilde{X} B, \tag{5}$$

where $B = [b_1 \cdots b_M]^T$, and $\widetilde{X} = [x_1 \cdots x_M]$. In the same philosophy as above, if $\widetilde{X}$ is a nonsingular square matrix, (5) can be solved by

$$B = \left(\widetilde{X}\right)^{-1} y, \tag{6}$$

or, otherwise, B can be solved by

$$B = \left(\widetilde{X}^T \widetilde{X} + \gamma I\right)^{-1} \widetilde{X}^T y, \tag{7}$$

where γ is a positive small value constant, and it is usually set to 0.01, and I is the identity matrix.

When we obtain the coefficients b_i for each of the nearest neighbors, the contribution of each of the classes to the testing image will be measured, and the classification output will be based on the distance between the contribution and the testing image. If the nearest neighbors $x_s \cdots x_t$ are from the rth class $(r \in C)$, and the linear contribution to approximate the testing sample by this class is defined as

$$g_r = b_s x_s + \cdots + b_t x_t. \tag{8}$$

The measurement of the distance between the testing sample and the rth class samples in the M-nearest neighbors is calculated by the deviation of g_r from y, such as

$$D_r = \left\| y - g_r \right\|^2, \quad r \in C. \tag{9}$$

It is clear that a smaller value of D_r means a better approximation of the training samples from the rth class for the testing sample, and thus the rth class will have a higher possibility over other classes to be the target class. However, if outliers are considered, a threshold must be applied to the classification output to differentiate the members of class from outliers, such as

$$D_k = \min D_r < T \quad (k, r \in C; T \in [0, +\infty)), \tag{10}$$

where T is the threshold. If $D_k \geq T$, the testing sample will be regarded as an outlier and therefore will be rejected. Only when $D_k < T$, the testing sample y can be classified to the kth class with the smallest deviation from y.

In the second phase of the T-TPTSR, the solution in (6) or (7) finds the coefficients for the linear combination of the M-nearest neighbors to approximate the testing sample, and the training class with the minimum deviation of the approximation will be considered as the target class for the testing sample. However, the value of the minimum deviation must be less than the threshold T. If the minimum distance between the testing sample and the member class's approximations is greater than the threshold T, the testing sample will be classified as an outlier and thus rejected. However, if the value of the minimum deviation of the linear combinations to an outlier is less than the threshold T, this outlier will be classified into the member class with the minimum deviation, and a misclassification will occur. Likewise, if a testing image belongs to a member class, but the minimum deviation from the linear combinations of each of the classes is greater than the threshold T, this testing image will be classified as an outlier, and a false alarm is resulted. Since the samples used in the T-TPTSR method are all normalized in advanced, the value of D_r in (9) will be within a certain range, such that $0 \leq D_r \leq s$, where $s \approx 1$, and therefore it is practical to determine a suitable threshold for the identification task before the testing.

3. The T-GR, T-PCA, and T-LDA Methods for Outlier Rejection

As a performance comparison with the T-TPTSR method, in the following section, we also introduce the modified versions of the GR, PCA, and LDA methods, respectively, for outlier rejection and member classification in face recognition.

3.1. The T-GR Method. The thresholded global representation (T-GR) method is essentially the T-TPTSR method with all the training samples that are selected as the M-nearest neighbors (M is selected as the number of all the training samples), and it also finds the target class directly by calculating the best representing sample class for the testing image.

In the T-GR method, the testing sample is represented by the linear combination of all the training samples, and the classification is not just based on the minimum deviation of the linear contribution from each of the classes to the testing sample, but also based on the value of the minimum deviation. If the minimum deviation is greater than the threshold applied, the testing sample will be identified as an outlier.

3.2. The T-PCA Method. The PCA method is based on linearly projecting the image space onto a lower-dimensional feature space, and the projection directions are obtained by maximizing the total scatter across all the training classes [24, 25]. Again, we assume that there are L classes and n training images, $x_1, x_2, \ldots, x_n$, each of which is m-dimensional, where $n < m$. If a linear transformation is introduced to map the original m-dimensional image space into an l-dimensional feature space, where $l < m$, the new feature vector $u_i \in R^l$ can be written in the form of

$$u_i = W^T x_i \quad (i = 1, 2, \ldots, n), \tag{11}$$

where $W^T \in R^{m \times l}$ is a matrix with orthonormal columns. If the total scatter matrix S^T is defined as

$$S^T = \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T, \tag{12}$$

where $\mu \in R^m$ is the mean of all the training samples, we can see that, after applying the linear transformation W^T, the scatter of all the transformed feature vectors $u_1, u_2, \ldots, u_n$ is $W^T S^T W$, which can be maximized by finding a projection direction W_m, such as

$$\begin{aligned} W_m &= \arg\max_W W^T S^T W \\ &= [w_1, w_2 \cdots w_l], \end{aligned} \tag{13}$$

where w_i $(i = 1, \ldots, l)$ is the set of m-dimensional eigenvectors of S^T corresponding to the l biggest eigenvalues. During the recognition process, both the testing sample y and all the training samples are projected into the new feature space via W_m before the distance between them is calculated, such as

$$\begin{aligned} D_i &= \left\| W_m^T y - W_m^T x_i \right\|^2 \\ &= \left\| W_m^T (y - x_i) \right\|^2 \quad (i = 1, 2, \ldots, n). \end{aligned} \tag{14}$$

In the thresholded PCA method, the testing sample y will be classified to the class whose member has the minimum distance D_i, but this distance must be less than the threshold T, such that

$$D_k = \min D_i < T \quad (k, i = 1, 2, \ldots, n;\ T \in [0, +\infty)). \tag{15}$$

The testing sample y whose corresponding minimum distance D_k is less than the threshold T will be classified as an outlier and therefore rejected; otherwise y will be classified into the class with x_k.

3.3. The T-LDA Method. The LDA is a class-specific linear method for dimensionality reduction and simple classifiers in a reduced feature space [26–29]. The LDA method also finds a direction to project the training images and testing images into a lower dimension space, on the condition that the ratio of the between-class scatter and the within-class scatter is maximized.

Likewise, if there are L classes and n training images, $x_1, x_2, \ldots, x_n$, each of which is m-dimensional, where $n < m$, and in the ith class there are N_i samples $(i = 1, 2, \ldots, L)$, the between-class scatter matrix can be written as

$$S_b = \sum_{i=1}^{L} N_i (\mu_i - \mu)(\mu_i - \mu)^T, \tag{16}$$

and the within-class scatter matrix can be defined as

$$S_w = \sum_{i=1}^{L} \sum_{j=1}^{N_i} (x_j - \mu_i)(x_j - \mu_i)^T, \tag{17}$$

where μ_i is the mean image of the ith class, and μ is the mean of all the samples. It is noted that S_w must be nonsingular in order to obtain an optimal projection matrix W_m with the orthonormal columns to maximize the ratio of the determinant of the projected S_b and projected S_w, such that

$$\begin{aligned} W_m &= \arg\max_W \frac{\left| W^T S_b W \right|}{W^T S_w W} \\ &= [w_1 w_2 \cdots w_l], \end{aligned} \tag{18}$$

where w_i $(i = 1, \ldots, l)$ is the set of m-dimensional generalized eigenvectors of S_b and S_w corresponding to the l biggest eigenvalues, such as

$$S_b w_i = \lambda_i S_w w_i, \quad (i = 1, 2, \ldots, l), \tag{19}$$

where λ_i $(i = 1, \ldots, l)$ is the l generalized eigenvalues. Since there are the maximum number of $L - 1$ nonzero generalized eigenvalues available, the maximum l can only be $L - 1$.

The distance between the projection of the testing sample y and the training samples with W_m in the new feature space is calculated as

$$\begin{aligned} D_i &= \left\| W_m^T y - W_m^T x_i \right\|^2 \\ &= \left\| W_m^T (y - x_i) \right\|^2 \quad (i = 1, 2, \ldots, n). \end{aligned} \tag{20}$$

If the sample x_k's projection into the feature space has a minimum distance from the projection of the testing sample y, the testing sample will be classified into the same class as x_k, such that

$$D_k = \min D_i < T \quad (k, i = 1, 2, \ldots, n;\ T \in [0, +\infty)), \tag{21}$$

where T is a threshold to screen out the outliers. For the threshold LDA method, all the target members' projection distance D_i must be less than T, or otherwise they will be classified as outliers and rejected.

4. Experimental Results

In this experiment, we test the performance of the T-TPTSR, the T-GR, the T-PCA, and the T-LDA methods for outlier rejection and member classification, respectively. One of the measurement criteria for comparing the performance of these methods is to find the minimum overall classification error rate. During the classification task, an optimal threshold T can be found for the above methods so that the overall classification error rate is minimized. The overall classification error rate is calculated based on three classification error rates, such as the misclassifications among member's classes (when the testing sample is a member and $D_k < T$, but misclassified as another class), the misclassifications of a member to outlier's group (when the testing sample is a member but $D_k > T$, and thus misclassified), and misclassifications for outliers (when the testing sample is an outlier but $D_k < T$, and therefore accepted wrongly as a member). If $\mathrm{ERR}_{\mathrm{overall}}(T)$ represents the overall classification error rate as a function of the threshold T, $\mathrm{ERR}_{\mathrm{member}}(T)$ denotes the classification error rate for errors that occurred among members (misclassifications recorded for testing

samples from member's group versus the total number of testing samples from member's group), and $\mathrm{ERR}_{\mathrm{outlier}}(T)$ is the misclassification rate for outliers (classification errors recorded for testing samples from the outlier's group versus the total number of testing outliers), their relationship can be written as

$$\mathrm{ERR}_{\mathrm{overall}}(T) = \mathrm{ERR}_{\mathrm{member}}(T) + \mathrm{ERR}_{\mathrm{outlier}}(T). \quad (22)$$

It is noted that the value of $\mathrm{ERR}_{\mathrm{member}}$ varies with the threshold T, and when $T = 0$, $\mathrm{ERR}_{\mathrm{member}}$ takes the value of 100%, and it generally decreases when the value of T increases until it reaches a constant classification error rate. The classification error rate for outlier also changes its value according to the threshold T, however, $\mathrm{ERR}_{\mathrm{outlier}} = 0\%$ when $T = 0$, and its value increases until reaching 100%. The minimum $\mathrm{ERR}_{\mathrm{overall}}(T)$ can be found between the range of $T = 0$ and $T = T_m$, where $\mathrm{ERR}_{\mathrm{member}}(T)$ becomes a constant, or $\mathrm{ERR}_{\mathrm{overall}}(T)$ reaches 100%, such that

$$\mathrm{ERR}_{\mathrm{opt}} = \min \mathrm{ERR}_{\mathrm{overall}}(T), \quad T \in [0, +\infty). \quad (23)$$

The value of $\mathrm{ERR}_{\mathrm{opt}}$ is an important criterion showing the performance of a classifier for both of outlier rejection and member recognition.

Another measuring criterion for measuring the performance of the thresholded classifiers is the receiver operation characteristics (ROC) curve, which is a graphical plot of the true positive rate (TPR) versus the threshold T in the application of thresholded classification for outlier rejection. We firstly define the true positive detection rate for the outliers, $\mathrm{TPR}_{\mathrm{outlier}}(T)$, and it can be written in the form of the classification error rate for the outliers, such that

$$\mathrm{TPR}_{\mathrm{outlier}}(T) = 100\% - \mathrm{ERR}_{\mathrm{outlier}}(T), \quad T \in [0, +\infty). \quad (24)$$

We also define the false alarm rate caused in the member' group as a function of the threshold, $\mathrm{ERR}_{\mathrm{FA}}(T)$, which is the number of errors recorded for misclassifying a member to an outlier versus the number of testing samples from the member's group. An optimal classifier for outlier rejection and member classification needs to find a suitable threshold T so that the $\mathrm{TPR}_{\mathrm{outlier}}(T)$ can be maximized as well as the $\mathrm{ERR}_{\mathrm{FA}}(T)$ can be minimized. Therefore, the following function $D_{O\text{-}F}(T)$ is defined for this measurement, such that

$$\begin{aligned} D_{O\text{-}F}(T) &= \mathrm{TPR}_{\mathrm{outlier}}(T) - \mathrm{ERR}_{\mathrm{FA}}(T) \\ &= 100\% - \mathrm{ERR}_{\mathrm{outlier}}(T) \\ &\quad - \mathrm{ERR}_{\mathrm{FA}}(T), \quad T \in [0, +\infty). \end{aligned} \quad (25)$$

It is obvious that $D_{O\text{-}F}(T)$ is required to be maximized so that a classifier can be optimized for both outlier rejection and member classification, such that

$$D_{\mathrm{opt}} = \max D_{O\text{-}F}(T), \quad T \in [0, +\infty), \quad (26)$$

and the value of D_{opt} is an important metric for comparing the performance of classifier for outlier rejection analysis.

Figure 1: Part of the face images from the Feret database for testing.

The minimum overall classification error rates $\mathrm{ERR}_{\mathrm{opt}}$ and the maximum difference of the true positive outlier recognition rate and the false-alarm rate D_{opt} are essentially the same performance assessment metric for a classifier with outlier rejection. The difference is that the overall classification error rate represents the efficiency of member classification, while $D_{O\text{-}F}$ and D_{opt} show the performance of outlier rejection. In the following experiment, we test and compare the minimum overall classification error rates $\mathrm{ERR}_{\mathrm{opt}}$ and the maximum D_{opt} of the T-TPTSR, T-GR, T-PCA, and T-LDA methods, respectively, and based on these two criteria we find the optimal classifier for outlier rejection and member classification.

In our experiment, we test and compare the performance of the above methods using the online face image databases Feret [30, 31], ORL [32], and AR [33], respectively. These databases provide face images from different faces with different facial expression and facial details under different lighting conditions. The Feret database provides 1400 face images from 200 individuals for the training and testing, and there are 7 face images from each of the classes. In the AR database, there are totally 3120 face images from 120 people, each of which provides 26 different facial details. For the ORL database, there are 400 face images from 40 different individuals, each of which has 10 face images.

In this experiment, the training set and the testing set are selected randomly from each of the individuals. For each of the databases, the people included are divided into two groups and one is member's group and the other is outlier's group. For individuals chosen as the member's class, the training samples are prepared by selecting some of their images from the database, and the rest of the images are taken as the testing set. For the outliers that is supposed to be outside the member's group, there is no training set for the classification, and all the samples included in the outlier's group are taken as the testing set.

We firstly test the Feret database with the above outlier rejection methods. The Feret database is divided into two groups, 100 members from the 200 individuals are randomly selected into the member's group, and the rest of the 100 individuals are the outliers in the test. For each of the 100

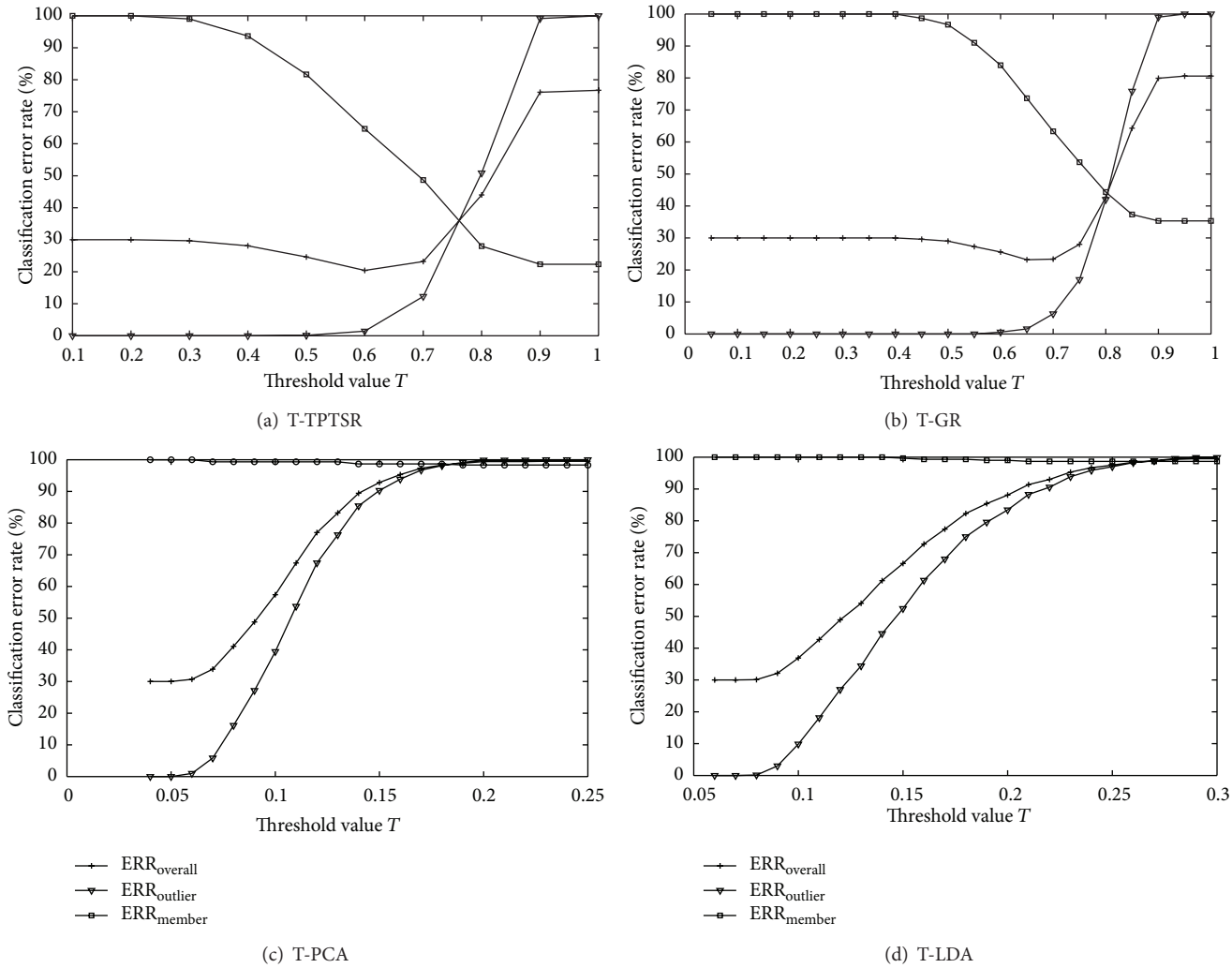


Figure 2: Classification error rates for outliers, members, and overall of (a) the T-TPTSR method, (b) the T-GR method, (c) the T-PCA method, and (d) the T-LDA method, respectively, on the Feret database.

member classes, 4 images out of 7 are selected randomly as the training set, and the rest of the 3 images are for the testing set. For the 100 individuals in the outlier's group, all 7 images from each of them are the testing set for the classification task. Therefore, there are 400 training images and 1000 testing images in this test, and, among the testing images, there are 300 images from member's group and 700 images from outlier's group. Figure 1 shows part of the member and outlier's images from the Feret database for the testing, and all the images have been resized to a 40×40-pixel image by using a downsampling algorithm [34]. Since the number of classes in the Feret database is much more than the ORL and AR databases, also the number of training images is less, and the resolution of the images is lower, the testing with the Feret database would be more challenging and the result is generally regarded as more convincing.

In the test of the T-TPTSR method with the Feret database, the number of nearest neighbors M selected for the first-phase processing is 60 (according to the empirical data, the optimal number M is selected about 10~15% of the number of training samples). In the test with the above methods, the threshold value T varies from 0 to a constant that can result in 100% of $ERR_{outlier}$ with the interval of 0.1 or 0.5, where all outliers are accepted as members. Figures 2(a)~2(d) show different classification error rates of the above methods as the function of the threshold T, respectively. It can be seen that the ERR_{opt} values of the T-TPTSR method and the T-GR method are much lower than the T-PCA and T-LDA methods, and the ERR_{member} curves of the T-TPTSR and T-GR decrease from 100% to a much lower constant than those of the T-PCA and T-LDA when the threshold T increases. The second row of Table 1 lists all the ERR_{opt} values shown in Figure 2, and we can see that the T-TPTSR method achieves the lowest overall classification error rate. Figure 3 shows the ROC curves of the T-TPTSR, T-GR, T-PCA and T-LDA methods, respectively, and the third row of Table 1

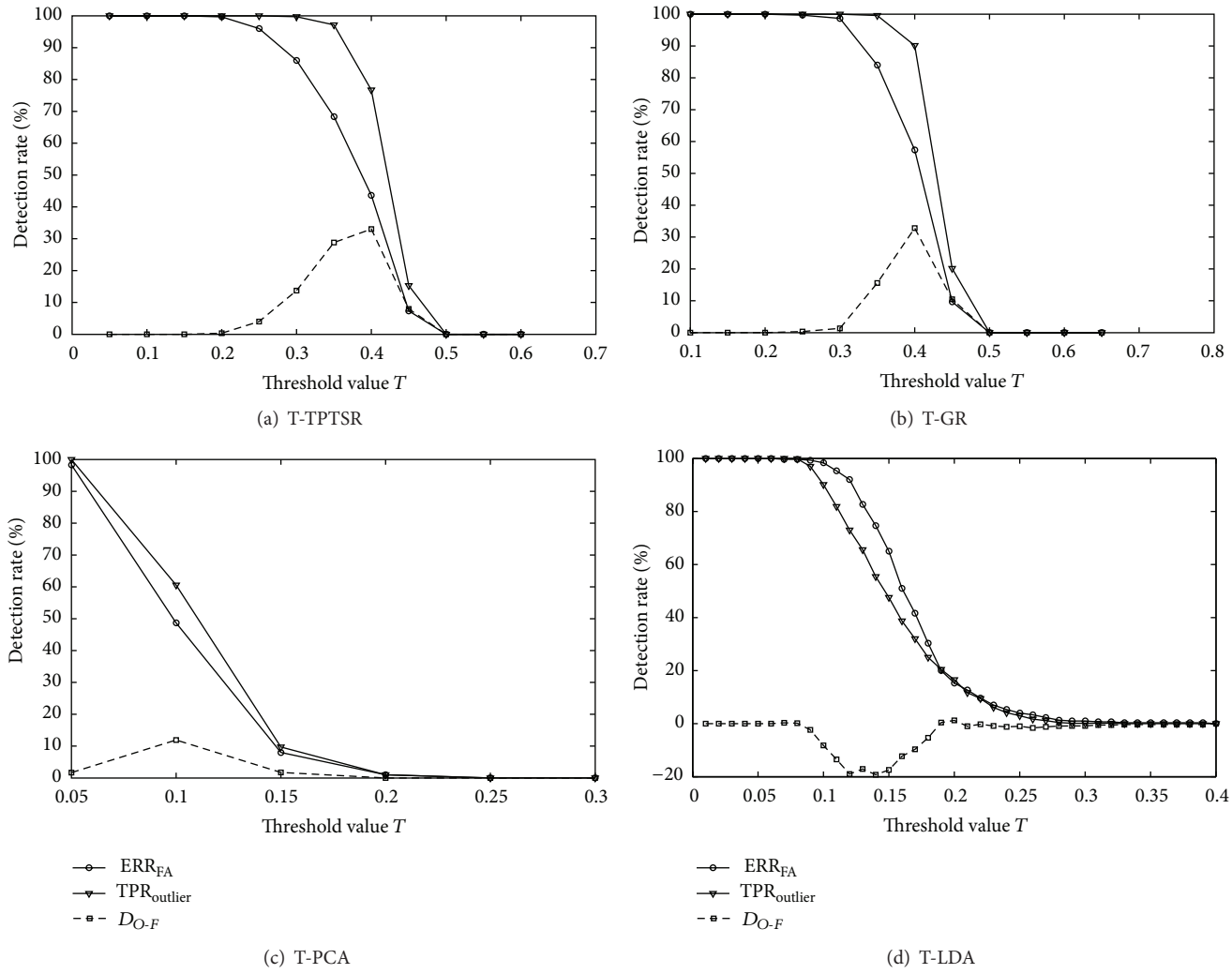


Figure 3: ROC curves for (a) T-TPTSR method, (b) T-GR method, (c) T-PCA method, and (d) T-LDA method, respectively, on the Feret database.

gives details of all the D_{opt} values shown in Figure 3. It can be seen that the T-TPTSR also has a higher value of D_{opt} than other methods.

For the testing with the AR database, we randomly selected 80 classes as the member and the rest of the 40 people are taken as outliers. For each of the members, 13 images are selected randomly from the 26 images as the training set, and the rest of the 13 images are included in the testing set. Hence, there are 1040 training images and 2080 testing images in this test, and in the testing set, there are 1040 member's images and 1040 outlier's images. Figure 4 shows part of the member's and outlier's images from the AR database, and the images for training and testing have been downsized to be a 40×50-pixel image [34].

When we test the T-TPTSR method with the AR database, the number of nearest neighbors M selected is 150. Table 2 describes the ERR_{opt} values and D_{opt} values of the T-TPTSR, T-GR, T-PCA, and T-LDA methods, respectively, when tested with the AR database. It is obvious from the ERR_{opt} values and D_{opt} values that the T-TPTSR method outperforms the T-GR, the T-PCA, and the T-LDA methods in the outlier rejection and member classification applications.

Table 1: Minimum overall classification error rate and maximum ROC difference for T-TPSR, T-GR, T-PCA, and T-LDA methods, respectively, on the Feret database.

Methods	T-TPTSR	T-GR	T-PCA(150)	T-LDA(149)
ERR_{opt} (%)	20.4	23.2	30.0	30.0
D_{opt} (%)	33.0	32.8	11.9	1.24

T-PCA(150) indicate that the T-PCA used 150 transform axes for feature extraction, and T-LDA(119) means that the T-LDA used 119 transform axes for feature extraction. Tables 2 and 3 show the method and number of transform axes used in the same way.

We also test the above methods with the ORL face image database. There are totally 40 classes in the ORL database, and we select 30 random classes to be the members and

Figure 4: Part of the face images from the AR database for testing.

Figure 5: Part of the face images from the ORL database for testing.

Table 2: Minimum overall classification error rate and maximum ROC difference for T-TPSR, T-GR, T-PCA, and T-LDA methods, respectively, on the AR database.

Methods	T-TPTSR	T-GR	T-PCA(1040)	T-LDA(79)
ERR_{opt} (%)	27.2	30.2	33.0	50.0
D_{opt} (%)	45.5	41.8	43.4	21.8

Table 3: Minimum overall classification error rate and maximum ROC difference for T-TPSR, T-GR, T-PCA, and T-LDA methods, respectively, on the ORL database.

	T-TPTSR	T-GR	T-PCA(200)	T-LDA(29)
ERR_{opt} (%)	21.2	24.0	22.8	60.0
D_{opt} (%)	58.6	57.3	57.3	30.0

the other 10 individuals to be the outliers. Among the 30 members, 5 images out of 10 for each of the members are selected randomly as the training samples, and the rest of the 5 images are the testing samples. So in the test, we have 150 training images and 250 testing images, and, in the testing set, there are 150 member's images and 100 outlier's images. Figure 5 shows some sample images from the ORL database, and the images used are also resized to 46×56 [34].

The number of nearest neighbors selected for the T-TPTSR method for the ORL database is 40. Table 3 gives the details of the ERR_{opt} values and D_{opt} values of the four methods, respectively. It can be seen that the T-TPTSR method also shows better performance than all the T-GR, T-PCA, and T-LDA methods, and it has been confirmed that the T-TPTSR method is the optimal solution among them for outlier rejection and member classification.

It is noted that, in the test with the AR and ORL databases, the performance of the T-TPTSR, the T-GR, and the T-PCA are comparable. This is because, under redundant and reasonable resolution sample situation, the performance of the T-PCA method is close to the T-TPTSR and T-GR methods. However, when the T-PCA method is tested with a small number of training samples and low-resolution images, like the Feret database, the advantages of the T-TPTSR method are very obvious.

The criterion we use for judging, whether a sample is an outlier or not, is to measure the distance between the testing sample and the selected target class. If this distance is greater than the threshold, this sample will be classified as an outlier. In T-TPTPR method, the first-phase process finds a local distribution close to the testing sample in the wide sample space by selecting M-nearest samples. In the second-phase processing of the T-TPTSR method, the testing sample is classified based on the distance between the testing sample and the closest class among the M-nearest neighbors. If the testing sample is an outlier, the measure of distance will only be limited within the local distribution within the sample space, and, therefore, the measurement is not confused with other training samples that happen to be close to the outlier.

By applying a suitable threshold, a classifier can reject the outliers and classify the members with the minimum overall classification error rate and the maximum gap between the outlier detection rate and false alarm rate for members. The T-TPTSR method linearly representing the testing sample with the training samples and the distance between the testing sample and the target class are measured by calculating the difference between the testing sample and the weighted contribution of the class in the linear representation. In our test above, the T-TPTSR method achieves the best performance in outlier rejection as well as member classification. This is because in the T-TPTSR the two-phase linear representation of the testing sample results in a closer approximation and assessment by the training samples. Thus, the distance between the testing sample and the target class can be minimized, and the distance between the testing sample and an outlier can be maximized, leading to a better overall classification rate and greater ratio of outlier recognition rate versus the false alarm rate.

5. Conclusion

This paper introduces the modified versions of four useful approaches in face recognition, the T-TPTSR method, the T-GR method, the T-PCA method, and the T-LDA method, for

the application of outlier rejection as well as member classification. Their performance is tested with three different online face image databases, the Feret, AR, and ORL databases, respectively. The results show that the T-TPTSR method achieves the lowest overall classification error rate as well as the greatest difference between the outlier detection rate and false-alarm rate. Even the T-PCA method may achieve comparable performance with the T-TPTSR method under ideal sample conditions, the test result of the T-PCA method is generally poor under bad sample conditions. The T-TPTSR method achieves the best performance in outlier rejection as well as member classification because of the two-phase linear representation of the testing sample with the training samples.

Acknowledgment

Financial supports from The Research Fund for the Doctoral Program of Higher Education (no. 20122302120072) to initiate this research project are gratefully acknowledged.

References

[1] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, 1990.

[2] Y. Xu, D. Zhang, J. Yang, and J. Y. Yang, "An approach for directly extracting features from matrix data and its application in face recognition," *Neurocomputing*, vol. 71, no. 10–12, pp. 1857–1865, 2008.

[3] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[4] Y. Xu and D. Zhang, "Represent and fuse bimodal biometric images at the feature level: complex-matrix-based fusion scheme," *Optical Engineering*, vol. 49, no. 3, Article ID 037002, 2010.

[5] S. W. Park and M. Savvides, "A multifactor extension of linear discriminant analysis for face recognition under varying pose and illumination," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 158395, 11 pages, 2010.

[6] Z. Fan, Y. Xu, and D. Zhang, "Local linear discriminant analysis framework using sample neighbors," *IEEE Transactions on Neural Networks*, vol. 22, no. 7, pp. 1119–1132, 2011.

[7] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

[8] C. Cattani, R. Badea, S. Chen, and M. Crisan, "Biomedical signal processing and modeling complexity of living systems," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 298634, 2 pages, 2012.

[9] V. Vural, G. Fung, B. Krishnapuram, J. G. Dy, and B. Rao, "Using local dependencies within batches to improve large margin classifiers," *Journal of Machine Learning Research*, vol. 10, pp. 183–206, 2009.

[10] Z. Y. Liu, K. C. Chiu, and L. Xu, "Improved system for object detection and star/galaxy classification via local subspace analysis," *Neural Networks*, vol. 16, no. 3-4, pp. 437–451, 2003.

[11] Y. Yang, D. Xu, F. Nie, S. Yan, and Y. Zhuang, "Image clustering using local discriminant models and global integration," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2761–2773, 2010.

[12] Z. Lai, Z. Jin, J. Yang, and W. K. Wong, "Sparse local discriminant projections for face feature extraction," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 926–929, August 2010.

[13] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.

[14] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[15] Y. Shi, D. Dai, C. Liu, and H. Yan, "Sparse discriminant analysis for breast cancer biomarker identification and classification," *Progress in Natural Science*, vol. 19, no. 11, pp. 1635–1641, 2009.

[16] M. Dikmen and T. S. Huang, "Robust estimation of foreground in surveillance videos by sparse error estimation," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08)*, December 2008.

[17] S. Chen and Y. Zheng, "Modeling of biological intelligence for SCM system optimization," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 769702, 10 pages, 2012.

[18] Q. Guan, B. Du, Z. Teng, J. Gillard, and S. Chen, "Bayes clustering and structural support vector machines for segmentation of carotid artery plaques in multicontrast MRI," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 549102, 6 pages, 2012.

[19] S. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.

[20] S. Chen and X. Li, "Functional magnetic resonance imaging for imaging neural activity in the human brain: the annual progress," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 613465, 9 pages, 2012.

[21] Z. Lai, W. Wong, Z. Jin, J. Yang, and Y. Xu, "Sparse approximation to the eigensubspace for discrimination," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 12, pp. 1948–1960, 2012.

[22] Y. Xu, D. Zhang, J. Yang, and J. Y. Yang, "A two-phase test sample sparse representation method for use with face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 9, pp. 1255–1262, 2011.

[23] Y. L. Chen and Y. F. Zheng, "Face recognition for target detection on PCA features with outlier information," in *Proceedings of the 50th Midwest Symposium on Circuits and Systems (MWSCAS '07)*, pp. 823–826, August 2007.

[24] L. Sirovitch and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519–524, 1987.

[25] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[26] B. Scholkopf and A. Smola, *Learning with Kernels*, MIT Press, Cambridge, Massm USA, 2002.

[27] K.-R. Muller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–201, 2001.

[28] D. Tao and X. Tang, "Kernel full-space biased discriminant analysis," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '04)*, pp. 1287–1290, June 2004.

[29] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. J. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 212–220, 2007.

[30] P. Jonathon Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[31] P. J. Phillips, "The Facial Recognition Technology (FERET) Database," http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.

[32] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[33] http://cobweb.ecn.purdue.edu/?aleix/aleix?face?DB.html.

[34] Y. Xu and Z. Jin, "Down-sampling face images and low-resolution face recognition," in *Proceedings of the 3rd International Conference on Innovative Computing Information and Control (ICICIC '08)*, pp. 392–395, June 2008.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 619385, 8 pages
http://dx.doi.org/10.1155/2013/619385

Research Article

Computational Approach to Seasonal Changes of Living Leaves

Ying Tang,[1,2] Dong-Yan Wu,[1,2] and Jing Fan[1,2]

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

Correspondence should be addressed to Jing Fan; fanjing@zjut.edu.cn

Received 10 December 2012; Accepted 17 January 2013

Academic Editor: Carlo Cattani

This paper proposes a computational approach to seasonal changes of living leaves by combining the geometric deformations and textural color changes. The geometric model of a leaf is generated by triangulating the scanned image of a leaf using an optimized mesh. The triangular mesh of the leaf is deformed by the improved mass-spring model, while the deformation is controlled by setting different mass values for the vertices on the leaf model. In order to adaptively control the deformation of different regions in the leaf, the mass values of vertices are set to be in proportion to the pixels' intensities of the corresponding user-specified grayscale mask map. The geometric deformations as well as the textural color changes of a leaf are used to simulate the seasonal changing process of leaves based on Markov chain model with different environmental parameters including temperature, humidness, and time. Experimental results show that the method successfully simulates the seasonal changes of leaves.

1. Introduction

The seasonal changes of trees vary the appearances of trees through seasons, which include shapes and textures of the leaves, flowers, and fruits. Among these, the change of leaves constitutes the most important part of the seasonal changes of trees. In this paper, we focus on how to compute the leaf changing during different seasons.

As we observe the changes of leaves from spring to winter, most leaves become withered and curled up due to the influences of environmental factors [1]. Besides, the leaves usually turn from green to yellow during the aging process and finally fall off to the ground. According to the above observation, the seasonal changes of leaves are simulated in terms of their geometric deformations as well as their textural colors transitions. There is a lot of research work done in simulating 3D shape changes of leaves the occurring during the withering process of leaves. Most of these methods generate the 3D deformation of leaves based on the changes of veins [2–7]. For veins-driven methods [3, 4, 6, 7], each vertex in the 3D model of a leaf is deformed to the nearest vertex in the interactively generated veins and deformations are controlled by dragging some vertices in the veins. These methods involve, much user interaction to extract the skeleton of the leaf, and the generated results are not realistic enough. The method proposed by Chi et al. [8] combines the veins with a double-layered model of the leaf and simulates the deformation process more realistically. However, this method is computationally intensive and difficult to implement due to the complex computation. In this paper, we propose a new improved method using mass-spring model and grayscale mask map to simulate the deformation process of leaves with simplified computations and realistic results.

In order to simulate textural colors of leaves, the Phong lighting model with a diffuse component derived from leaf pigments is adopted to directly compute the reflections on the surfaces of leaves [9]. Other methods use the technique of texture mapping to produce the leaves' appearances, and the textures can be changed to reflect the appearance changes of leaves [10]. In our method, we apply multiple textures to represent appearance changing of leaves in different seasons.

In order to efficiently simulate the seasonal changes of leaves, we combine the changes of geometric shape and textural color of the above methods in our algorithm to produce the results. The Markov chain model is used to show the state transfer of leaves in the dynamic growing process of trees. The following sections are arranged as follows. In Section 2, the related work is introduced. We describe the modeling of

three-dimensional leaves in Section 3. Section 4 focuses on the implementation of geometrical changes of leaves based on improved mass-spring model. In Section 5, the Markov chain-based method is described to compute different states of leaves combining the texture and geometry changes. We show our experimental results in Section 6 and conclusion in Section 7.

2. Related Work

The work related to the simulation of seasonal changes of leaves includes leaf modeling, leaf deformation, and leaf appearances rendering. For leaf modeling, there are L-system-based and image-based methods. The L-system-based methods model leaves with self-similarity [11, 12]. As for image-based modeling methods [13, 14], usually the feature points on the edge of the leaf are extracted from the scanned leaf image, and the geometric shape of the leaf is represented by the triangular meshes produced by Delaunay algorithm [15]. According to the botanical characteristics of the leaf, Dengler and Kang claim that leaf shapes have a close relationship with leaf veins [16] which is used to generate the shapes of leaves. Runions et al. present the biologically motivated method to construct leaf veins with user interaction [17]. Besides user interaction, the leaf veins are generated by fixing the start points and setting the control points of veins according to the sum of a fixed value and a random parameter between zero and ten [18]. Chi et al. [8] introduce an improved method to construct the leaf vein skeleton which generates the main vein and the branch vein separately, and the leaf model is built by a double-layered mass-spring model. These methods produce the relatively complex leaf models which reflect the characteristics of leaf's geometric shapes. In this paper, we generate the optimized triangular mesh to represent the leaf model by two steps. In the first step, the key points on the edge of the leaf are obtained through user interaction. Then, the optimized leaf triangular mesh is generated by improved Delaunay algorithm in the second step. Instead of generating the leaf veins explicitly in the modeling procedure, we emphasize leaf veins with a user-specified mask in the process of leaf deformation.

The leaves gradually become withered and curled up during the transitions of different seasons. The deformation of geometric shapes of leaves is very important to simulate the seasonal changes. The 3D deformation algorithms are mainly classified into two categories, which are free-form-based deformation methods [19] and physically based deformation methods [20]. Free-form-based deformation methods are widely used in the field of computer animation and geometric modeling [21]. These kinds of methods embed the objects into a local coordinate space and transform the local space to make the objects deformed. There are two common physically based deformation methods: skeleton-based method and mass-spring-based method. The deformation method based on skeleton is relatively simple [7] and produces more realistic deformation results of leaves. However, it requires much human interaction. Mass-spring model is more frequently used in fabric deformation [22]. Tang and Yang [23] adopt the mass-spring model to generate the deformation of leaves, in which the mesh of the leaf is not optimized, and the deformation effects are relatively unnatural and difficult to control. Double mass-spring model proposed by Chi et al. [8] is capable of simulating the changes of leaves more realistically. However, it is complex and difficult to be implemented.

In order to simulate color changes of leaf surfaces in various environmental conditions, Phong lighting model considering leaf's pigments [9] and the technique of texture mapping [24] have been adopted. The texture images of leaves can be obtained by scanning real leaves [25] or texture synthesis [26]. Desbenoit et al. [10] applies open Markov chain model to decide which texture images are mapped to certain leaves to simulate the aging process of the leaves. In this paper, we also adopt the Markov chain model to statistically determine the distribution of leaves, textures on the tree under the influence of environmental factors including temperature and humidness.

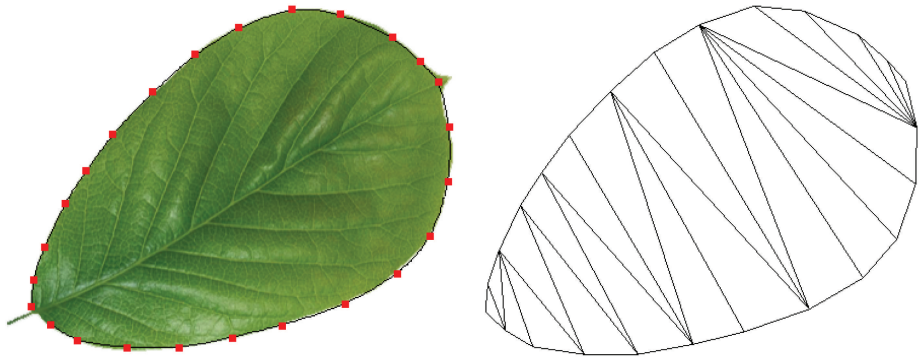
3. Modeling Three-Dimensional Leaves

In this paper, we apply the image-based approach to model the geometric shapes of three-dimensional leaves [27, 28]. First, the key points on the edge of the leaf are obtained through user interaction, and then the triangular mesh of the leaf is constructed by Delaunay triangulation through incremental insertion of points [29, 30]. Finally, the optimization procedure is employed to compute the high quality mesh with even-sized triangles.

Instead of adopting the automatic edge detection methods to extract the leaf contour, we provide the interface to make the user interactively select the edge points of the leaf. After the selection of edge points, the smooth B-spline curve running through these points is automatically generated to approximate the leaf edges [31]. The B-spline edge which passes through the user-selected points is shown in Figure 1(a), from which we find that the curve represents the real leaf edge well. If more control points are selected, the edge is more accurate. The generated B-spline curve is sampled to get the key points which are to be used in Delaunay triangulation.

The Delaunay triangulation method is usually used to generate a triangulated irregular network (TIN) [32]. The Delaunay triangles are a set of connected but not overlapping triangles, and the circumscribed circle of the triangles does not contain any other point in the same region. Unfortunately, the initially triangular mesh generated with key points on the edge usually contains some long and narrow triangles, as shown in Figure 1(b). The leaf mesh with such bad quality triangles would make the leaf deformation unnatural. Instead, we need to generate a high quality leaf mesh with even-sized triangles. So we optimize the triangular mesh based on the subdivision method in [33]. An even-sized triangular mesh is obtained by repeating the following two steps: (1) relocate the vertex position; (2) modify the connection properties of triangles.

The high-resolution triangular mesh produces more natural and smooth deformations. However, more triangles in the mesh would lead to more time to compute the



(a) (b)

Figure 1: (a) The B-spline curve with key points selected by the user; (b) the Delaunay triangulated mesh of the leaf.

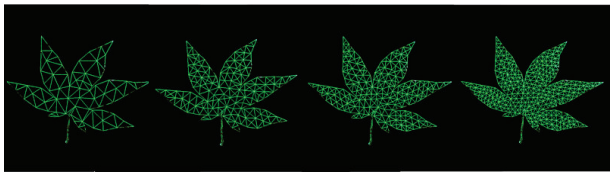


Figure 2: Triangular meshes of the maple leaf produced by a different number of iterations.

deformation. According to the triangulation algorithm, the subdivision level of triangular mesh is related to the number of iterations. Usually, we set the number of iterations to be 160 in our implementation, which is enough to produce the subdivided triangular mesh capable of natural deformation within acceptable time. In Figure 2, we show the triangular mesh models of the maple leaf produced by a different number of iterations.

4. Deformations of Leaves Based on Improved Mass-Spring Model

Leaves become slowly curled up as the season changes. This phenomenon is mainly caused by the different structures of the upper and bottom surfaces of a leaf, which have different amounts of contraction during the dehydration process. To take into account the differences between the upper and bottom internal structures of a leaf, we introduce the improved mass-spring model to make leaf deformation more realistic.

4.1. Numerical Calculation and Constraints. The mass-spring model is widely used in the simulation of the deformation of soft fabrics [34]. This model consists of two important parts: a series of virtual particles and the corresponding light springs of natural length nonequal to zero. The deformation of the object is determined by the displacements of particles after they are stressed. The springs connecting the particles constrain the movement of particles. The triangular mesh model of a leaf can be used as the mass-spring model, where the mesh vertices are regarded as particles and the edges are as springs [8].

There are internal and external forces acting on the springs, and we denote the joined forces as $F_{i,j}(t)$. The force distribution is computed by Newton's laws of motion, and explicit Euler's method is adopted to find the numerical solution of the model. The equations to compute the acceleration, particle velocity, and particle displacement are listed as follows:

$$\begin{aligned} a_{i,j}(t+\Delta t) &= \frac{1}{\mu_{i,j}} F_{i,j}(t), \\ v_{i,j}(t+\Delta t) &= v_{i,j}(t) + \Delta t \cdot a_{i,j}(t+\Delta t), \\ P_{i,j}(t+\Delta t) &= P_{i,j}(t) + \Delta t \cdot v_{i,j}(t+\Delta t). \end{aligned} \tag{1}$$

In the above equations, the mass of a particle is denoted as $\mu_{i,j}$, the acceleration is denoted as $a_{i,j}$, the velocity of a particle is denoted as $v_{i,j}$, and the particle's displacement is denoted as $P_{i,j}$. The time step is denoted as Δt, the value of which is important in computing the desirable deformation. The time step needs to be small enough to ensure the stability of the numerical calculation. Otherwise, dramatic changes of particle positions would be incurred by large time step values.

Actually, the deformation curve of a leaf under forces is not ideally linear. If we directly compute the deformation with the above equations, the problem of "over elasticity" would occur, that is, the deformation of the springs would exceed 100%. To overcome this problem, we adopt the method of constraining velocities to constrain the deformation of the springs [35]. The basic idea is as follows. Particle u and particle v are the ends of spring s.$V_\mu(t)$ and $V_v(t)$, respectively, represent the velocity of particle u and particle v at time t. Assume that the relative velocity between the two particles is $V_{\mu,v}(t)$, and the relative position is $P_{\mu,v}(t)$, the new relative position after one time step $P_{u,v}(t+\Delta t)$ is computed by constraining the velocity of the particle. If $P_{u,v}(t+\Delta t)$ satisfies (2), the velocity is updated [35]. Otherwise, it is not updated

$$P_{u,v}(t+\Delta t) = \left|P_{u,v}(t) + V_{u,v}(t+\Delta t)\cdot \Delta t\right| \le (1+\tau_c)\cdot L. \tag{2}$$

(a) (b)

Figure 3: (a) The texture of a maple leaf; (b) mask map of the maple model.

In (2), L presents the natural length of the spring without any forces exerted, and τ_C is the threshold of deformation. This equation guarantees that when the value of τ_c is set to be 0.1, the maximum deformation length of the spring does not exceed 10 percent of the natural length. In other words, the difference between $P_{u,v}(t + \Delta t)$ and $P_{u,v}(t)$ should be within 10 percent of the natural length.

4.2. Deformation. The key of shape deformation is to compute the changes of the position of each particle. If each particle has the same mass value, the relative displacements in directions x, y, and z only depend on the joint force in each direction. For a relatively high-resolution mesh model with nearly even-sized triangles, the joint forces between most particles and its adjacent particles would not differ enough to make desirable deformations. Thus, the uniform mass of all particles is not in favor of generating the nonuniform deformation results relative to different leaf regions, for example, the regions near edges usually undergo more deformation than the center regions. To enhance the change of the relative displacement of each particle and generate the adaptively deformed results for different leaf regions, we adaptively allocate the mass values to different particles in our improved deformation model.

According to Newton's law of motion $F = ma$, for the same force F, the smaller the object's mass m is, the larger the acceleration a is. So we can control the deformation of leaves by setting different masses of the particle's. We introduce the *mask map* to adaptively control the particles masses. The *mask map* is generated according to the texture image of the leaf. Suppose that we have a texture image of a leaf called leaf1.bmp which is obtained by scanning the real leaf. We select out the leaf region from the texture and paint different grayscale colors to this region. The intensities of the painted pixels are in proportion to the particle's masses. For example, if we try to set a smaller mass value for a particle, we can paint this pixel in black or an other color close to black. A maple leaf is shown in Figure 3(a). According to our observations of natural maple leaves, the regions around the leaf corner and close to petiole usually undergo more deformation than other regions. So we paint these regions in black or darker gray values while other regions in brighter gray values as shown in Figure 3(b). Different *mask maps* map different masses to the same particles, which results in different deformation results. The corresponding *mask map* needs to be generated based on the natural deformation pattern of the specific leaf.

According to the texture coordinates of the particles of triangular mesh, we find in the mask map the pixels which correspond to particles in the leaf mesh model. The gray values of pixels in the mask map are mapped to the value of particle masses m by the following:

$$m = \begin{cases} 0.5, & \text{gray} = 0 \\ \ln(\text{gray} + 1), & \text{gray!} = 0 \end{cases} \tag{3}$$

In (3), the mass value is computed as logarithm of the grayscale value, which makes the change of the masses more gentle and smooth compared with the changes of grayscale values. Such mass distribution is more amenable to yield natural deformation of leaves.

The detailed steps to implement deformation process are shown as follows.

(1) Generate the *mask map* to determine the mass distribution of the leaf.

(2) Initialize parameter values in our improved mass-spring model. Set the initial velocity and acceleration of particles to be zero. Initialize masses of the particles according to the *mask map*.

(3) Establish constraints among particles. The connection between particles (i.e., the mesh topology) determines what other particles directly exert forces on the current particle for the computation of displacements. The constraints are built by three steps as follows.

Step 1. Find the adjacent triangle faces of current particle. Adjacent faces are those triangles which include a current particle as one of their vertices.

Step 2. Find the adjacent particles of a current particle. The other two vertices in adjacent triangles are the adjacent particles of a current particle.

Figure 4: Several deformations using the mask map in Figure 3(b).

Step 3. Establish the constraints. Set a flag value for each particle to describe whether this particle had been traversed, and initialize the flag value as false. If one particle is traversed, set its flag value as true. Set the constraints between this particle and its adjacent particles if they are not traversed. Thus, all particles are traversed only once, and the constraints are set without duplication. When this particle is moved, the particles having constraints move with it too.

(4) Exert the force, and compute the change of position of each particle by numerical calculation in one time step.

(5) Repeat the numerical calculation in each time step to obtain the new velocities and accelerations, and update particle positions accordingly to produce deformation effects at different time steps.

For example, the deformation results at different time steps of the maple leaf under the *mask map* in Figure 3(b) are showed in Figure 4 (the first model is the original mesh model).

The deformation results in Figure 4 show that the leaf regions with darker gray values are deformed more than the regions with brighter gray values. The masses of those regions with darker gray values are smaller so that they move more distances under forces. The regions with brighter gray values have larger masses which make them move much more slowly. Different movements of particles distributed over the leaf surfaces produce the adaptive deformation results over the leaf surface. If we paint the veins white or bright gray values, we can get the deformation result in which the veins are kept unmoved and two-side regions around veins become curly. With this method, we can control the leave's deformation flexibly. For the same leaf model, we can generate different deformation results by different *mask maps*. In Figure 6, we show the different deformation results for the same leaf model for a different *mask map* in Figure 5. Therefore, in order to achieve desirable deformations, we can construct the corresponding *mask map* to make the leaves deformed as expected.

Figure 5: Another mask map of the maple leaf model.

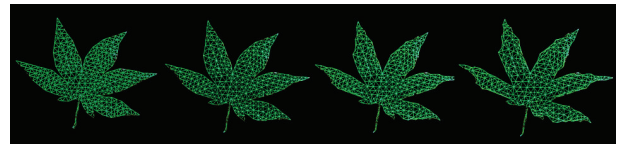


Figure 6: Different deformation results of the maple leaf for mask map shown in Figure 5.

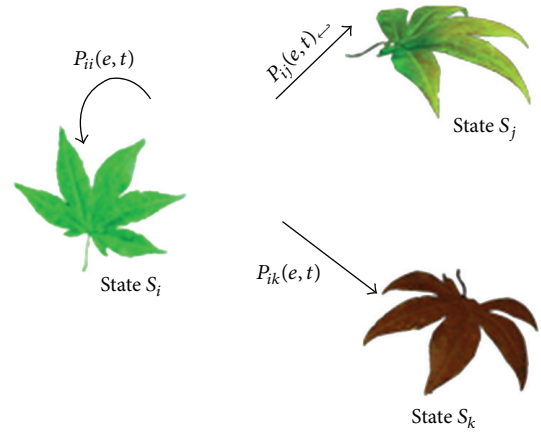


Figure 7: Transition relationship for Markov chain model.

5. Textural and Geometric Changes

To simulate the seasonal changes of leaves, we need to take the transitions of textural colors of leaves into account besides geometric deformations. The whole seasonal changing process of leaves can be regarded as the sequences of a series of discrete states. The leaves transform from one state to the other with certain probabilities conditioned by environmental factors. This transformation can be approximated by Markov chain model [10].

Markov chain model has two properties. (1) The state of the system at time $t + 1$ is only related to the state at time t and has nothing to do with the states at a previous time. (2) Transformation of the state from time t to time $t + 1$ has nothing to do with the value of t. The leaf changing process can be regarded as the Markov chain. Different texture images as well as the deformed geometric shapes are

Figure 8: Seven texture states of a maple model.

organized to constitute different states in the Markov chain. We simulate various distributions of leaves on the tree by the randomness of the Markov chain model. The environmental factors including temperature and humidness are used as the conditions to determine the probability to transfer from one state to another. By setting different environmental parameters, we get the seasonal appearances of trees with the corresponding distributions of leaves.

The leaf's state is denoted as S_x, where $0 \le x < n$ and n represent the total number of possible states of leaves. Assume that we have three states S_i, S_j, and S_k and the transition relationship among these three states are shown in Figure 7. It shows that for the state S_i at time t, it may evolve to states S_j and S_k or remain in the original state at time $t+1$ with certain probabilities.

The arc $P_{ii}(e,t)$ in Figure 7 represents the possibility that a leaf at a given state S_i stays in the same state at the next time. It is defined as the probability of keeping self-state. The function of this probability is denoted as follows [10]:

$$P_{ii}(e,t) = e^{-\lambda_i(e)t}, \quad 0 \le i \le n, \tag{4}$$

$$\lambda_i(e) = \frac{\ln 2}{\tau_i(e)}. \tag{5}$$

Function $\tau_i(e)$ is the bilinear interpolation of the temperature and humidness.

The probability that the leaf transfers to other states is denoted as $1 - P_{ii}(e,t)$. $P_{ij}(e,t)$ is defined as the probability of the leaf at state S_i transferring to another state S_j, and it is computed by (6) as follows:

$$P_{ij}(e,t) = \left(1 - P_{ii}(e,t)\right) X_{ij}(e), \quad 0 \le i < n, \ i \ne j. \tag{6}$$

Function $X_{ij}(e)$ is the bilinear interpolation of four constants between zero and one. These four constants correspond to the transition possibilities in the four extreme cases: wet and cold, wet and warm, dry and cold, and dry and warm. The values of these constants are interactively specified by users.

The parameters of time, temperature, and humidness are set by users. Taking the maple leaves in Figure 8, for example, we use three specific combinations of textures and shapes for each season. For instance, three main states are used to represent leaves in summer, which are texture 2 in Figure 8 combined with the first deformation in Figure 4, texture 3 combined with the second deformation, and texture 4 combined with the third deformation.

Several states which combine changes of textures and shapes in different seasons are showed in Figure 9. Given the combinations of states, we calculate the transition probabilities of leaves according to the specific temperature and humidness set for certain seasons and get the corresponding leave's distributions in that season.

Figure 9: The basic triangular mesh model of the maple leaf, and seven states combining textures and geometric deformations.

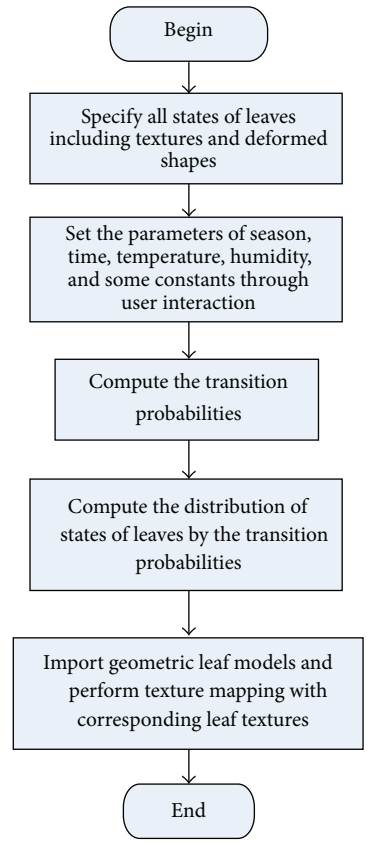


Figure 10: Seasonal changing process of leaves based on Markov-chain model.

To summarize, the seasonal changing process of leaves under certain environmental parameters is showed in Figure 10.

Figure 11: Tree growing process based on L-system.

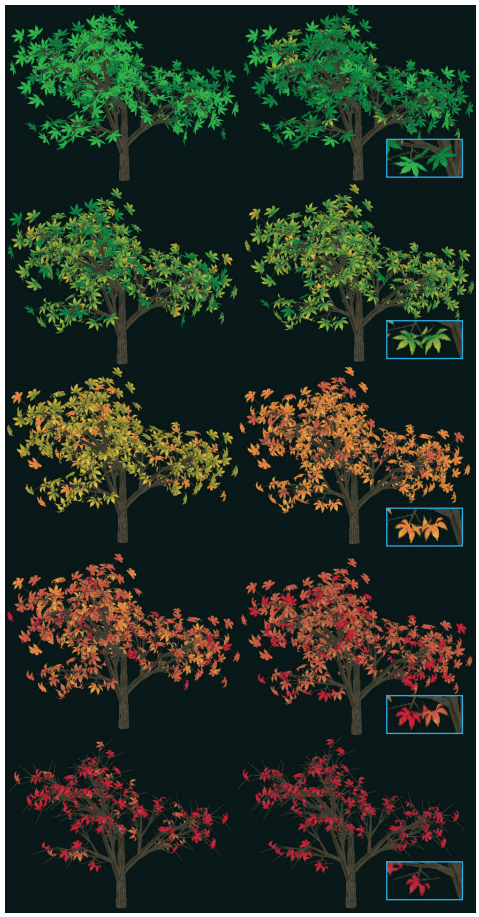


Figure 12: Seasonal changes of a maple tree based on Markov chain model.

6. Results

To produce the results of seasonal changes of trees, we grow the leaves on the trees and simulate their distributions for different seasons. In order to get the 3D model of the tree, we adopt the L-system method to produce the trunks and branches of the tree. The trunks and branches of the tree are drawn with quadratic surface, and the leaves grown on branches are modeled as triangular meshes. In Figure 11, we model the tree and its growth through the iteration of the L-system, and the leaves grown on the tree are shown. To simulate leaves, seasonal changes, we distribute various leaves on the tree under different environments based on Markov chain model. Figure 12 shows some seasonal changes of the maple tree, and the enlarged picture at the lower right corner show the change of the individual leaf more clearly.

7. Conclusion

In this paper, we propose a computational approach to simulate the seasonal changes of living leaves by combining the changes in geometric shapes and textural colors. First, the key points are selected on the leaf image by user interaction. Then, the triangular mesh of the leaf is constructed and optimized by improved Delaunay triangulation. After the models of leaves have been obtained, the deformations of leaves are computed by improved mass-spring models. The seasonal changes of trees under different environmental parameters are computed based on Markov chain. The improved mass-spring model is based on the user-specified *mask map* which adaptively determines the masses of particles on the leaf surface.

In the future, we are interested in the following work.

(1) Work on how to generate the mask map more naturally according to the characteristics of the deformations of leaves.

(2) Intend to simulate the dynamic procedure of the leaves falling onto ground out of gravity.

(3) Develop a more precise model to compute the colors of leaves which takes into account of the semitransparency of leaves.

Acknowledgments

This work is supported by National Natural Science Foundation of China (61173097, 61003265), Zhejiang Natural Science Foundation (Z1090459), Zhejiang Science and Technology Planning Project (2010C33046), Zhejiang Key Science and Technology Innovation Team (2009R50009), and Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology.

References

[1] C. Cattani, R. Badea, S. Chen, and M. Crisan, "Biomedical signal processing and modeling complexity of living systems," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 298634, 2 pages, 2012.

[2] Q. Xu, *Research on techniques of mesh deformation [Ph.D. thesis]*, Zhejiang University, 2009.

[3] P. Prusinkiewicz, L. Mündermann, R. Karwowski, and B. Lane, "The use of positional information in the modeling of plants," in *Proceedings of the Computer Graphics Annual Conference (SIGGRAPH 2001)*, pp. 289–300, August 2001.

[4] L. Mündermann, P. MacMurchy, J. Pivovarov, and P. Prusinkiewicz, "Modeling lobed leaves," in *Proceedings of the Computer Graphics International (CGI'03)*, pp. 60–65, July 2003.

[5] S. Y. Chen, "Cardiac deformation mechanics from 4D images," *Electronics Letters*, vol. 43, no. 11, pp. 609–611, 2007.

[6] S. M. Hong, B. Simpson, and G. V. G. Baranoski, "Interactive venation-based leaf shape modeling," *Computer Animation and Virtual Worlds*, vol. 16, no. 3-4, pp. 415–427, 2005.

[7] S. L. Lu, C. J. Zhao, and X. Y. Guo, "Venation skeleton-based modeling plant leaf wilting," *International Journal of Computer Games Technology*, vol. 2009, Article ID 890917, 8 pages, 2009.

[8] X. Y. Chi, B. Sheng, Y. Y. Chen, and E. H. Wu, "Physically based simulation of weathering plant leaves," *Chinese Journal of Computers*, vol. 32, no. 2, pp. 221–230, 2009.

[9] M. Braitmaier, J. Diepstraten, and T. Ertl, "Real-time rendering of seasonal influenced trees," in *Proceedings of the Theory and Practice of Computer Graphics*, pp. 152–159, Bournemouth, UK, June 2004.

[10] B. Desbenoit, E. Galin, S. Akkouche, and J. Grosjean, "Modeling autumn sceneries," in *Proceeding of the Eurographics*, pp. 107–110, 2006.

[11] P. Prusinkiewicz and A. Lindennmyer, *Algorithmic Beauty of Plants*, Springer, Berlin, Germany, 1990.

[12] S. B. Zhang and J. Z. Wang, "Improvement of plant structure modeling based on L-system," *Journal of Image and Graphics*, vol. 7, no. 5, pp. 457–460, 2002.

[13] L. Quan, P. Tan, G. Zeng, L. Yuan, J. D. Wang, and S. B. Kang, "Image-based plant modeling," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 599–604, 2006.

[14] P. Tan, G. Zeng, J. D. Wang, S. B. Kang, and L. Quan, "Image-based tree modeling," in *Proceedings of the ACM SIGGRAPH 2007*, New York, NY, USA, August 2007.

[15] L. P. Chew, "Guaranteed-quality triangular meshes," Tech. Rep. TR-89-983, Department of Computer Science, Cornell University, 1989.

[16] N. Dengler and J. Kang, "Vascular patterning and leaf shape," *Current Opinion in Plant Biology*, vol. 4, no. 1, pp. 50–56, 2001.

[17] A. Runions, M. Fuhrer, B. Lane, P. Federl, A. G. Rolland-Lagan, and P. Prusinkiewicz, "Modeling and visualization of leaf venation patterns," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 702–711, 2005.

[18] Z. J. Ma and Y. M. Jiang, "Chinar leaf simulation," *Computer Simulation*, vol. 26, no. 2, 2009.

[19] T. W. Sederberg and S. R. Parry, "Free-form deformation of solid geometric models," *Computer Graphics*, vol. 20, no. 4, pp. 151–160, 1986.

[20] L. H. de Figueiredo, J. de Miranda Gomes, D. Terzopoulos, and L. Velho, "Physically-based methods for polygonization of implicit surfaces," in *Proceedings of the Graphics Interface '92*, pp. 250–257, May 1992.

[21] G. R. Liu, J. H. Lin, X. D. Liu, and F. R. Zhao, "Free-form definition based on three-dimensional space," *Microelectronics and Computer*, vol. 25, no. 7, 2008.

[22] X. Provot, "Deformation constraints in a mass-spring model to describe rigid cloth behavior," in *Proceedings of the Graphics Interface Conference '95*, pp. 147–154, May 1995.

[23] Y. Tang and K. F. Yang, "Research on visualization of deformation of three-dimensional leaves," *Computer Simulation*, vol. 28, no. 5, 2011.

[24] N. Chiba, K. Ohshida, K. Muraoka, and N. Saito, "Visual simulation of leaf arrangement and autumn colours," *Journal of Visualization and Computer Animation*, vol. 7, no. 2, pp. 79–93, 1996.

[25] N. Zhou, W. Dong, and X. Mei, "Realistic simulation of seasonal variant maples," in *Proceedings of the 2nd International Symposium on Plant Growth Modeling and Applications (PMA'06)*, pp. 295–301, Beijing, China, November 2006.

[26] X. Y. Chi, B. Sheng, M. Yang, Y. Y. Chen, and E. H. Wu, "Simulation of autumn leaves," *Journal of Software*, vol. 20, no. 3, pp. 702–712, 2009.

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

[28] J. Zhang, S. Chen, S. Liu, and Q. Guan, "Normalized weighted shape context and its application in feature-based matching," *Optical Engineering*, vol. 47, no. 9, Article ID 097201, 2008.

[29] B. A. Lewis and J. S. Robinson, "Triangulation of planar regions with applications," *The Computer Journal*, vol. 21, no. 4, pp. 324–332, 1978.

[30] G. Macedonio and M. T. Pareschi, "An algorithm for the triangulation of arbitrarily distributed points: applications to volume estimate and terrain fitting," *Computers and Geosciences*, vol. 17, no. 7, pp. 859–874, 1991.

[31] 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.

[32] V. J. D. Tsai, "Delaunay triangulations in TIN creation: an overview and a linear-time algorithm," *International Journal of Geographical Information Systems*, vol. 7, no. 6, pp. 501–524, 1993.

[33] L. Markosian, J. M. Cohen, T. Crulli, and J. Hughes, "Skin: a constructive approach to modeling free-form shapes," in *Proceedings of the SIGGRAPH Conference'99*, pp. 393–400, 1999.

[34] H. Liu, C. Chen, and B. L. Shi, "Simulation of 3D garment based on improved spring-mass model," *Journal of Software*, vol. 14, no. 3, pp. 619–627, 2003.

[35] X. P. Sun, W. W. Zhao, and X. D. Liu, "Dynamic cloth simulation based on velocity constraint," *Computer Engineering and Applications*, vol. 44, no. 31, pp. 191–194, 2008.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 672509, 5 pages
http://dx.doi.org/10.1155/2013/672509

Research Article

Reliable RANSAC Using a Novel Preprocessing Model

Xiaoyan Wang,[1] Hui Zhang,[2] and Sheng Liu[1]

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

Correspondence should be addressed to Xiaoyan Wang; xw292@cam.ac.uk

Received 8 December 2012; Revised 8 January 2013; Accepted 17 January 2013

Academic Editor: Carlo Cattani

Geometric assumption and verification with RANSAC has become a crucial step for corresponding to local features due to its wide applications in biomedical feature analysis and vision computing. However, conventional RANSAC is very time-consuming due to redundant sampling times, especially dealing with the case of numerous matching pairs. This paper presents a novel preprocessing model to explore a reduced set with reliable correspondences from initial matching dataset. Both geometric model generation and verification are carried out on this reduced set, which leads to considerable speedups. Afterwards, this paper proposes a reliable RANSAC framework using preprocessing model, which was implemented and verified using Harris and SIFT features, respectively. Compared with traditional RANSAC, experimental results show that our method is more efficient.

1. Introduction

Feature matching is a basic problem in computer vision. Corresponding to local features has become the dominant paradigm for structure from motion [1, 2], image retrieval [3], and medical image processing [4]. It is a crucial issue to correspond to the features accurately and efficiently [5, 6]. Most applications are built upon a general pipeline consisting of steps for extracting features from images, matching them to obtain correspondences, and applying some forms of geometric verification to reject the outliers. The geometric verification is extremely critical for the pipeline's success. It has been proven that RANSAC [7] is the best method of choice for this pipeline [8]. However, there are two obvious shortcomings in RANSAC processing. On one hand, it is time-consuming. On the other hand, when the sampling time is restricted artificially, the selected matching pairs may not be correct.

Consequently, numerous extensions for RANSAC have been proposed to speed up different RANSAC stages, such as SCRANSAC [8], optimal randomized RANSAC [9], and other improved methods [10–12]. However, even with these extensions, the geometric verification is still a major bottleneck in applications. In addition, most of the improved methods cost considerable implementation runtime and are difficult to tune for optimal performance.

This paper proposes a fast and simple RANSAC framework based on a preprocessing model. It can result in a reduced correspondence set with a higher inlier percentage, on which RANSAC will converge faster to a correct solution. This model can successfully acquire a subset E with higher probability being inliers from the initial corresponding set P. Then, a reliable fundamental matrix $\mathbf{F}$ or a homography matrix $\mathbf{H}$ can be estimated from subset E. Owing to E with higher inliers ratio, the estimated $\mathbf{H}$ or $\mathbf{F}$ is more reliable. Finally, the outliers in set P can be rejected according to the estimated $\mathbf{H}$ or $\mathbf{F}$. Comparing with other improved methods, the proposed approach in this paper can achieve similar speedup while being considerably simpler to implement.

The rest of this paper is organized as follows. In Section 2, this paper discusses RANSAC for outlier rejection and introduces preprocessing model, including its motivation and algorithm flowchart. In Section 3, a novel RANSAC framework based on Preprocessing Model is proposed. Section 4

presents the experimental results and data analysis. The last part is a summarization of this paper.

2. Outlier Rejection

RANSAC has become the most popular tool to solve the geometric estimation problems in datasets containing outliers, which was first proposed by Fischler and Bolles in 1981 [7]. It is a nondeterministic algorithm with a purpose that it can produce a reasonable result only with a certain probability.

2.1. RANSAC. RANSAC operates in a hypothesized-and-verified framework. The basic assumption of RANSAC algorithm is that the data consists of "inliers", that is, the data whose distribution can be explained by some set of model parameters. And "outliers" are the data which do not fit the model. The outliers probably result from errors of measurement, unreasonable assumptions, or incorrect calculations. RANSAC randomly samples a minimal subset s of size from the initial set in order to hypothesize a geometric model. This model is then verified against the remaining correspondences, and the number of inliers, that is, of correspondences consistent with the model, is determined as its score. RANSAC achieves its goal by iteratively selecting a random subset of the original data, which are hypothetical inliers. This procedure is iterated until a certain termination criterion is met. In confidence p, ensure that at least one sampling within N times sampling, the elements are all inliers. The equation is

$$N = \frac{\log(1-p)}{\log(1-\varphi^s)}, \tag{1}$$

where s is the mean of the minimal size of sampling subset to hypothesize the geometric model, and φ represents the probability of a point being an inlier.

The iteration ensures a bounded runtime as well as a guarantee on the quality of the estimated result. As mentioned above, there are some limits in RANSAC processing. Time-consuming is the most urgent problem, especially when the initial inliers rate is low. Hence, this paper proposes a novel RANSAC framework with a preprocessing model to improve it.

2.2. Preprocessing Model. The main effort of this preprocessing model is to explore a reduced set with reliable correspondences from initial matching dataset and estimate the geometric model. This model can be divided into the following two steps.

2.2.1. Selecting Reliable Corresponding Pairs. When verifying hypotheses in RANSAC, the corresponding pairs are categorized into inliers and outliers. Since the number of samples taken by RANSAC depends on the inlier ratio, it is desirable to reduce the fraction of outliers in the matching set. Selecting a reduced set with higher inlier ratio is the first step of this preprocessing model. Our approach is motivated by the observation that extracting and exploring a subset E with higher probability being inliers is an efficacious idea to improve the runtime of RANSAC. The idea underlying the preprocessing model is to use relaxation technique [13] to acquire a reduced set of more confident correspondences. It leads to a significant speedup of the RANSAC procedure for two reasons. First, RANSAC only needs to operate on a substantially smaller set E for verifying model hypotheses. Second, the additional constraints enforced in relaxation method lead to an increased inlier ratio in reduced set E. This directly affects the number N of iterations. Hence, the preprocessing model converges faster to a correct solution.

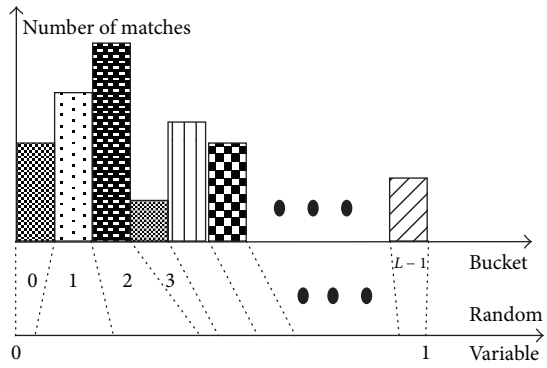


Figure 1: Monte Carlo sampling method.

2.2.2. Fundamental Matrix F Estimation. Zhang et al. [13] used LMedS technique to discard false matches and estimate fundamental matrix. However, when the inlier ratio is less than 50%, the result estimated by LMedS method may be unreliable. RANSAC is one of the robust methods for fundamental matrix estimation, which can obtain robust result even when the outlier ratio is more than 50%.

RANSAC is a stochastic optimization method, whose efficiency can be improved by Monte Carlo sampling method [14]. This method is shown in Figure 1. However, the sampling results may be very close to each other. Such a situation should be avoided because the estimation result may be instable and useless. The bucketing technique [14] is used to achieve higher stability and efficiency, which is shown in Figure 2. It works as follows. The min and max of the coordinates of the points are calculated in the first image. The region of the image is then divided into $b \times b$ buckets (shown in Figure 2). To each bucket is attached a set of feature points, and indirectly a set of correspondences, which fall into it. Those buckets which have no matches attached are excluded. In order to estimate fundamental matrix F, a subsample of 8 points should be generated. It is selected in 8 mutually different buckets, and then one match in each selected bucket is randomly selected.

Therefore, the fundamental matrix F can be estimated accurately and efficiently. This is the second step of the preprocessing model.

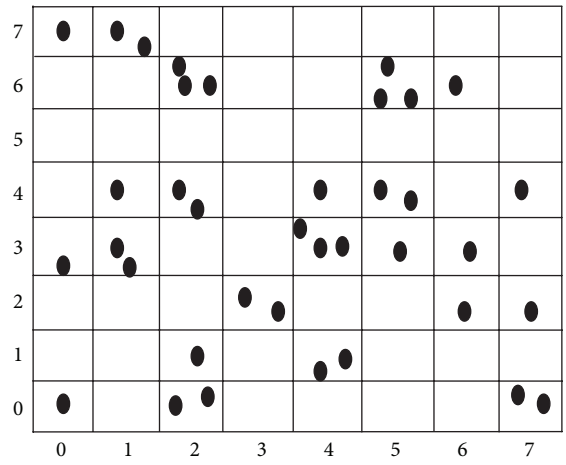


Figure 2: Bucketing technique.

(1) Computation of the reduced set E from initial matching set P
If $(q^* \geq q)$, store this pair in dataset E.
(2) RANSAC application
do
2.1 select the minimal sample s *in set* E
2.2 compute solution(s) for **F**
while $p = 1 - (1 - \varphi_{\text{red}}{}^s)^N < p_0$, **compute and store H(F).**
(3) Compute the hypothesis's support on full set P with matrix **H** or **F**

Algorithm 1: RANSAC with preprocessing model.

3. RANSAC Framework with Preprocessing Model

An improved RANSAC algorithm with preprocessing model is proposed in this section. This model can be easily integrated into the RANSAC procedure. The main idea is to suppose knowing some matching pairs being inliers with high probability, which are put into subset E $(E \subset P)$. Therefore, if RANSAC operates in subset E with the same confidence, it can calculate closer to the correct fundamental matrix **F** (or homography matrix **H**) with much less time of iteration. Thus, the preprocessing model can achieve the speedups in the whole RANSAC procedure. The steps of our framework are described as in Algorihm 1.

In Algorithm 1, q^* is the threshold of relaxation iteration. In this paper, q is set to 60. p_0 is the RANSAC threshold parameter, which is usually set to 95%. Let φ_{red} denote the ratio of inliers to all correspondences in set E. Then, the probability p that in N steps RANSAC ensures that at least one sampling within times N sampling, the elements are all inliers follow as $p = 1 - (1 - \varphi_{\text{red}}{}^s)^N$. Once matrix **F** is obtained in set E, we can additionally compute the hypothesis's support on the whole set P. In our experiments, we however only perform this last step to report the inlier numbers.

4. Experiment and Analysis

In the following, this paper experimentally evaluates the improved RANSAC and compares it with a classical approach. As we know, Harris and SIFT features are most commonly used in correspondence [15, 16]. In order to evaluate an approach comprehensively, choose both Harris and SIFT feature in initial corresponding. The environment of the experiments is Matlab R2010. Computer configuration is 2.10 G (CPU) and 4.00 G (RAM). The experimental images in this paper are from open databases: Visual Geometry Group, Peter Kovesi's home page, and the internet.

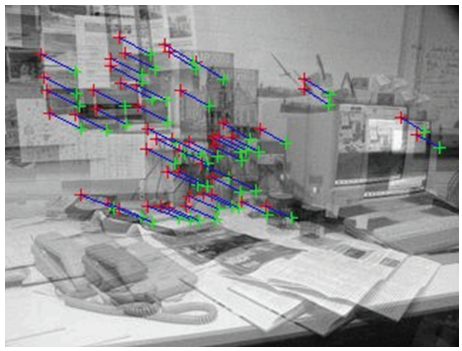
4.1. Experiment Based on Harris Feature. In the experiments based on Harris feature, this paper chooses match-by-correlation algorithm to obtain the initial matching set P. Then, the proposed RANSAC framework is operated on set P. The consequent of the Preprocessing Model directly determines the effect of the whole procedure. The reliable set E can be acquired by adjusting the model parameters.

Figure 3 is the comparison between our approach and the traditional RANSAC. Figure 3(a) shows the matching result calculated by our improved RANSAC. The result of traditional RANSAC method in the same experimental condition is shown in Figure 3(b). The numbers of iterations in Figures 3(a) and 3(b) are 260 and 361, respectively. 51/140 means extracting 51 inliers from 140 initial putative matching set. From the comparison, it is obvious that the result of our approach is better. The most important is that the iteration times are reduced. Thus, it can improve the runtime of RANSAC successfully. Compared with other improved RANSAC algorithms, our RANSAC framework can achieve the same result, while it is simpler to implement and the sampling times are reduced.

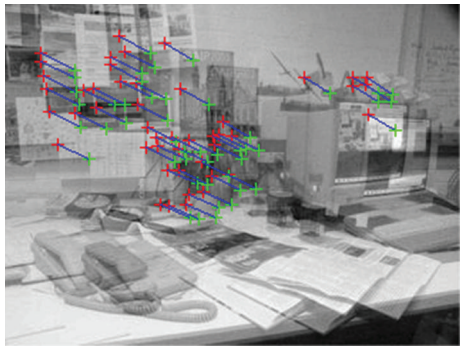
4.2. Experiment Based on SIFT Feature. Currently, SIFT is a popular and reliable algorithm to detect and describe local features in images. However, the initial matching by SIFT still exists in outliers. In this section, this paper uses the proposed approach to reject the outliers for the initial corresponding based on SIFT. The object is a model of scalp, which is usually used in biomedical modeling. The results are shown in Figure 4. Figure 4(a) is the result of initial matching by SIFT, and the number of pairs is 68. Figure 4(b) shows the result of our proposed RANSAC, the number of inliers is 50, and iteration times are 14. Figure 4(c) illuminates the result of classical RANSAC in the same experimental condition, the number of inliers is 42, and iteration times are 31.

From the comparison results in Figure 4, it can be found that our method is more effective for outlier rejection. Moreover, the iteration times are reduced to almost 45%. It is the most important benefit of our approach.

In conclusion, this paper argues that our method can be generally used in outlier rejection, no matter which kind of

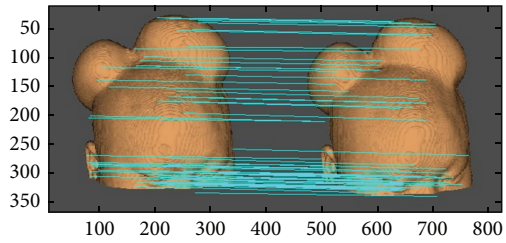


(a) 51/140

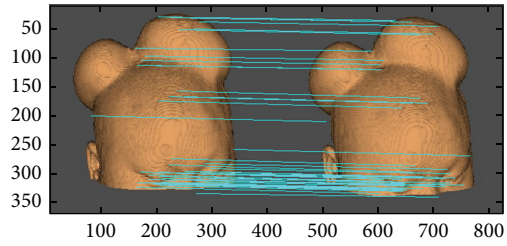


(b) 47/140

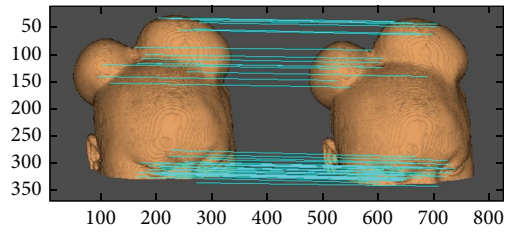
Figure 3: Comparison between our proposed RANSAC and traditional RANSAC.



(a) Initial matching by SIFT



(b) Result of our RANSAC approach



(c) Result of classical RANSAC

Figure 4: Results of the proposed method and classical RANSAC for correspondences based on SIFT.

feature is used. Moreover, the preprocessing model is adaptive for the condition of low-matching rate.

4.3. Analysis. As is shown above, the proposed RANSAC succeeds in reducing the iteration times. Our framework's success owes to the preprocessing model, which is effective for selecting the reliable corresponding pairs. Figure 5 illustrates the comparison of iteration times operating RANSAC in subset E and set P. It is obvious that there are huge differences especially when the initial matching rate is low. The main reason of the differences is that the elements of set E are much more reliable and with less scale. Through experimental statistics, it can be found that in the case of $\varphi \leq 0.6$, the proposed RANSAC needs much less iterations than direct RANSAC processing does. While if the condition of φ is selected in $0.6 \leq \varphi \leq 0.9$, the two methods usually have the same time complexity. Therefore, our model is beneficial to screen a reliable matching set E from the initial set P with lower matching rate φ and can reduce the followup of RANSAC iterations successfully.

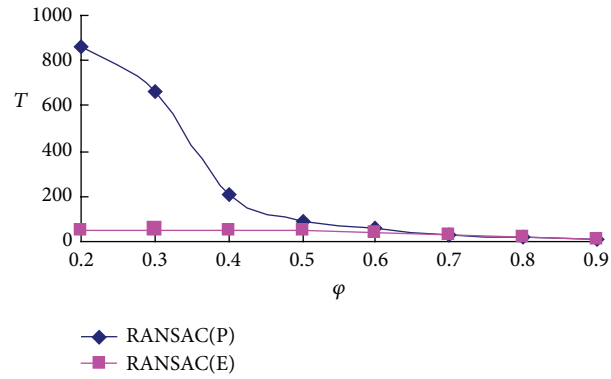


Figure 5: The number of iterations for RANSAC in set E and set P at the condition of different initial matching rates, T represents the iteration time of RANSAC, and φ means the initial matching rate.

5. Conclusion

In this paper, a novel framework was presented for improving RANSAC's efficiency in geometric matching applications. The improved RANSAC is based on Preprocessing Model that lets RANSAC operate on a reduced set of more confident correspondences with a higher inlier ratio. Compared with classic screening model, this model is simpler and efficient in implement, especially in the case of low-initial matching rate. The experimental results show that our approach can reduce much more iteration times especially when the initial matching rate is lower than 60%. In addition, the experiments were operated on two current features: Harris and SIFT. Therefore, it can be concluded that the proposed RANSAC framework is applicable.

In conclusion, this paper makes the following contributions: (1) this paper proposed a RANSAC framework which does not only rely on appearance but takes into account the quality of neighboring correspondences in the image space; (2) preprocessing model was introduced for selecting reduced set with higher inlier ratio, which improves runtime.

Acknowledgments

This work was supported by State Scholarship Fund from China Scholarship Council (no. 2011833105), Research Project of Department of Education of Zhejiang Province (no. Y201018160), Natural Science Foundation of Zhejiang Province (nos. Y1110649 and 61103140), and Commonweal Project of Science and Technology Department of Zhejiang Province (nos. 2012C33073 and 2010C33095), China.

References

[1] 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.

[2] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3D," *ACM Transactions on Graphics*, vol. 25, pp. 835–846, 2006.

[3] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, vol. 1–8, pp. 1545–1552, New York, NY, USA, June 2007.

[4] S. Chen, M. Zhao, G. Wu, C. Yao, and J. Zhang, "Recent advances in morphological cell image analysis," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 101536, 10 pages, 2012.

[5] S. Chen, Z. Wang, H. Tong, S. Liu, and B. Zhang, "Optimal feature matching for 3D reconstruction by combination of global and local information," *Intelligent Automation and Soft Computing*, vol. 17, no. 7, pp. 957–968, 2011.

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

[7] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, A. F. Martin and F. Oscar, Eds., pp. 726–740, Morgan Kaufmann, New York, NY, USA, 1987.

[8] T. Sattler, B. Leibe, and L. Kobbelt, "SCRAMSAC: improving RANSAC's efficiency with a spatial consistency filter," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 2090–2097, October 2009.

[9] O. Chum and J. Matas, "Optimal randomized RANSAC," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1472–1482, 2008.

[10] F. Mufti, R. Mahony, and J. Heinzmann, "Robust estimation of planar surfaces using spatio-temporal RANSAC for applications in autonomous vehicle navigation," *Robotics and Autonomous Systems*, vol. 60, pp. 16–28, 2012.

[11] L. Zhang, Z. Liu, and J. Jiao, "An improved RANSAC algorithm using within-class scatter matrix for fast image stitching," in *Image Processing: Algorithms and Systems IX*, J. T. Astola and K. O. Egiazarian, Eds., vol. 7870 of *Proceedings of SPIE*, San Francisco, Calif, USA, January 2011.

[12] J. Civera, O. G. Grasa, A. J. Davison, and J. M. M. Montiel, "1-point RANSAC for extended Kalman filtering: application to real-time structure from motion and visual odometry," *Journal of Field Robotics*, vol. 27, no. 5, pp. 609–631, 2010.

[13] Z. Zhang, R. Deriche, O. Faugeras, and Q. T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 87–119, 1995.

[14] Z. Zhang, "Determining the epipolar geometry and its uncertainty: a review," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 161–195, 1998.

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

[16] S. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 874517, 5 pages
http://dx.doi.org/10.1155/2013/874517

Research Article

Plane-Based Sampling for Ray Casting Algorithm in Sequential Medical Images

Lili Lin,[1] Shengyong Chen,[1] Yan Shao,[2] and Zichun Gu[2]

[1] *School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China*
[2] *Department of Plastic and Reconstructive Surgery, Sir Run Run Shaw Hospital, Medical College, Zhejiang University, Hangzhou 310016, China*

Correspondence should be addressed to Shengyong Chen; sy@ieee.org

Received 9 December 2012; Accepted 28 December 2012

Academic Editor: Carlo Cattani

This paper proposes a plane-based sampling method to improve the traditional Ray Casting Algorithm (RCA) for the fast reconstruction of a three-dimensional biomedical model from sequential images. In the novel method, the optical properties of all sampling points depend on the intersection points when a ray travels through an equidistant parallel plan cluster of the volume dataset. The results show that the method improves the rendering speed at over three times compared with the conventional algorithm and the image quality is well guaranteed.

1. Introduction

Modeling three-dimensional (3D) volume of biomedical tissues from 2D sequential images is an important technique to highly improve the diagnostic accuracy [1]. Volume rendering refers to the process that maps the 3D discrete digital data into image pixel values [2]. It can be classified into two categories: one is direct volume rendering which generates images by compositing pixel values along rays cast into a 3D image, and the other one is indirect volume rendering which visualizes geometry element graphics extracted from the volume data [3]. The importance of volume rendering is resampling and synthesizing image [4]. Ray casting, splatting, and shear-warp are the three popular volume rendering algorithms now [5].

Ray Casting Algorithm (RCA) is a direct volume rendering algorithm. The traditional RCA is widely used for it can precisely visualize various medical images with details of boundary and internal information from sequential images, while real-time rendering with traditional RCA is still an obstacle due to its huge computation.

In recent years, numerous techniques have been proposed to accelerate the rendering speed. In general, there are three primary aspects, including hardware-based, parallel, and software-based acceleration algorithms. Liu et al. [6] proposed a method combined that Graphics Processing Unit (GPU) and octree encoding and accelerated RCA at a rate of 85 times. Wei and Feng [7] presented a GPU-based real-time ray casting method for algebraic B-spline surfaces via iterative root-finding algorithms. Zhang et al. [8] accelerated RCA on Compute Unified Device Architecture (CUDA), which can perform more samplings within a ray segment using cubic B-spline.

However, both hardware-based and parallel techniques are inseparable from the development of computer hardware. By comparison, software-based algorithms can be quickly transplanted among different machines. What is more, they can show flexibility of the procedure and reflect the thoughts of researchers. Yang et al. [9] sampled points based on all intersection points at which the ray transacts with the voxel. All intersections in a voxel depend on four vertexes on one face. However, the condition whether two intersection points were on adjacent or opposite surface in a voxel was neglected. Ling and Qian [10] used a bounding volume method to avoid casting the viewing rays that do not intersect with the volume. Since such situation can be judged quickly by comparing the world coordinates of sampling point with the volume dataset, it did not obviously speed up the rendering process.

Recently, Qian et al. [11] replaced the sampling points with intersection points when rays travel through three groups of parallel planes along three orthometric axes to reduce the rendering time. However, it cannot guarantee the image density when the distance between adjacent parallel planes far surpasses the sampling interval.

This paper proposes an improved RCA to speed the rendering process. The main idea is, when the ray travels through one group of equidistant parallel planes of the volume, intersection points are obtained. Then the properties of sampling points between adjacent intersection points can be calculated by the formula of definite proportion and separated points. By this method, a small number of intersection points are considered; meanwhile the method does not sacrifice the sampling density.

2. Ray Casting Algorithm

2.1. Ray Casting Algorithm Overview. The traditional RCA involves two steps: (1) assign optical properties such as color and opacity to all 3D discrete vertexes according to their gray value, and (2) apply a sampling and composing process. For each output image pixel in sequence, do the following.

(i) Cast the ray through the volume from back to front.

(ii) Sample the color c_i and opacity a_i at each regular sampling point along the ray.

(iii) Set the color of the current output pixel according to

$$c_{\text{out}} = \sum_{i=0}^{n-1} c(i) \prod_{j=0}^{i-1} 1 - a(j) = c_0 + c_1(1 - a_0) + c_2(1 - a_1)(1 - a_0) + \cdots. \quad (1)$$

The rendering time is mainly comprised of four parts in the above-mentioned rendering process [11]. They are converting gray value into optical property (about 30%), computing position of sampling points (about 3%), sampling optical properties (about 39%), and compositing properties into output pixel color (about 6%). The time proportion of sampling is the highest. Moreover, the time ratio of four parts is not constant. The greater the sampling data is, the larger the proportion of sampling time is. Therefore, sampling has a direct impact on speed of RCA.

2.2. Traditional Sampling Method. Traditionally, the optical property of each sampling point depends on eight vertexes of its voxel by trilinear interpolation [12, 13]. In detail, there are four steps for the sampling one point. First, locate its voxel and convert the world coordinates of sampling point into voxel's local coordinates. The following three steps are processes of linear interpolations along three different axes in order. The interpolation diagram of Ray Casting Algorithm is shown in Figure 1.

For example, to sample point $S(x, y, z)$ in white circle (Figure 1), first obtain the voxel (i, j, k) and local coordinates (x_n, y_n, z_n) of S, which are expressed in (2). Then the optical property of four points (F_1, F_2, F_3, F_4) on the plane through

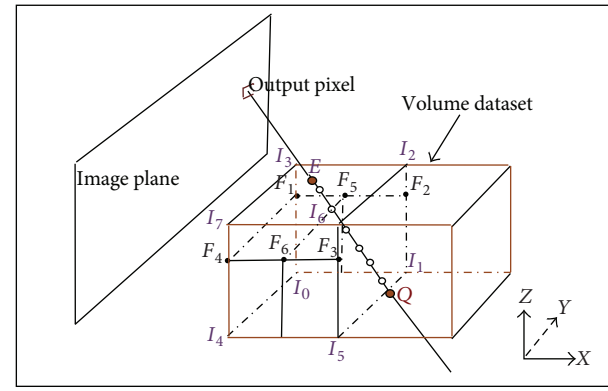


Figure 1: Interpolation for ray casting.

S is deduced according to eight vertexes ($I_0 \sim I_8$) along z-axis. The next property of two points (F_5, F_6) forming the line segment through S is computed along x-axis. At last S is obtained along y-axis by definite proportional division point formula.

In Figure 1, assume the pixel spacing along x-, y-, z- axes is Δx, Δy, and Δz, respectively, with $I_0(x_i, y_j, z_k)$:

$$i = \left[\frac{x}{\Delta x}\right], \quad j = \left[\frac{y}{\Delta y}\right], \quad k = \left[\frac{z}{\Delta z}\right],$$
$$x_i = i \times \Delta x, \quad y_j = j \times \Delta y, \quad z_k = k \times \Delta z, \quad (2)$$
$$x_n = \frac{x - x_i}{\Delta x}, \quad y_n = \frac{y - y_j}{\Delta y}, \quad z_n = \frac{z - z_k}{\Delta z},$$

where operator $[\cdot]$ represents taking the floor integral.

The property F of S can be calculated by F_5 and F_6, which are obtained by F_1, F_2, F_3, and F_4. The relationship between them is shown in

$$F_1 = I_0 + z_n \times (I_3 - I_0), \quad F_2 = I_1 + z_n \times (I_2 - I_1),$$
$$F_3 = I_5 + z_n \times (I_6 - I_5), \quad F_4 = I_4 + z_n \times (I_7 - I_4),$$
$$F_5 = F_1 + x_n \times (F_2 - F_1), \quad F_6 = F_4 + x_n \times (F_3 - F_4),$$
$$F = F_5 + y_n \times (F_6 - F_5). \quad (3)$$

According to the above equations, 17 additions and 16 multiplications are executed for sampling each point such as S (see Figure 1), including 3 additions and 9 multiplications to locate the voxel (i, j, k) and get the local coordinates. In Figure 1, there are 6 sampling points in two voxels, 102 additions, and 96 multiplications performed. To simplify the calculation of sampling process, a new RCA based on plane clusters sampling is proposed.

2.3. Proposed Plan-Based Sampling Method. The basic idea of the plan-based sampling method is to acquire all sampling points based on intersection points when ray travels through a group of parallel planes in the volume data field.

The sampling process, specifically, consists of three steps. First, intersections and the corresponding plane are obtained based on some necessary initial conditions. Then the optical property of all the intersection points is obtained by linear interpolation according to vertexes on plane clusters. The optical property of sampling points between intersection points along the ray is computed by definite proportion and separated point formula.

Assuming that the direction vector of ray is $\zeta = (r, m, n)$ and the extent of gridding volume data is $Ex \times Ey \times Ez$, with the spacing $\Delta x, \Delta y, \Delta z$ along x-, y-, z- axes, respectively, the three plane clusters are as follows:

$$\begin{aligned} X_i &= i\Delta x \quad (i = 0, 1, 2, \ldots, Ex - 1), \\ Y_j &= j\Delta y \quad (j = 0, 1, 2, \ldots, Ey - 1), \\ Z_k &= k\Delta z \quad (k = 0, 1, 2, \ldots, Ez - 1). \end{aligned} \tag{4}$$

Parallel plane clusters along y axis are selected. Let the origin point of ray be $O(x_o, y_o, z_o)$. The ray intersects with plane Y_j at entry point $E(V_i, V_j, V_k)$ and E belongs to the voxel (i, j, k). The coordinates of E and voxel (i, j, k) are deduced next. The derivation is shown as follows. Since

$$V_j = y_o + m \times t_j = j\Delta y \quad (j = 0, 1, 2, \ldots, Ey - 1), \tag{5}$$

where t_j means the distance from O to E along ray, the value of j can be obtained from

$$j = \left[\frac{y}{\Delta y} \right]. \tag{6}$$

Therefore,

$$t_j = \frac{j\Delta y - y_o}{m}, \tag{7}$$

and V_i, V_k *of* $E(V_i, V_j, V_k)$ can be expressed as follows:

$$V_i = x_o + r \times t_j; \qquad V_k = z_o + n \times t_j. \tag{8}$$

Considering that E belongs to voxel (i, j, k), then i and k are expressed as follows:

$$\begin{aligned} i &= \left[\frac{V_i}{\Delta x} \right], \\ k &= \left[\frac{V_k}{\Delta z} \right]. \end{aligned} \tag{9}$$

Therefore, when $j$is given, $E(V_i, V_j, V_k)$, i and k can be obtained through the above equations.

From the mathematical derivation, when original position, direction vector, and the extent of volume data are given, all the intersections and associated voxels can be quickly obtained.

In Figure 1, the property I_E of entry point E can be computed by the property (I_0, I_1, I_3) of three vertexes on voxel (i, j, k), that is,

$$I_E = I_0 + (I_1 - I_0)\left(\frac{V_i}{\Delta x} - i \right) + (I_3 - I_0)\left(\frac{V_k}{\Delta z} - k \right). \tag{10}$$

Table 1: Comparison of two sampling methods.

Objects and sizes	Head $512 \times 512 \times 295$	Heart $512 \times 512 \times 41$
Spacing (mm × mm × mm)	0.486 × 0.486 × 0.700	0.318 × 0.318 × 2.000
Sampling distance (mm)	0.3	0.3
Time by the traditional (s)	58.274	7.192
Time by the proposed (s)	17.158	2.043
Acceleration rate	3.606	3.52

In the same way, the property I_Q of exit point Q can be obtained. At last the property I_S is expressed as follows:

$$I_S = I_E + \frac{t - t_j}{t_{j+1} - t_j}\left(I_Q - I_E \right). \tag{11}$$

In addition, when one component of the direction vector ζ is zero, a plane cluster along another axis can be chosen. If two components are zero, the plane clusters along the third axis are taken into account.

2.4. Comparison of Two Sampling Methods. In the new RCA sampling process, only intersection points on a plane cluster along one axis need to be considered without converting coordinates. While in the conventional sampling process, the world coordinates of each sampling point are converted into voxel's local coordinates and computed by trilinear interpolation [14, 15].

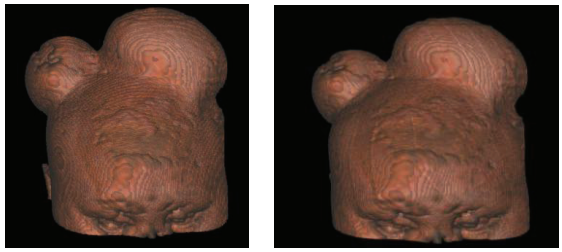
As is shown in Figure 1, there are 6 sampling points between E and Q. 15 additions and 19 multiplications are executed to sample E and Q, and 24 additions and 12 multiplications are run to sample six points based on E and Q. Totally, 39 additions and 31 multiplications are taken compared with 102 additions and 96 multiplications with trilinear interpolation. Furthermore, not all vertexes are referred because some vertexes (such as I_4, I_7, I_2 in Figure 1) are not used as reference by the new method. Thus, in theory, the calculation amount is reduced to less than one third on the whole.

3. Experiments and Analysis

3.1. Data. Experiments are carried out on head CT sequences and heart CT sequences. Both sequences are scanned by Siemens spiral CT. The detail information is shown in Table 1. Taking head for an example, the extents are $512 \times 512 \times 295$, and the pixel spacing is 0.486 mm, 0.486 mm and 0.700 mm along x-, y-, z- axis, respectively. The sampling distance along ray is 0.3 mm.

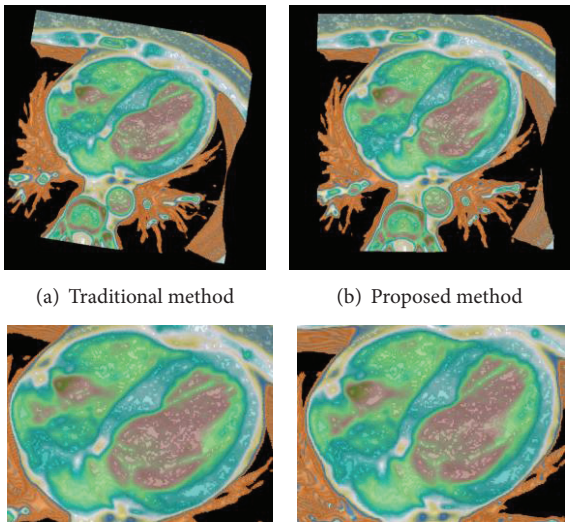
3.2. Results. The reconstructed results of two datasets are shown in Figures 2 and 3. The rendering time of the data is shown in Table 1. For example, it takes 17.158 seconds to render the head sequences with the new sampling method, while 58.274 seconds using the traditional method.

3.3. Analysis. The new sampling method does not consult all 3D vertexes of the volume data. For this reason, it is



(a) Traditional method (b) Proposed method

Figure 2: Head images of ray casting.



(a) Traditional method (b) Proposed method

(c) the details of (a) (d) the details of (b)

Figure 3: Heart images with ray casting.

a question whether the image quality can be guaranteed. It can be seen in Figures 2 and 3 that images reconstructed by RCA based on plan cluster sampling method are almost the same as those based on traditional trilinear interpolation in RCA. They can clearly show the details of the boundary and internal information of the volume with the new sampling method. Therefore, the image quality can be well ensured.

By comparing the amount of computation (39/102-31/96) in the two sampling methods, the new method can reduce the amount of traditional one to about one third. It can be seen that the total rendering time (Table 1) using new method is less than one third of that using conventional trilinear interpolation. It indicates that the time saved to inquire the property of the vertexes not for reference should not be underestimated.

Moreover, it is shown that the acceleration rate of the head images is higher than that of the heart images. The main difference between them is that the spacing of head CT sequences is denser than the heart data. Therefore, the denser the data is, the more efficient the new method is.

4. Conclusion

This paper presented a novel RCA based on a parallel plan cluster sampling method. The proposed method can efficiently speed up the sampling process at more than three times and still clearly display the boundary and internal information of the volume; thus the image quality is well guaranteed. In addition, the comparison of acceleration rate indicates that the new method is more effective for dataset with denser spacing. The new method can meet the real-time requirements of interactive rendering.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61105073, 61173096, and 61103140) and the Science and Technology Department of Zhejiang Province (R1110679 and 2010C33095).

References

[1] C. Cattani, R. Badea, S. Y. Chen, and M. Crisan, "Biomedical signal processing and modeling complexity of living systems," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 298634, 2 pages, 2012.

[2] Y. Mishchenko, "Automation of 3D reconstruction of neural tissue from large volume of conventional serial section transmission electron micrographs," *Journal of Neuroscience Methods*, vol. 176, no. 2, pp. 276–289, 2009.

[3] B. Lee, J. Yun, J. Seo, B. Shim, Y. G. Shin, and B. Kim, "Fast high-quality volume ray casting with virtual samplings," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1525–1532, 2010.

[4] S. Y. Chen and X. Li, "Functional magnetic resonance imaging for imaging neural activity in the human brain: the annual progress," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 613465, 9 pages, 2012.

[5] N. Max, "Optical models for direct volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, no. 2, pp. 99–108, 1995.

[6] B. Q. Liu, G. J. Clapworthy, F. Dong, and E. C. Prakash, "Octree rasterization accelerating high-quality out-of-core GPU volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, no. 99, pp. 1–14, 2012.

[7] F. F. Wei and J. Q. Feng, "Real-time ray casting of algebraic B-spline surfaces," *Computers & Graphics*, vol. 35, no. 4, pp. 800–809, 2011.

[8] C. G. Zhang, P. Xi, and C. X. Zhang, "CUDA-based volume ray-casting using cubic B-spline," in *Proceedings of the International Conference on Virtual Reality and Visualization (ICVRV '11)*, pp. 84–88, November 2011.

[9] A. R. Yang, C. X. Lin, and J. Z. Luo, "A ray-casting approach based on rapid direct interpolation," *Control & Automation*, vol. 26, no. 7, pp. 8–10, 2010.

[10] L. Tao and Z. Y. Qian, "An improved fast ray casting volume rendering algorithm of medical image," in *Proceedings of the 4th International Conference on Biomedical Engineering and Informatics (BMEI '11)*, pp. 109–112, 2011.

[11] Y. Qian, X. Zhang, and J. Lai, "Improved ray casting algorithm," *Computer Engineering and Design*, vol. 32, no. 11, pp. 3780–3783, 2011.

[12] J. Meyer-Spradow, T. Ropinski, J. Mensmann, and K. Hinrichs, "Voreen: a rapid-prototyping environment for ray-casting-based volume visualizations," *IEEE Computer Graphics and Applications*, vol. 29, no. 6, pp. 6–13, 2009.

[13] H. R. Ke and R. C. Chang, "Ray-cast volume rendering accelerated by incremental trilinear interpolation and cell templates," *The Visual Computer*, vol. 11, no. 6, pp. 297–308, 1995.

[14] B. Lee, J. Yun, J. Seo, B. Shim, Y. G. Shin, and B. Kim, "Fast high-quality volume ray casting with virtualsamplings," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1525–1532, 2010.

[15] A. Knoll, Y. Hijazi, R. Westerteiger, M. Schott, C. Hansen, and H. Hagen, "Volume ray casting with peak finding and differential sampling," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1571–1578, 2009.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2012, Article ID 125321, 7 pages
doi:10.1155/2012/125321

Research Article

Self-Adaptive Image Reconstruction Inspired by Insect Compound Eye Mechanism

Jiahua Zhang,[1] Aiye Shi,[1] Xin Wang,[1] Linjie Bian,[2] Fengchen Huang,[1] and Lizhong Xu[1]

[1] *College of Computer and Information Engineering, Hohai University, Nanjing, Jiangsu 211100, China*
[2] *College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China*

Correspondence should be addressed to Lizhong Xu, lzhxu@hhu.edu.cn

Received 23 November 2012; Accepted 17 December 2012

Academic Editor: Sheng-yong Chen

Inspired by the mechanism of imaging and adaptation to luminosity in insect compound eyes (ICE), we propose an ICE-based adaptive reconstruction method (ARM-ICE), which can adjust the sampling vision field of image according to the environment light intensity. The target scene can be compressive, sampled independently with multichannel through ARM-ICE. Meanwhile, ARM-ICE can regulate the visual field of sampling to control imaging according to the environment light intensity. Based on the compressed sensing joint sparse model (JSM-1), we establish an information processing system of ARM-ICE. The simulation of a four-channel ARM-ICE system shows that the new method improves the peak signal-to-noise ratio (PSNR) and resolution of the reconstructed target scene under two different cases of light intensity. Furthermore, there is no distinct block effect in the result, and the edge of the reconstructed image is smoother than that obtained by the other two reconstruction methods in this work.

1. Introduction

The classical reconstruction methods include the nearest neighbor algorithm, bilinear interpolation, and bicubic interpolation algorithm [1, 2]. According to existing research, the reconstruction accuracy of bilinear interpolation is higher than that of the nearest neighbor algorithm, and the former can get better image reconstruction results. However, the reconstructed image by bilinear interpolation appears saw-tooth and blurring sometimes [3]. Although the reconstruction results of bicubic interpolation are better than the others, they always lose efficiency and take much more time. As a compromise, bilinear interpolation is often used for research. These algorithms can improve the reconstruction quality of the original image to some extent. However, only the correlation between the local and global pixels is considered in these algorithms. Interpolation-based reconstruction methods do improve the effect of image reconstruction, but they destroy the high-frequency detailed information of the original image [4, 5].

Some studies have found that insects have a relatively broad living environment, for instance, the mantis shrimp can live between 50 m and 100 m depth underwater. In such living environment, the light condition changes dramatically, due to the combined effect of sunlight and water media. To adapt to the changing environment, this species, whose ommatidia structure is fixed, must regulate the light acceptance angle adaptively [6, 7]. Through the joint action of the lens and the rhabdome, the mantis shrimp has different degrees of overlapping images in the whole region of the ommatidia. The ommatidia get the different optical information depending on the different lighting conditions. Under the light and the dim environment conditions, the mantis shrimp can regulate the length of rhabdome and lens through relaxing or contracting the myofilament. Based on the biological mechanism above, the ommatidia visual field can be narrowed or expanded to get a relatively stable number of incoming photons and a better spatial resolution. Ultimately, the imaging system can reach balance between the visual field and the resolution [8], as shown in Figure 1. According to Schiff's [9] research, the imaging angle and visual field of the mantis shrimp ommatidia both change while the light intensity condition changes. For instance, the ommatidia visual field is 5° under dim-adapted pattern, but

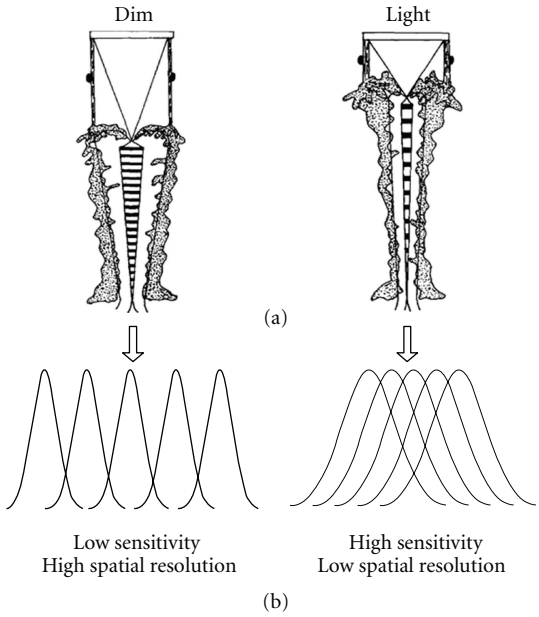


Figure 1: Light-dim adaptive regulatory mechanism of ommatidia. (a) Structure adaptation in ommatidia visual system. (b) Adaptation in the view-field of ommatidia and compound eyes.

the corresponding visual field will be only 2° under bright-adapted pattern, and some other species also have similar characteristics [10–14].

Recently, the compressed sensing theory provides a new approach for computer vision [15–17], image acquisition [18, 19], and reconstruction [20–22]. This method can get the reconstruction results as effectively as the traditional imaging systems do, or even higher quality (in resolution, SNR, etc.), with fewer sensors, lower sampling rate, less data volume, and lower power consumption [23–27]. According to the compressed sensing theory, the compressive sampling can be executed effectively if there is a corresponding sparse representation space. Currently, the compressed sensing theory and application of the independent-channel signal have been developed in-depth, such as single-pixel camera imaging [28].

By the combined insect compound eye imaging mechanism with compressed sensing joint sparse model (JSM-1) model [29–32], we use the spatial correlation of multiple sampled signals to get the compressive sampling and reconstruction. Inspired by the light-dim self-adaptive regulatory mechanism of insect compound eyes (ICE), this paper proposes an ICE-based adaptive reconstruction method (ARM-ICE). The new method can execute multiple compressive sampling on the target scene. According to the environment light intensity, it can regulate the sampling visual field to control imaging. The simulation results show that, in contrast to the image-by-image reconstruction and bilinear interpolation algorithm, the new method can reconstruct the target scene image under two kinds of light intensity conditions with higher-peak signal-to-noise ratio (PSNR). The new method also improves the resolution and detailed information of reconstruction.

In the first section, we describe the imaging control mechanism of insect compound eyes, compressed sensing theory, and current research of bionic compound eyes imaging system. Section 2 demonstrates the ARM-ICE imaging system pattern from three aspects: visual field self-adaptive adjusting, sampling, and reconstruction. Section 3 completes the ARM-ICE system simulation under the dim and light conditions and then analyzes the imaging results and the comparison of relevant parameters. In Section 4, we conclude with possible topics for future work.

2. Compressed Sensing-Based Arm-Ice Imaging System Pattern

Figure 2 shows an ARM-ICE imaging system pattern. The purple lines represent the light environment visual field, while the blue lines represent the dim environment visual field. The target scene is imaged, respectively, by the compound eye lens array. The isolation layer is composed by multichannel opening shade blocks, which can be controlled. And each port of shade blocks is connected to a corresponding little lens of compound eye lenses. This structure sets a number of independent controllable light-sensitive cells. Each port of isolation layer opens at different time. The feedback signal controls them to regulate the relative position to make the light from target scene to the n light-sensitive cells. The corresponding area is sparsely sampled in the digital micromirror device. Measurement data can be obtained in the imaging plane. Ultimately, the processor reconstructs the target scene according to the k-sparse property of data sensed on the wavelet basis Ψ and the uncorrelated measurement matrix Φ.

2.1. Arm-ICE Visual Field Self-Adaptive Regulation. According to the biological research, in the insect compound eyes system under different light intensities, the angle of imaging and the visual field change accordingly [33–37]. Inspired by this self-adaptive ability, this paper mimics the insect compound eye system on its imaging control mechanism based on light intensity sensitivity, to expand or narrow the scope of visual field and overlapping field by regulating the position of the lenses.

According to the results of biological research, the relationship between light intensity, imaging pore size, and other factors can be described as (1), hereby to regulate the lenses position to achieve the overlap visual field [12]

$$\Delta\rho_T = \frac{0.530}{\upsilon_{\max}}\sqrt{\ln cN_p - \frac{1}{2}\ln\left[N_p + \sigma_D^2\right]}, \quad (1)$$

where $\Delta\rho_T$ indicates the visual field range, $\upsilon_{\max}$ indicates the maximum detectable spatial frequency, which can be regarded as a constant, c is the mean contrast of the scene, N_p indicates the number of the photons captured by an input port, and σ_D^2 shows the total variance for environmental light intensity.

From (1), the visual field can be calculated according to the $\upsilon_{\max}$ set while the light intensity changes. Based on

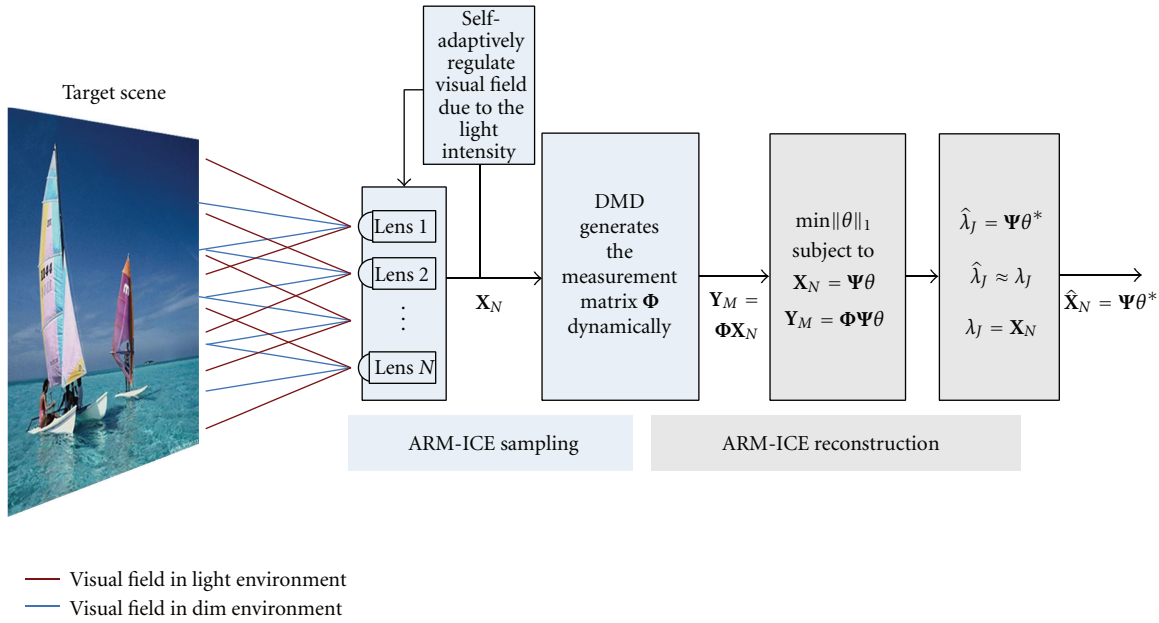


Figure 2: ARM-ICE imaging system pattern.

the biological principle above, the visual field range can be regulated according to the environment light intensity.

2.2. Compressive Sampling. The digital micromirror device (DMD) senses the optical information from the lenses array, and then makes sparse sampling. The principle is inner product the optical signal from the lenses array perception $\mathbf{X}(m)$ and DMD measurement basis vector $\boldsymbol{\varphi}(m)$, and make the result as the output voltage $(v)m$ of the DMD device at the moment m. The output voltage $v(m)$ of the photodiode can be expressed as the inner product of the desired image x with a measurement basis vector [26, 28, 29]:

$$v(m) \propto \langle \mathbf{X}(m), \boldsymbol{\varphi}(m) \rangle + O_{\mathrm{DC}}, \tag{2}$$

where the value of $\boldsymbol{\varphi}(m)$ is related to the position of DMD micro-mirror; when the micromirror turns $+10^\circ$, $\phi_i(m) = 1$; when the micromirror turns -10°, $\phi_i(m) = 0$. O_{DC} is the direct current offset, which can be measured by setting all mirrors to -10°.

Based on the principle of measurement matrix of a single DMD device, we can use the DMD device array to get sparse signals of image system. The compound eye lenses and the isolation layer constitute n light-sensitive independent cells, each of which is controlled by the isolation layer to open at different time. The array jointly senses the target scene data X_i:

$$X_i = X_{i,C} + X_{i,S}, \tag{3}$$

where $X_{i,C}$ expresses the common information of the perception data and $X_{i,S}$ expresses the specific information of each lens. Vector $\mathbf{X}_N = (X_1, X_2, \ldots, X_N)^T$ indicates the perception data from n light-sensitive units. The perception data can be regarded as k-sparse on wavelets basis $\mathbf{\Psi}$ due to the spatial correlation:

$$\mathbf{X}_N = \mathbf{\Psi}\boldsymbol{\theta}, \tag{4}$$

where $\boldsymbol{\theta} = (\lambda_0, \gamma_0, \gamma_1, \ldots, \gamma_{J-1})^T$ is the sparse vector coefficient, consisting of the high-frequency subset $\gamma_0, \gamma_1, \ldots, \gamma_{J-1}$ (γ_k is subset at scale $J - k$) and the low-frequency subset λ_0 of wavelet transform. After light-sensitive lenses obtain X_N, k-sparse signal X_N is used to generate M measurement data of the image plane from the $M \times N$ measurement matrix $\mathbf{\Phi}$ on the DMD device:

$$\mathbf{Y}_M = (Y_1, Y_2, \ldots, Y_M)^T = \mathbf{\Phi}\mathbf{X}_N, \tag{5}$$

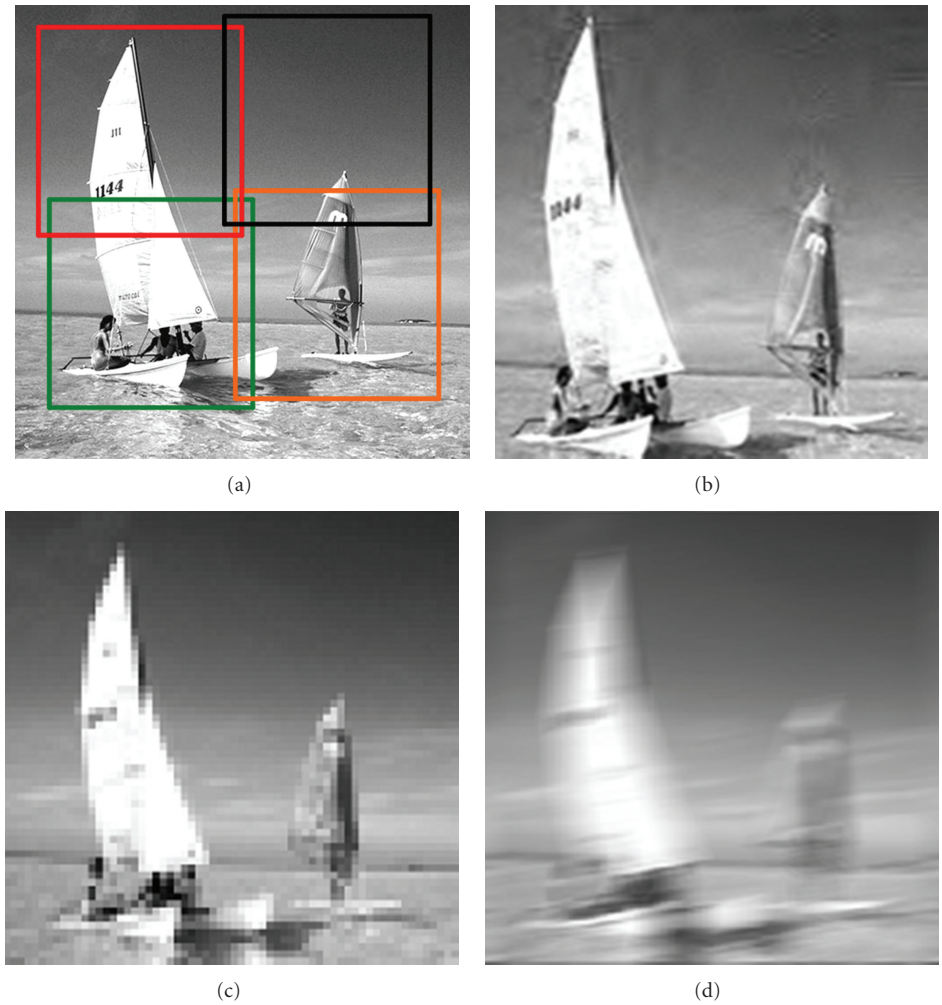
where matrix $\mathbf{\Phi}$ is a 0-1 matrix, which consists of the output voltage $v(m)$ of the DMD device in (2) at the moment m. Equation (5) can also be described as follows:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix} = \begin{bmatrix} \Phi_1 & & & 0 \\ & \Phi_2 & & \\ & & \ddots & \\ 0 & & & \Phi_M \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}. \tag{6}$$

2.3. Joint Reconstrucion. According to the multichannel captured data, which are k-sparse on wavelet basis and the inconsistency of the measurement matrix $\mathbf{\Phi}$ with the wavelet basis $\mathbf{\Psi}$, the processor runs the decoding algorithm to reconstruct the target scene:

$$\min \|\theta\|_0, \quad \text{subject to } \mathbf{Y}_M = \mathbf{\Phi}\mathbf{\Psi}\theta. \tag{7}$$

The optimized sparse solution θ^* can be gotten by solving the issue of optimizing l_0 norm. The reconstruction of captured data from each lens can be indicated as follows:



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

Figure 3: ARM-ICE imaging results and comparison under strong light: (a) target scene, whose brightness value is 144.8527 Nits; (b) ARM-ICE reconstructed image, whose PSNR is 41.9113 dB; (c) result of bilinear interpolation reconstruction, whose PSNR is 34.9112 dB; (d) result of image-by-image reconstruction, whose PSNR is 27.8246 dB.

$\hat{X}_N = (\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_N)^T = \Psi\theta^*$. An important issue during the reconstruction process is how to calculate the wavelet basis Ψ. Assume the set of captured data X_N is already known, and $\lambda_J = X_N$. Each light-sensitive sensor captures the target scene from different views, so its obtained data can be divided into two parts: the common part $\lambda_{J,P}$ and the particular part $\lambda_{J,D}$. T indicates the lifting wavelet transform after J times' recursion:

$$\text{for } k = J \text{ to } 1 \begin{cases} \lambda_{k-1} = \lambda_{k,P} + U(\gamma_{k-1}), \\ \gamma_{k-1} = \lambda_{k,D} - P(\lambda_{k,P}), \\ T(\lambda_k) = (\lambda_{k-1}, \gamma_{k-1}), \end{cases} \tag{8}$$

where λ_{k-1} is the low-frequency coefficient set, γ_{k-1} is the high-frequency coefficient set, P is the linear prediction operator, and U is the linear update operator. Using the spatial correlation of captured data, $\lambda_{k,D}$ can be calculated by $\lambda_{k,P}$. γ_{k-1} contains fewer information relatively.

For λ_k, after k times' recursive lifting wavelet transform:

$$T^k(\lambda_k) = \{\lambda_0, \hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_{k-1}\}. \tag{9}$$

After resetting the wavelet coefficients which are under threshold value in γ_i, the sparsely structured $\hat{\gamma}_i$ can be used to reconstruct the original signal λ_k exactly. Assuming that $T^{-k}(\bullet)$ is a lifting wavelet inverse transform, as the linear prediction operator and the linear update operator are both linear operations, therefore$T^k(\bullet)$ and $T^{-k}(\bullet)$ are both linear transforms. $T^{-k}(\bullet)$ can be expressed as follows:

$$\begin{aligned} T^{-K}(\lambda_0, \hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_{k-1}) &= \hat{\lambda}_k, \\ \hat{\lambda}_K &= \Psi\theta^* \approx \lambda_k, \end{aligned} \tag{10}$$

where $\theta^* = (\lambda_0, \hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_{k-1})^T$. Since $\lambda_J = \mathbf{X}_N$, the initial data $\hat{\mathbf{X}}_N = \Psi\theta^*$ can be reconstructed exactly.

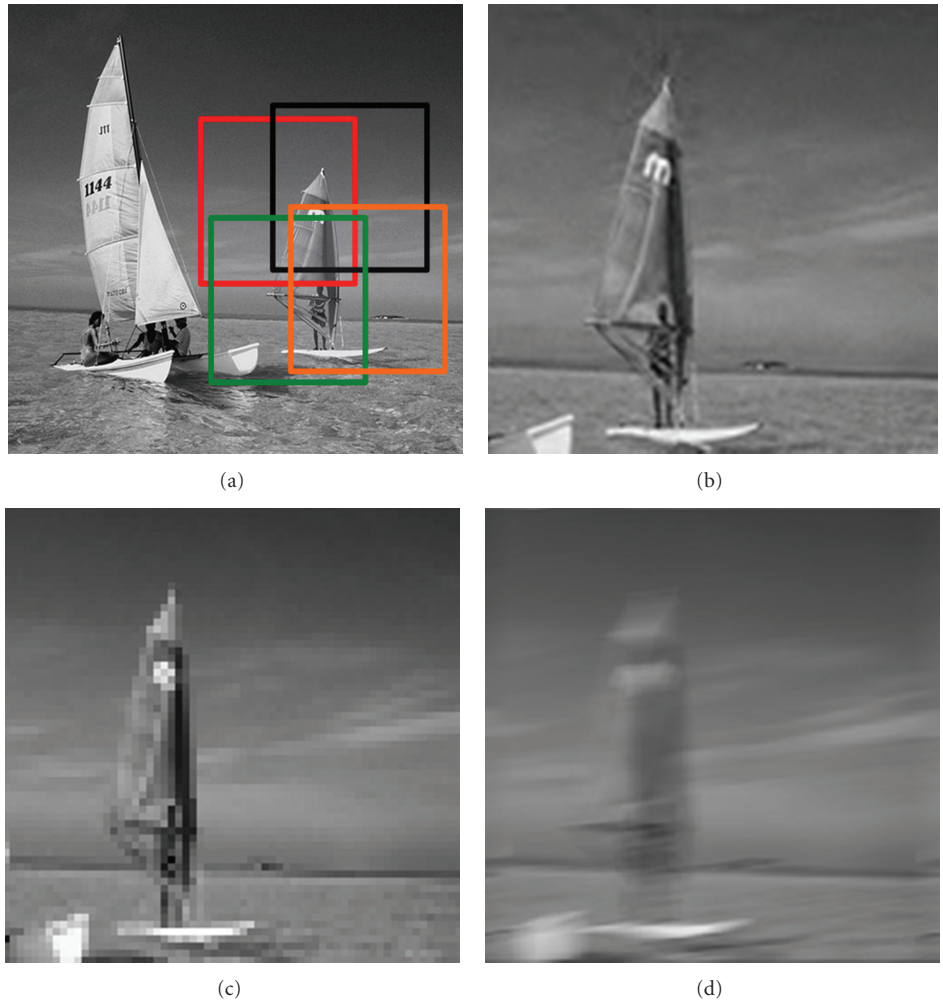


Figure 4: ARM-ICE imaging results and comparison under low light: (a) target scene, whose brightness value is 103.3661 Nits; (b) ARM-ICE reconstructed image, whose PSNR is 44.4705 dB; (c) result of bilinear interpolation reconstruction, whose PSNR is 36.5021 dB; (d) result of image-by-image reconstruction, whose PSNR is 29.5852 dB.

3. Four-Channel Arm-ICE Imaging System Pattern Simulation

According to the ARM-ICE visual field self-adaptive adjustment mechanism under different surrounding light intensities described in Section 2.1, in this section, we simulate a four-channel ARM-ICE imaging system. When the surrounding light intensity turns strong, the lenses array regulates their relative positions according to (1) automatically. The simulation results are shown in Figure 3; Figure 3(a) is the target scene under strong illumination environment, whose brightness value is 144.8527 Nits. Figure 3(b) is the joint reconstruction image from photoelectric coupler array, and its reconstructed PSNR is 41.9113 dB. Figure 3(c) is a reconstructed image by linear interpolation method, and its PSNR is 27.8246 dB under the same sampling rate as ARM-ICE. Figure 3(d) is an image-by-image reconstruction, and its PSNR is 27.8246 dB under the same sampling rate as ARM-ICE.

When the surroundings are dim, the compound eye lenses array contracts to the central area, sacrificing the visual field to improve the reconstruction resolution of target scene. The simulation results are shown in Figure 4. Figure 4(a) is the target scene under the dim conditions whose brightness value is 103.3661 Nits. Put the brightness values into (1) and calculate the lenses' positions at the moment. Figure 4(b) is the joint reconstruction image from photoelectric coupler array, and its reconstructed PSNR is 44.4705 dB. Figure 4(c) is the reconstructed image by linear interpolation method. PSNR is 36.5021 dB at the same sampling rate. Figure 4(d) is the reconstruction result of image-by-image, whose PSNR is 29.5852 dB.

From the reconstruction effect, the result of linear interpolation method is superior to the result reconstructed by image-by-image. However, there is still obvious block effect, and lack of smoothness at the edge direction. Correspondingly, the image reconstructed by ARM-ICE has a significant improvement in resolution. From Figures 3 and 4,

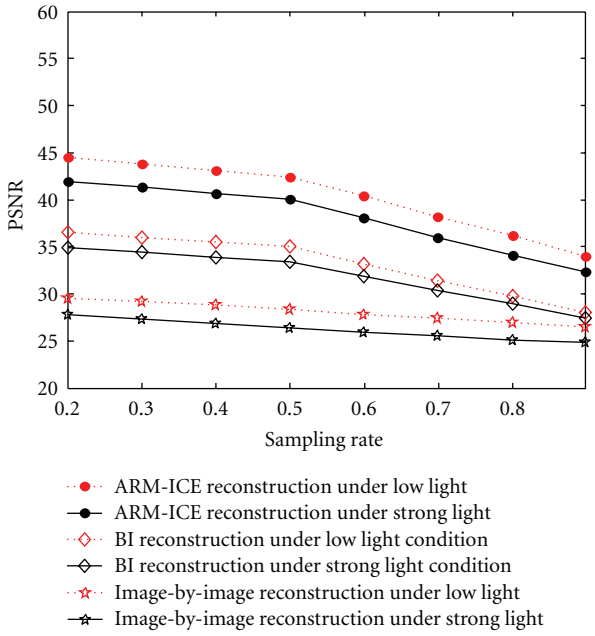


Figure 5: The comparison of PSNR-Sampling rates under low light and strong light conditions.

we can see that there is no distinct block effect in the result and the edges of the reconstructed image are smoother compared to the results of the other two reconstruction methods studied in this work.

Figure 5 is the comparison of PSNR-Sampling rates under low light and strong light conditions (144.8527 Nits). The three black lines in the figure show the comparison results under the strong light condition, in which the black dotted line shows the result of ARM-ICE, the black diamond line shows the result of bilinear interpolation, and the black five-pointed star-shaped line shows the result of image-by-image reconstruction. It can be concluded from the figure that the PSNR of ARM-ICE is higher than bilinear interpolation and image-by-image reconstruction under different sampling rates under the strong light condition.

The three red lines in the figure show the comparison obtained under the low light condition (103.3661 Nits), in which the red dotted line shows the result of ARM-ICE reconstruction, the red diamond line shows the result of bilinear interpolation, and the red five-pointed star-shaped line shows the result of image-by-image reconstruction. It can be seen from the figure that when the target scene is under low light condition, the PSNR of ARM-ICE at different sampling rates is higher than bilinear interpolation and image-by-image reconstruction.

4. Conclusion

Inspired by the imaging mechanism and the adaptive regulatory regulation mechanism of the insect compound eyes, this paper proposes a reconstruction method, which regulates the scale of the sampling area adaptively according to the surrounding light intensity condition. The imaging system pattern of the new method can complete the multichannel independent sampling in the target scene almost at the same time. Meanwhile, the scale of the sampling area and the optical signal redundancy can be regulated adaptively to achieve the imaging control. Compared with the traditional methods, the resolution of the reconstructed image by ARM-ICE method has been significantly improved. The reconstructed image with the proposed method has three features: higher resolution, no distinct block effect, and smooth edge.

Simulation results indicate that the new method makes the PSNR of the reconstructed image higher under two kinds of light conditions. However, the reconstruction quality under low light conditions is improved by the proposed algorithm at the cost of the scale of the visual field. Therefore, the key issue in the future work would be how to reconstruct high-resolution large scenes in low light conditions.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (No. 61263029 and No. 61271386). The authors thank Wang Hui, a graduate student of Hohai University, for helping in research work.

References

[1] R. C. Kenneth and R. E. Woods, *Digital Image Processing*, Publishing House of Electronics Industry, Beijing, China, 2002.

[2] F. G. B. D. Natale, G. S. Desoli, and D. D. Giusto, "Adaptive least-squares bilinear interpolation (ALSBI): a new approach to image-data compression," *Electronics Letters*, vol. 29, no. 18, pp. 1638–1640, 1993.

[3] L. Chen and C. M. Gao, "Fast discrete bilinear interpolation algorithm," *Computer Engineering and Design*, vol. 28, p. 15, 2007.

[4] S. Y. Chen and Z. J. Wang, "Acceleration strategies in generalized belief propagation," *IEEE Transactions on Industrial Informatics*, vol. 8, p. 1, 2012.

[5] N. M. Kwok, X. P. Jia, D. Wang et al., "Visual impact enhancement via image histogram smoothing and continuous intensity relocation," *Computers & Electrical Engineering*, vol. 37, p. 5, 2011.

[6] L. Z. Xu, M. Li, A. Y. Shi et al., "Feature detector model for multi-spectral remote sensing image inspired by insect visual system," *Acta Electronica Sinica*, vol. 39, p. 11, 2011.

[7] F. C. Huang, M. Li, A. Y. Shi et al., "Insect visual system inspired small target detection for multi-spectral remotely sensed images," *Journal on Communications*, vol. 32, p. 9, 2011.

[8] H. Schiff, "A discussion of light scattering in the Squilla rhabdom," *Kybernetik*, vol. 14, no. 3, pp. 127–134, 1974.

[9] B. Dore, H. Schiff, and M. Boido, "Photomechanical adaptation in the eyes of Squilla mantis (Crustacea, Stomatopoda)," *Italian Journal of Zoology*, vol. 72, no. 3, pp. 189–199, 2005.

[10] B. Greiner, "Adaptations for nocturnal vision in insect apposition eyes," *International Review of Cytology*, vol. 250, pp. 1–46, 2006.

[11] A. Horridge, "The spatial resolutions of the apposition compound eye and its neuro-sensory feature detectors: observation versus theory," *Journal of Insect Physiology*, vol. 51, no. 3, pp. 243–266, 2005.

[12] H. Ikeno, "A reconstruction method of projection image on worker honeybees' compound eye," *Neurocomputing*, vol. 52–54, pp. 561–566, 2003.

[13] J. Gál, T. Miyazaki, and V. B. Meyer-Rochow, "Computational determination of refractive index distribution in the crystalline cones of the compound eye of Antarctic krill (*Euphausia superba*)," *Journal of Theoretical Biology*, vol. 244, no. 2, pp. 318–325, 2007.

[14] 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.

[15] V. Cevher, P. Indyk, L. Carin, and R. Baraniuk, "Sparse signal recovery and acquisition with graphical models," *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 92–103, 2010.

[16] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *IEEE Transactions on Signal Processing*, vol. 6, 2011.

[17] L. Z. Xu, X. F. Ding, X. Wang, G. F. Lv, and F. C. Huang, "Trust region based sequential quasi-Monte Carlo filter," *Acta Electronica Sinica*, vol. 39, no. 3, pp. 24–30, 2011.

[18] J. Treichler and M. A. Davenport, "Dynamic range and compressive sensing acquisition receivers," in *Proceedings of the Defense Applications of Signal Processing (DASP '11)*, 2011.

[19] 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.

[20] 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.

[21] C. Hegde and R. G. Baraniuk, "Sampling and recovery of pulse streams," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1505–1517, 2011.

[22] A. Y. Shi, L. Z. Xu, and F. Xu, "Multispectral and panchromatic image fusion based on improved bilateral filter," *Journal of Applied Remote Sensing*, vol. 5, Article ID 053542, 2011.

[23] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[24] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[25] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[26] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[27] L. Z. Xu, X. F. Li, and S. X. Yang, "Wireless network and communication signal processing," *Intelligent Automation & Soft Computing*, vol. 17, pp. 1019–1021, 2011.

[28] D. Takhar, J. N. Laska, M. B. Wakin et al., "A new compressive imaging camera architecture using optical-domain compression," in *Computational Imaging IV*, vol. 6065 of *Proceedings of SPIE*, January 2006.

[29] D. Baron, B. Wakin, and S. Sarvotham, "Distributed Compressed Sensing," Rice University, 2006.

[30] D. Baron and M. F. Duarte, "An information-theoretic approach to distributed compressed sensing," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, vol. 43, Allerton, Ill, USA, 2005.

[31] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proceedings of the 39th Asilomar Conference on Signals, Systems and Computers*, pp. 1537–1541, November 2005.

[32] M. B. Wakin, S. Sarvotham, and M. F. Duarte, "Recovery of jointly sparse signals from few random projections," in *Proceedings of the Workshop on Neural Information Proccessing Systems*, 2005.

[33] S. Chen, Y. Zheng, C. Cattani, and W. Wang, "Modeling of biological intelligence for SCM system optimization," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 769702, 10 pages, 2012.

[34] C. Cattani, S. Y. Chen, and G. Aldashev, "Information and modeling in complexity," *Mathematical Problems in Engineering*, vol. 2012, Article ID 868413, 3 pages, 2012.

[35] S. Y. Chen and X. L. Li, "Functional magnetic resonance imaging for imaging neural activity in the human brain: the annual progress," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 613465, 9 pages, 2012.

[36] C. Cattani, "On the existence of wavelet symmetries in Archaea DNA," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 673934, 21 pages, 2012.

[37] X. H. Wang, M. Li, and S. Chen, "Long memory from Sauerbrey equation: a case in coated quartz crystal microbalance in terms of ammonia," *Mathematical Problems in Engineering*, vol. 2011, Article ID 758245, 9 pages, 2011.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2012, Article ID 549102, 6 pages
doi:10.1155/2012/549102

Research Article

Bayes Clustering and Structural Support Vector Machines for Segmentation of Carotid Artery Plaques in Multicontrast MRI

Qiu Guan,[1] Bin Du,[1] Zhongzhao Teng,[2] Jonathan Gillard,[2] and Shengyong Chen[1]

[1] *College of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China*
[2] *Department of Radiology, University of Cambridge, Hills Road, Cambridge CB2 0SP, UK*

Correspondence should be addressed to Shengyong Chen, sy@ieee.org

Received 6 October 2012; Accepted 19 November 2012

Academic Editor: Carlo Cattani

Accurate segmentation of carotid artery plaque in MR images is not only a key part but also an essential step for in vivo plaque analysis. Due to the indistinct MR images, it is very difficult to implement the automatic segmentation. Two kinds of classification models, that is, Bayes clustering and SSVM, are introduced in this paper to segment the internal lumen wall of carotid artery. The comparative experimental results show the segmentation performance of SSVM is better than Bayes.

1. Introduction

Cardiovascular diseases (CVDs) are the leading cause of death globally according to the recent statistics of the World Health Organization. Atherosclerosis, a kind of systematic inflammatory disease, is estimated to be responsible for CVDs to a great extent. Therefore, there are considerable interests in characterizing atherosclerotic plaques for proper treatment planning. Research in the past 20 years indicates that plaque vulnerability is very relative to its structure, such as the lumen condition, atherosclerotic components within the plaque [1–5].

As the fundamental step, artery wall should be segmented accurately. Meanwhile, explicit detection of wall is very important to locate each component inside the plaque correctly, which is also very significant for the subsequent procedures such as component analysis.

Automated analysis of plaque composition in the carotid arteries has been presented by many researchers. Different imaging techniques always bring out distinct characteristic of image, which will restrict different applicable approach to approach of segmentation. Among current standard imaging techniques in clinical, in vivo multicontrast MRI technique has been generally validated to be used to quantify the composition of plaque effectively [6]. Most segmentation methods based on this kind of imaging technique are generally based on manual extraction of numerous contours. Automatic segmentation not only makes the combination of different multicontrast-weighted MR Image possible, but also can further make full use of the advantages of different image to improve the accurate rate of classification of component within lumen. Other impressive experiments are also carried out by taking use of model-based clustering and fuzzy clustering [7], maximum-likelihood classifier and nearest-mean classifier [8], morphology-enhanced probability maps [9], and k-means clustering [10]. Most of these methods are based on voxel-wise statistical classification, and the manual analysis cannot be completely replaced by them. An automatic method which was used to segment the carotid artery plaques in CT angiography (CTA) [11] has potential to replace the manual analysis. Firstly, the vessel lumen was segmented. Subsequently, classifier was trained to classify each pixel. However, this algorithm is needed to be improved to deal with the multicontrast-weighted MR Image. Furthermore, in order to provide a more accurate and objective ground truth, a simultaneous segmentation and registration model [12] is necessary in registration. This method is an active contour model based on simultaneous segmentation and registration which is belong to mutual-information-based registration [13]. Therefore, researches concerning segmentation of plaques are essential.

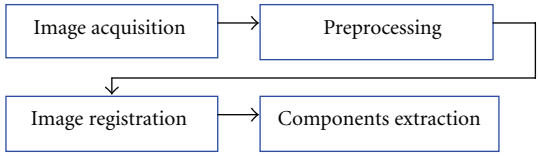


Figure 1: Flow of operations.

The paper is organized as follows. Significance of studying carotid artery plaque and current research contributions are briefly presented in Section 1. Section 2 is mainly focus on describing major and special preprocessing such as ill-illumination uniforming and image registration. Two kinds of model used to segment the wall boundary are descried in detailed in Section 3. Section 4 focuses on two algorithms to segment the lumen, and a conclusion and further work are presented in Section 5.

2. Testing Image Set

The complete process of plaque analysis system is organized as below, which is composed of four modules. Firstly, carotid artery region should be separated from the original MRI image and then move on to the preprocessing parts including noise removal and illumination uniform. After that, the lumen and the outer wall in the images are obtained in turn. The latter operations are related with extracting and modeling essential plaque components, and mechanical analysis based on FSI (fluid-structure interaction) theory will be also introduced to estimate the risk extent of a plaque. The steps in Figure 1 will be discussed in detail in this paper.

2.1. Acquisition of Testing Image Set. Images used in our research are acquired by a MRI scanner named GE SIGNA. Taking Figure 2(a) for instance, it can be found that carotid arteries marked by two rectangles are closely surrounded by other tissues as muscles, fat, bones, and other vessels in the 512 mm × 512 mm MRI image. In order to handle carotid artery alone as shown in Figure 2(b), small ROI of each artery region should be firstly segmented from the original scanning image by picking out the artery centroid which size is 81 mm × 81 mm. The reduction of interested region effectively avoids disturbing from other tissues and also improves the computing speed.

The detail of MRI acquisition has already been published in [14]. Briefly speaking, patients undergo high resolution MRI of their carotid arteries in a 1.5 Tesla MRI system (named as Signa HDx GE Healthcare, Waukesha, WI, USA) with a 4-channel phased-array neck coil (named as PACC, Machnet BV, Elde, The Netherlands). Artifact resulted from movement is minimized by using a dedicated vacuum-based head restraint system (VAC-LOK Cushion, Oncology Systems Limited, UK). It is used to fix the head and neck of patient in a comfortable position to avoid occurrence of artefact. After an initial coronal localizer sequence is sampled and tested, 2-dimensional (2D) axial time-of-fight (TOF) MR angiography is performed to identify the location of the carotid bifurcation and the region of maximum stenosis. Axial images are acquired through the common carotid artery 12 mm (4 slices) below the carotid bifurcation to a point 12 mm (4 slices) distal to the extent of the stenosis identified on the TOF sequence. This kind of method ensures that the whole region of carotid plaque is completely imaged.

To describe the characteristic of different MRI sequence, the following parameters are used: T1 weighted (repetition time/echo time: 1 × RR/7.8 ms) with fat saturation, T2 weighted (repetition time/echo time: 2 × RR/100 ms) with fat saturation, proton density weighted (repetition time/echo time: 2 × RR/7.8 ms) with fat saturation, and short-time inversion recovery (repetition time/echo time/inversion time: 2 × RR/46/150 ms). The window of view of each MR image is 10 cm × 10 cm, and size of data matrix is 512 × 512. The spatial resolution achieved of each pixel is 0.39 mm × 0.39 mm.

In Figure 2(a), two small ROIs marked by red rectangles are carotid arteries each size of RIO is 81 mm × 81 mm. Figure 2(b) is the amplified images of these two areas.

2.2. Preprocessing. Due to the inhomogeneity of coil, the intensity of each image should be adjusted to be relative uniform to obtain relative consistent gray scale for the subsequent segmentation based on clustering. The region (14 mm × 14 mm), which lies in the center of the vessel, is selected as the interesting region. The contrast of the image is increased by a linear transformation,

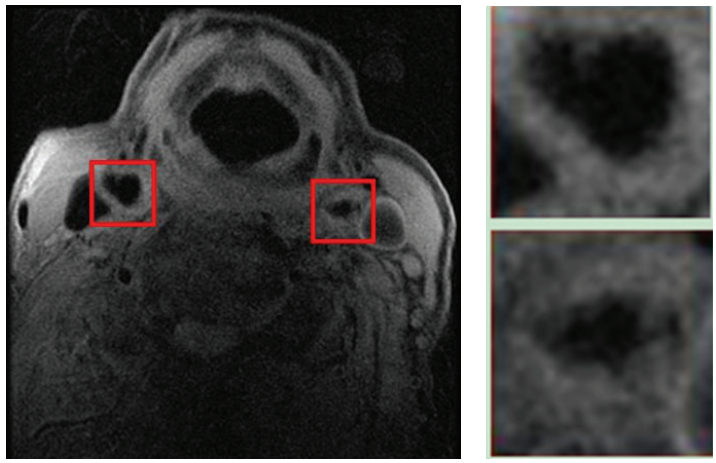
$$u_1 = \frac{u_0 - m}{M - m} \times 255, \tag{1}$$

where u_0 is the initial intensity, u_1 is adjusted intensity, and M and m are the maximum intensity and minimum intensity of the original image. The adjusted results of intensity uniform are shown in Figure 3.

2.3. Image Registration. According to the characteristics of MR image, the contour of lumen is clearly presented in the sequence of T1 which is blood suppressed for short. In Figure 4, mark two feature points in images (a) and (b) as red points. Normally, the luminal bifurcation and narrowest location are selected as marking points for registration.

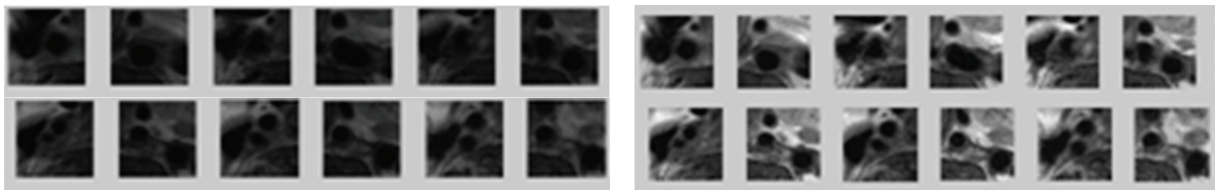
Generally speaking, the image is indistinct as shown in Figure 4. Therefore it is very difficult to mark feature points in some images. In order to deal with this problem, the registration method proposed in this paper is based on prior-constrained segmentation of carotid artery under DOG scale space. As seen from the name, the segmentation algorithm implies two parts. First, inspired by SIFT algorithm, the advantage of difference of Gaussian (DOG) scale space is introduced to catch the edges that seem ambiguous in the original image scale, which is the scale derivative of Gaussian scale space along the scale coordinate. Second, given a simple prior knowledge that the artery wall is near round, a given thickness of carotid artery wall is set to restrict the searching area. Prior shape is critical information for external wall segmentation. The steps to get the wall boundary are shown in Figure 5.

Then through minimizing the energy function using a gradient flow, we can achieve the goal of simultaneous



(a) (b)

Figure 2: ROI extraction: (a) original MRI image, (b) extracted images.



(a) (b)

Figure 3: Preprocessing of selected slices of MR images: (a) a set of original images, (b) resultant images after contrast normalization.

segmentation and registration [12]. On the one hard, this new method can reduce the influence of noise on the original images and lead to improved registration, on the other hand it also can improve the precision segmentation, especially for segmentation the blurred images.

Given two images I_1 and I_2; C_1 is the object contour of I_1, and C_2 is the object contour of I_2. Establish mapping $C_2 = g(C_1)$. The steps of simultaneous segmentation and registration method are listed as follows.

Step 1. Initialize C_1, g, and C_2.

Step 2. Optimize the registration parameters to obtain the optimal mapping function g.

Step 3. Evolute C_1 to obtain the optimum partition line of the current image I_1, and obtain the optimal split line of the current image I by $C_2 = g(C_1)$.

Step 4. Reach the maximum number of iterative steps, or before and after the two results of the iteration are less than the threshold value then the algorithm stops, ended; otherwise turn to Step 2.

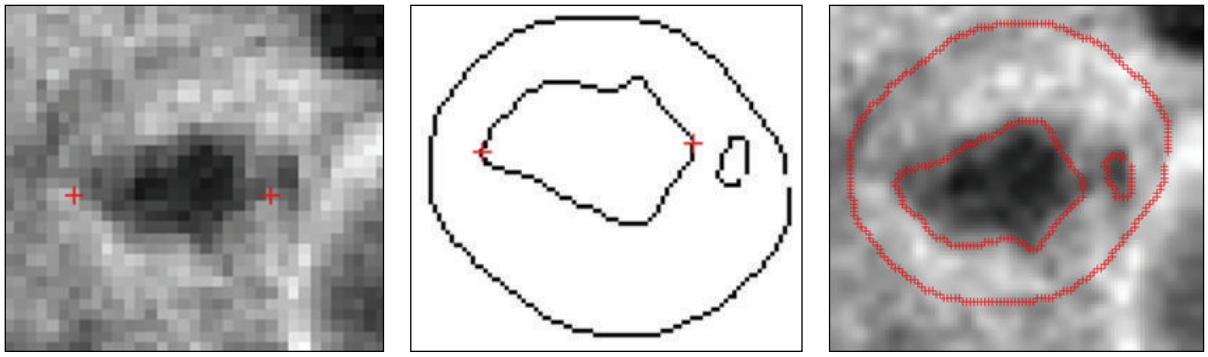
3. Modelling

To compare the results of different algorithm of modeling, two kinds of model which are based on Bayes classification algorithm and SSVM (structural support vector machines) are carried out in this paper.

3.1. Building of Training Set. From MRI slices with matching histological slices, slices 12 and 25 are selected to generate the training set for segmentation. Images of those two slices are manually segmented based on registered histological results and relative intensity. A total of 549 pixels (each pixel contains 4 densities representation with total 4 different contrast weight) are selected randomly in the investigation. From these segmentation results, each pixel is determined to belong to one of the 4 issue types including lipid (denoted as Z_1), normal issue (denoted as Z_2), calcification (denoted as Z_3), and others (including lumen or outer issue, denoted as Z_4). The training set is used to generate the probability function which is used to determine the probability of tissue type of each pixel in the model based on Bayes classification.

3.2. Model Based on Bayes Classification. The most important part of the segmentation algorithms is to determine the probabilities of each pixel. These probabilities represent the likelihood that the tissue of the pixel at the current location is lipid, calcification, normal issue, or others.

Maximum classifier is used to determine which issue type the pixel belongs to. Figure 6 gives the flow-chart of our maximum decision probability functional classifier. Where $\vec{I}$ is one pixel of multicontrast weighted MR images transformed



(a) (b) (c)

Figure 4: Handle marking points for registration: (a) MR images, (b) manual outline, (c) result of registration.

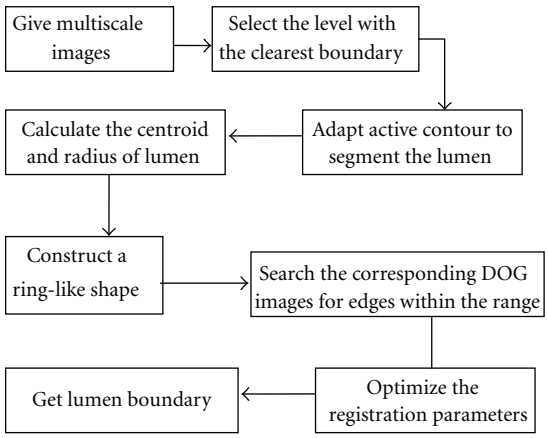


Figure 5: Flowchart of multiscale PCA.

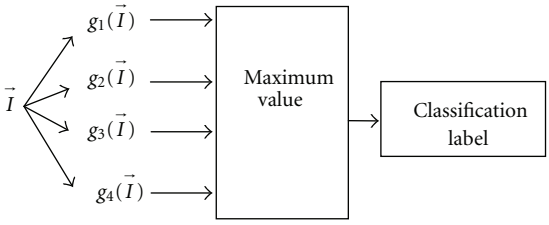


Figure 6: Flowchart of maximum decision probability functional classifier.

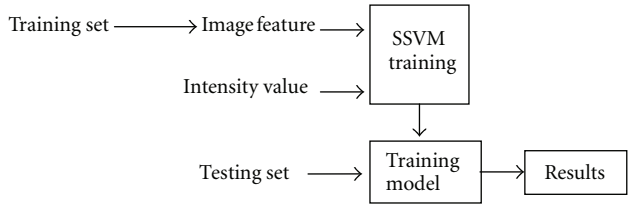


Figure 7: Flowchart of SSVM to obtain gray information.

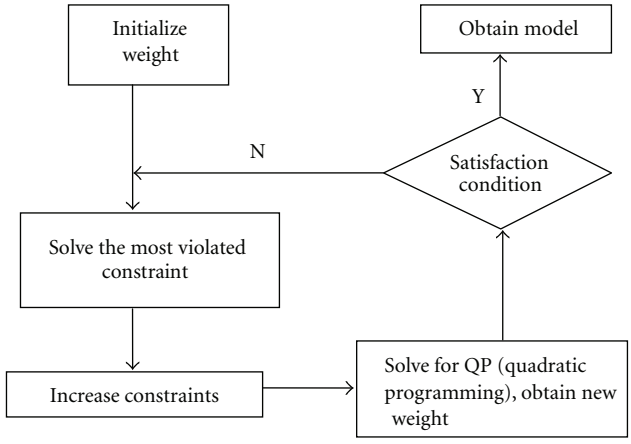


Figure 8: Flowchart of the iterative training of SSVM.

by preprocessing, $g_i(\vec{I})$ is the decision function, and $P(Z_i \mid \vec{I})$ is class-conditional probability density function (pdf). By comparing values of four functions, if $g_i(\vec{I})$ is the maximum probability value of one pixel, then pixel $\vec{I}$ belongs to Z_i and is labeled i.

3.3. Model Based on SSVM. Recently, structured prediction has already attracted much attention, and many approaches have also been developed based on it. Structured learning is one of the main approaches of structured prediction, which not only studies the problems with well-structured inputs and outputs but also reveals strong internal correlations. It is formulated as the learning of complex functional dependencies between multivariate input and output representations. Structured learning has significant impact in addressing important computer vision tasks. Figure 7 gives the flowchart of SSVM to obtain gray information. The flowchart of the iterative training of SSVM is given in Figure 8.

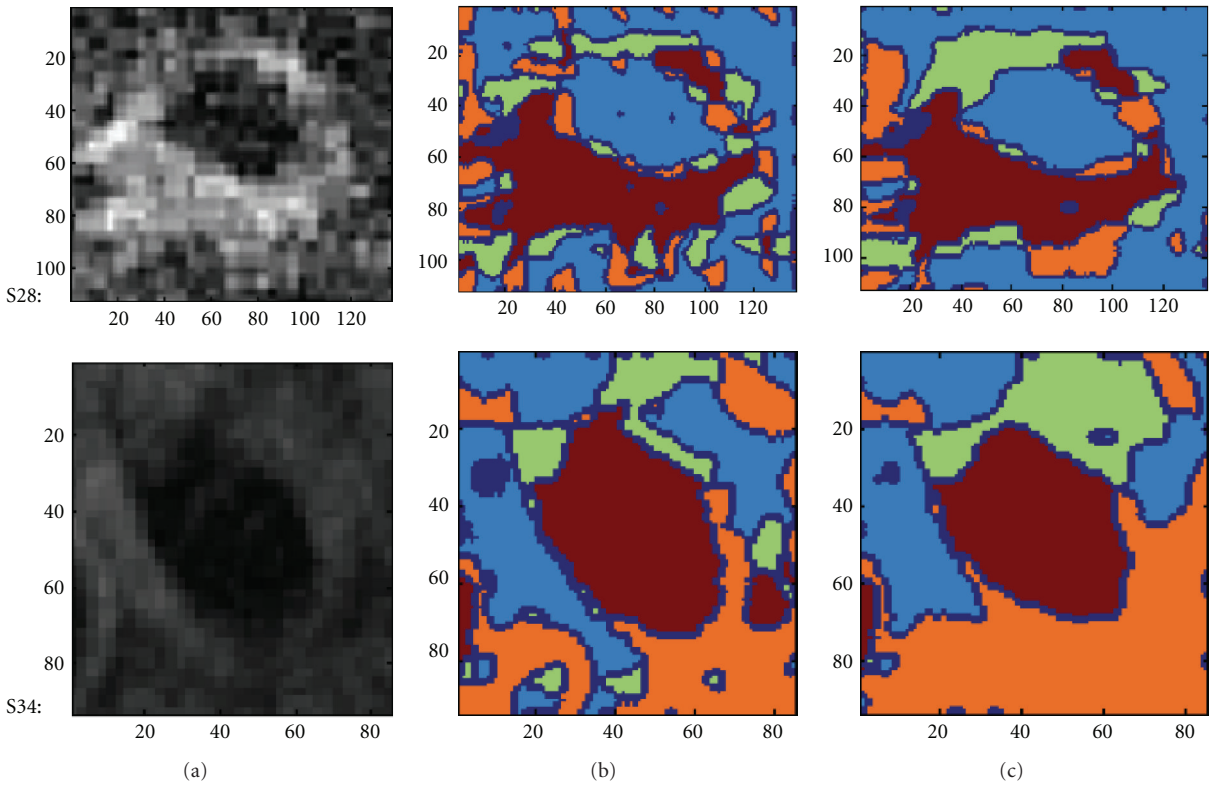


Figure 9: Two segmentation results of selected slice using multicontrast MR images: (a) testing MR images; (b) automatic segmentation results of Bayes classifier; (c) automatic segmentation results of SSVM process.

4. Comparison

The results of segmentation of slices 28 and 34 MR images based on Bayes and SSVM are illustrated in Figure 9.

As seen in Figure 9, the segmentation result in term of classification algorithm reveals that the performance of SSVM is much better than that of Bayes due to the former including structural information, and smoothing effect of segmentation of SSVM is also obvious.

The results presented by image are inadequate to make evaluations. Here a parameter named misclassification rate is defined to judge the accuracy of each algorithm.

In the experiment of this paper, a selected slice MR image is corrupted by global intensity varying from 20% to 40% and adding 1%–9% noise. Misclassification rate, an evaluating criterion, is defined as the ratio of misclassified pixels to total number of pixels of this class. It is formulated as (2) as follows:

$$e(i) = \frac{fp + fn}{n}, \tag{2}$$

where $e(i)$ is the misclassification rate of tissue i; fp is the false positive responses (pixel belongs to tissue i but is classified as other tissues); fn is the false negative responses (pixel does not belong to tissue i but is classified as tissue type i); n is the total number of pixels of tissue type i.

The misclassification rate of lumen obtained by Bayes and SSVM algorithm is listed in Table 1. From the statistics shown in Table 1, it can be seen that the misclassification rate caused by SSVM is much lower than that of Bayes. That stands for the performance of SSVM outperforms that of Bayes, especially while the level of noise is higher.

Table 1: Misclassification rate of lumen for Bayes and SSVM.

Noise	Misclassification rate	
	Bayes	SSVM
1%	3.5	2.6
3%	5.3	4.8
5%	6.5	6.3
7%	10.6	8.5
9%	16.9	9.6

5. Conclusion

To summarize, the work in this paper is focus on the first several steps of carotid artery plaque analysis, including preprocessing of MR image, model-based segmentation of lumen, plaque, and external wall. Two kinds of model, Bayes and SSVM, are separately constructed and applied to the detection of internal wall. Receivable boundaries can be both obtained by two algorithms, the results of experiment shows

the segmentation performance of SSVM is better than that of Bayes, especially, while the level of noise in image is higher.

But there are still some improvements need to be done in the future to break the limitations of the current work. Firstly, improve Bayes to better performance by increasing structural information. Secondly, introduce sequence image tracking technique in research to improve the performance of human interaction to specify the center of lumen. Further effort should focus on estimation of artery location in each MRI slice and take advantage of information gained from previous slice to pick out the artery centroid of current image. Moreover, several other algorithms need to be testified and compared with them when dealing with plaques.

Acknowledgments

The work was supported in part by the National Science Foundation of China (NSFC no. 61173096, 61103140, and 51075367), Doctoral Fund of Ministry of Education of China (20113317110001), and Zhejiang Provincial S and T Department (2010R10006, 2010C33095).

References

[1] Z. Teng, J. He, A. J. Degnan et al., "Critical mechanical conditions around neovessels in carotid atherosclerotic plaque may promote intraplaque hemorrhage," *Atherosclerosis*, vol. 223, no. 2, pp. 321–326, 2012.

[2] Z. Teng, A. J. Degnan, S. Chen, and J. H. Gillard, "Characterization of healing following atherosclerotic carotid plaque rupture in acutely symptomatic patients: an exploratory study using in vivo cardiovascular magnetic resonance," *Journal of Cardiovascular Magnetic Resonance*, vol. 13, no. 1, article 64, 2011.

[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, pp. 480–487, 2011.

[4] 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.

[5] 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.

[6] R. A. Trivedi, J. U-King-Im, M. J. Graves et al., "Multi-sequence in vivo MRI can quantify fibrous cap and lipid core components in human carotid atherosclerotic plaques," *European Journal of Vascular and Endovascular Surgery*, vol. 28, no. 2, pp. 207–213, 2004.

[7] I. M. Adame, R. J. van der Geest, B. A. Wasserman, M. A. Mohamed, J. H. C. Reiber, and B. P. F. Lelieveldt, "Automatic segmentation and plaque characterization in atherosclerotic carotid artery MR images," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 16, no. 5, pp. 227–234, 2004.

[8] S. E. Clarke, V. Beletsky, R. R. Hammond, R. A. Hegele, and B. K. Rutt, "Validation of automatically classifiedmagnetic resonance images for carotid plaque compositional analysis," *Stroke*, vol. 37, no. 1, pp. 93–97, 2006.

[9] F. Liu, D. Xu, M. S. Ferguson et al., "Automated in vivo segmentation of carotid plaque MRI with morphology-enhanced probability maps," *Magnetic Resonance in Medicine*, vol. 55, no. 3, pp. 659–668, 2006.

[10] C. Karmonik, P. Basto, K. Vickers et al., "Quantitative segmentation of principal carotid atherosclerotic lesion components by feature space analysis based on multicontrast MRI at 1.5 T," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 352–360, 2009.

[11] D. Vukadinovic, S. Rozie, M. van Gils et al., "Automated versus manual segmentation of atherosclerotic carotid plaque volume and components in CTA: associations with cardiovascular risk factors," *International Journal of Cardiovascular Imaging*, vol. 28, no. 4, pp. 877–887, 2012.

[12] Y. Chen, S. Thiruvenkadam, F. Huang, K. S. Gopinath, and R. W. Brigg, "Simultaneous segmentation and registration for functional MR images," in *Proceedings of the 16th International Conference on Pattern Recognition*, vol. 1, pp. 747–750, Québec, Canada, 2006.

[13] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[14] U. Sadat, R. A. Weerakkody, D. J. Bowden et al., "Utility of high resolution MR imaging to assess carotid plaque morphology: a comparison of acute symptomatic, recently symptomatic and asymptomatic patients with carotid artery disease," *Atherosclerosis*, vol. 207, no. 2, pp. 434–439, 2009.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2012, Article ID 291510, 5 pages
doi:10.1155/2012/291510

Research Article

Heavy-Tailed Prediction Error: A Difficulty in Predicting Biomedical Signals of $1/f$ Noise Type

Ming Li,[1] Wei Zhao,[2] and Biao Chen[2]

[1] *School of Information Science & Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China*
[2] *Department of Computer and Information Science, University of Macau, Padre Tomas Pereira Avenue, Taipa, Macau*

Correspondence should be addressed to Ming Li, ming_lihk@yahoo.com

Received 31 October 2012; Accepted 20 November 2012

Academic Editor: Carlo Cattani

A fractal signal $x(t)$ in biomedical engineering may be characterized by $1/f$ noise, that is, the power spectrum density (PSD) divergences at $f = 0$. According the Taqqu's law, $1/f$ noise has the properties of long-range dependence and heavy-tailed probability density function (PDF). The contribution of this paper is to exhibit that the prediction error of a biomedical signal of $1/f$ noise type is long-range dependent (LRD). Thus, it is heavy-tailed and of $1/f$ noise. Consequently, the variance of the prediction error is usually large or may not exist, making predicting biomedical signals of $1/f$ noise type difficult.

1. Introduction

Signals of $1/f$ noise type are widely observed in biomedical engineering, ranging from heart rate to DNA and protein, see, for example, [1–37], just to cite a few. Predicting such a type of signals is desired in the field [38–43]. A fundamental issue in this regard is whether a biomedical signal of $1/f$ noise type to be predicted is predicable or not.

The predictability of signals of non-$1/f$ noise type is well studied [44–48]. However, the predictability of $1/f$ noise is rarely reported, to our best knowledge. Since many phenomena in biomedical engineering are characterized by $1/f$ noise [1–37], the predictability issue of $1/f$ noise is worth investigating.

Note that minimizing the mean square error (MSE) of prediction is a commonly used criterion in both theory and practical techniques of prediction, see, for example, [49–68]. Therefore, a sufficient condition for a biomedical signal $x(t)$ to be predictable is that the variance of its predication error exists. If the variance of the predication error does not exist, on the contrary, it may be difficult to be predicted if not unpredictable. In the case of a signal being bandlimited, the variance of its predication error is generally finite. Consequently, it may be minimized and it is predictable. However, that is not always the case for biomedical signals of $1/f$ noise type.

Let $x(t)$ be a biomedical signal in the class of $1/f$ noise. Then, its PDF is heavy-tailed, and it is LRD, see, for example, Adler et al. [69], Samorodnitsky and Taqqu [70], Mandelbrot [71], Li and Zhao [72]. Due to that, here and below, the terms, $1/f$ noise, LRD random function, and heavy-tailed random function are interchangeable.

Let $p(x)$ be the PDF of a biomedical signal $x(t)$ of $1/f$ noise type. Then, its variance is expressed by

$$\mathrm{Var}[x(t)] = \int_{-\infty}^{\infty} (x - \mu_x)^2 p(x)dx, \tag{1}$$

where μ_x is the mean of $x(t)$ if it exists. The term of heavy tail in statistics implies that $\mathrm{Var}[x(t)]$ is large. Theoretically speaking, in general, we cannot assume that $\mathrm{Var}[x(t)]$ always exists [72]. In some cases, such as the Pareto distribution, the Cauchy distribution, α-stable distributions [72], $\mathrm{Var}[x(t)]$ may be infinite. That $\mathrm{Var}[x(t)]$ does not exist is particularly true for signals in biomedical engineering and physiology, see Bassingthwaighte et al. [33] for the interpretation of this point of view.

Recall that a prediction error is a random function as we shall soon mention below. Therefore, whether the prediction error is of $1/f$ noise, or equivalently, heavy-tailed, turns to be a crucial issue we need studying. We aim at, in this research, exhibiting that prediction error of $1/f$ noise is heavy-tailed

and accordingly is of $1/f$ noise. Thus, generally speaking, the variance of a prediction error of a biomedical signal $x(t)$ of $1/f$ noise type may not exist or large. That is a reason why predicting biomedical signals of $1/f$ noise type is difficult.

The rest of this paper is organized as follows. Heavy-tailed prediction errors occurring in the prediction of biomedical signals of $1/f$ noise type are explained in Section 2. Discussions are in Section 3, which is followed by con-clusions.

2. Prediction Errors of $1/f$ Noise Type

We use $x(n)$ to represent a biomedical signal in the discrete case for $n \in \mathbf{N}$, where $\mathbf{N}$ is the set of natural numbers. Let $x_N(n)$ be a given sample of $x(n)$ for $n = 0, 1, \ldots, N-1$. Denote by $x_M(m)$ the predicted values of $x(n)$ for $m = N, N+1, N+M-1$. Then, the prediction error denoted by $e(m)$ is given by

$$e(m) = \sum_{m=N}^{N+M-1} x(m) - x_M(m). \tag{2}$$

If one uses the given sample of $x(n)$ for $n = N, N+1, \ldots, 2N-1$ to obtain the predictions denoted by $x_M(m)$ for $m = 2N, 2N+1, 2N+M-1$, the error is usually different from (2), which implies that the error $e(m)$ is a random variable. Denote by $p(e)$ the PDF of $e(m)$. Then, its variance is expressed by

$$\mathrm{Var}[e(m)] = \sum_{m=N}^{N+M-1} (e-\mu_e)^2 p(e), \tag{3}$$

where μ_e is the mean of $e(m)$.

Let P be the operator of a predictor. Then,

$$x_M(m) = P x_N(n). \tag{4}$$

A natural requirement in terms of P is that $\mathrm{Var}[e(m)]$ should be minimized. Thus, the premise that $\mathrm{Var}[e(m)]$ can be minimized is that it exists.

It is obviously seen that $\mathrm{Var}[e(m)]$ may be large if $p(e)$ is heavy tailed. In a certain cases, $\mathrm{Var}[e(m)]$ may not exist. To explain the latter, we assume that $e(m)$ follows a type of heavy-tailed distribution called the Pareto distribution.

Denote by $p_{\mathrm{Pareto}}(e)$ the PDF of the Pareto distribution. Then [73], it is in the form

$$p_{\mathrm{Pareto}}(e) = \frac{ab^a}{e^{a+1}}, \tag{5}$$

where $e \geq b$, $a > 0$, and $b > 0$. The mean and variance of $e(m)$ are, respectively, expressed by

$$\begin{aligned} \mu_e &= \frac{ab}{a-1}, \\ \mathrm{Var}(e) &= \frac{ab^2}{(a-1)^2(a-2)}. \end{aligned} \tag{6}$$

The above exhibits that $\mathrm{Var}[e(m)]$ does not exist if $a = 1$ or $a = 2$ and if $e(m)$ follows the Pareto distribution.

Note that the situation that $\mathrm{Var}[e(m)]$ does not exist may not occur if $e(m)$ is light-tailed. Therefore, the question in this regard is whether $e(m)$ is heavy-tailed if a biomedical signal $x(n)$ is of $1/f$ noise. The answer to that question is affirmative. We explain it below.

Theorem 1. *Let $x(n)$ be a biomedical signal of $1/f$ noise type to be predicted. Then, its prediction error is heavy-tailed. Consequently, it is of $1/f$ noise.*

Proof. Let $r_{xx}(k)$ be the autocorrelation function (ACF) of $x(n)$. Then,

$$r_{xx}(k) = E[x(n)x(n+k)], \tag{7}$$

where k is lag and E the mean operator. Let $r_{MM}(k)$ be the ACF of $x_M(m)$. Then,

$$r_{MM}(k) = E[x_M(m)x_M(m+k)]. \tag{8}$$

Let $r_{ee}(k)$ be the ACF of $e(m)$. Then,

$$r_{ee}(k) = E[e(m)e(m+k)]. \tag{9}$$

Note that

$$\begin{aligned} r_{ee}(k) &= E[e(m)e(m+k)] \\ &= E\{[x(m) - x_M(m)][x(m+k) - x_M(m+k)]\} \\ &= E[x(m)x(m+k) + x_M(m)x_M(m+k) \\ &\qquad - x_M(m)x(m+k) - x(m)x_M(m+k)] \\ &= r_{xx}(k) + r_{MM}(k) - r_{Mx}(k) - r_{xM}(k). \end{aligned} \tag{10}$$

In the above expression, $r_{Mx}(k)$ is the cross-correlation between $x_M(m)$ and $x(m)$. On the other side, $r_{xM}(k)$ is the cross-correlation between $x(m)$ and $x_M(m)$. Since $r_{Mx}(k) = r_{xM}(k)$, we have

$$r_{ee}(k) = r_{xx}(k) + r_{MM}(k) - 2r_{xM}(k). \tag{11}$$

Recall that $x(m)$ is $1/f$ noise. Thus, it is heavy-tailed and hence LRD. Consequently, for a constant $c_1 > 0$, we have

$$r_{xx}(k) \sim c_1 k^{-\alpha} \quad (k \longrightarrow \infty) \text{ for } 0 < \alpha < 1. \tag{12}$$

On the other hand, the predicted series $x_M(m)$ is LRD. Thus, for a constant $c_2 > 0$, the following holds:

$$r_{MM}(k) \sim c_2 k^{-\beta} \quad (k \longrightarrow \infty) \text{ for } 0 < \beta < 1. \tag{13}$$

In (11), if $r_{xM}(k)$ is summable, that is, it decays faster than $r_x(k)$ or $r_M(k)$, it may be ignored for $k \to \infty$. In this case, $r_{ee}(k)$ is still non-summable. In fact, one has

$$r_{ee}(k) \sim \begin{cases} c_1 k^{-\alpha}, & 0 < \alpha < \beta < 1, \\ c_2 k^{-\beta}, & 0 < \beta < \alpha < 1, \\ (c_1 + c_2) k^{-\beta}, & \alpha = \beta. \end{cases} \quad (k \longrightarrow \infty), \tag{14}$$

On the other side, when $r_{xM}(k)$ is non-summable, $r_e(k)$ is non-summable too. In any case, we may write $r_{ee}(k)$ by

$$r_{ee}(k) \sim ck^{-\gamma} \quad (k \longrightarrow \infty) \text{ for } 0 < \gamma < 1. \tag{15}$$

Therefore, the prediction error $e(m)$ is LRD. Its PDF $p(e)$ is heavy-tailed according to the Taqqu's law. Following [72], therefore, $e(m)$ is a $1/f$ noise. This completes the proof. □

3. Discussions

The present result implies that cautions are needed for dealing with predication errors of biomedical signals of $1/f$ noise type. In fact, if specific biomedical signals are in the class of $1/f$ noise, the variances of their prediction errors may not exist or large [72]. Tucker and Garway-Heath used to state that their prediction errors with either prediction model they used are large [74]. The result in this paper may in a way provide their research with an explanation.

Due to the fact that a biomedical signal may be of $1/f$ noise, PDF estimation is suggested as a preparatory stage for prediction. As a matter of fact, if a PDF estimation of biomedical signal is light-tailed, its variance of prediction error exists. On the contrary, the variance of the prediction error may not exist. In the latter case, special techniques have to be considered [75–78]. For instance, weighting prediction error may be a technique necessarily to be taken into account, which is suggested in the domain of generalized functions over the Schwartz distributions [79].

4. Conclusions

We have explained that the prediction error $e(m)$ in predicting biomedical signals of $1/f$ noise type is usually LRD. This implies that its PDF $p(e)$ is heavy-tailed and $1/f$ noise. Consequently, $\mathrm{Var}[e(m)]$ may in general be large. In some cases [72], $\mathrm{Var}[e(m)]$ may not exist, making the prediction of biomedical signals of $1/f$ noise type difficult with the way of minimizing $\mathrm{Var}[e(m)]$.

Acknowledgments

This work was supported in part by the 973 plan under the Project Grant no. 2011CB302800 and by the National Natural Science Foundation of China under the Project Grant no. 61272402, 61070214, and 60873264.

References

[1] N. Aoyagi, Z. R. Struzik, K. Kiyono, and Y. Yamamoto, "Autonomic imbalance induced breakdown of long-range dependence in healthy heart rate," *Methods of Information in Medicine*, vol. 46, no. 2, pp. 174–178, 2007.

[2] S. Tong, D. Jiang, Z. Wang, Y. Zhu, R. G. Geocadin, and N. V. Thakor, "Long range correlations in the heart rate variability following the injury of cardiac arrest," *Physica A*, vol. 380, no. 1-2, pp. 250–258, 2007.

[3] N. V. Sarlis, E. S. Skordas, and P. A. Varotsos, "Heart rate variability in natural time and $1/f$ 'noise'," *Europhysics Letters*, vol. 87, no. 1, Article ID 18003, 2009.

[4] Z. R. Struzik, J. Hayano, R. Soma, S. Kwak, and Y. Yamamoto, "Aging of complex heart rate dynamics," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, pp. 89–94, 2006.

[5] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and Biological Engineering and Computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[6] J. H. T. Bates, G. N. Maksym, D. Navajas, and B. Suki, "Lung tissue rheology and $1/f$ noise," *Annals of Biomedical Engineering*, vol. 22, no. 6, pp. 674–681, 1994.

[7] J. M. Halley and W. E. Kunin, "Extinction risk and the $1/f$ family of noise models," *Theoretical Population Biology*, vol. 56, no. 3, pp. 215–230, 1999.

[8] M. C. Wichmann, K. Johst, M. Schwager, B. Blasius, and F. Jeltsch, "Extinction risk, coloured noise and the scaling of variance," *Theoretical Population Biology*, vol. 68, no. 1, pp. 29–40, 2005.

[9] Z. Yang, L. Hoang, Q. Zhao, E. Keefer, and W. Liu, "$1/f$ neural noise reduction and spike feature extraction using a subset of informative samples," *Annals of Biomedical Engineering*, vol. 39, no. 4, pp. 1264–1277, 2011.

[10] J. Ruseckas and B. Kaulakys, "Tsallis distributions and $1/f$ noise from nonlinear stochastic differential equations," *Physical Review E*, vol. 84, no. 5, Article ID 051125, 7 pages, 2011.

[11] F. Beckers, B. Verheyden, and A. E. Aubert, "Aging and nonlinear heart rate control in a healthy population," *American Journal of Physiology*, vol. 290, no. 6, pp. H2560–H2570, 2006.

[12] B. Pilgram and D. T. Kaplan, "Nonstationarity and $1/f$ noise characteristics in heart rate," *American Journal of Physiology*, vol. 276, no. 1, pp. R1–R9, 1999.

[13] P. Szendro, G. Vincze, and A. Szasz, "Pink-noise behaviour of biosystems," *European Biophysics Journal*, vol. 30, no. 3, pp. 227–231, 2001.

[14] G. Massiera, K. M. Van Citters, P. L. Biancaniello, and J. C. Crocker, "Mechanics of single cells: rheology, time dependence, and fluctuations," *Biophysical Journal*, vol. 93, no. 10, pp. 3703–3713, 2007.

[15] Y. Murase, T. Shimada, N. Ito, and P. A. Rikvold, "Effects of demographic stochasticity on biological community assembly on evolutionary time scales," *Physical Review E*, vol. 81, no. 4, Article ID 041908, 14 pages, 2010.

[16] T. Yokogawa and T. Harada, "Generality of a power-law long-term correlation in beat timings of single cardiac cells," *Biochemical and Biophysical Research Communications*, vol. 387, no. 1, pp. 19–24, 2009.

[17] T. Harada, T. Yokogawa, T. Miyaguchi, and H. Kori, "Singular behavior of slow dynamics of single excitable cells," *Biophysical Journal*, vol. 96, no. 1, pp. 255–267, 2009.

[18] A. Eke, P. Hermán, J. B. Bassingthwaighte et al., "Physiological time series: distinguishing fractal noises from motions," *Pflugers Archiv*, vol. 439, no. 4, pp. 403–415, 2000.

[19] B. J. West, "Fractal physiology and the fractional calculus: a perspective," *Frontiers in Fractal Physiology*, vol. 1, article 12, 2010.

[20] P. Grigolini, G. Aquino, M. Bologna, M. Luković, and B. J. West, "A theory of $1/f$ noise in human cognition," *Physica A*, vol. 388, no. 19, pp. 4192–4204, 2009.

[21] F. Grüneis, M. Nakao, Y. Mizutani, M. Yamamoto, M. Meesmann, and T. Musha, "Further study on $1/f$ fluctuations observed in central single neurons during REM sleep," *Biological Cybernetics*, vol. 68, no. 3, pp. 193–198, 1993.

[22] H. Sheng, Y.-Q. Chen, and T.-S. Qiu, "Heavy-tailed distribution and local long memory in time series of molecular motion

on the cell membrane," *Fluctuation and Noise Letters*, vol. 10, no. 1, pp. 93–119, 2011.

[23] B. J. West and W. Deering, "Fractal physiology for physicists: Lévy statistics," *Physics Report*, vol. 246, no. 1-2, pp. 1–100, 1994.

[24] W. Deering and B. J. West, "Fractal physiology," *IEEE Engineering in Medicine and Biology Magazine*, vol. 11, no. 2, pp. 40–46, 1992.

[25] B. J. West, "Physiology in fractal dimensions: error tolerance," *Annals of Biomedical Engineering*, vol. 18, no. 2, pp. 135–149, 1990.

[26] M. Joyeux, S. Buyukdagli, and M. Sanrey, "$1/f$ Fluctuations of DNA temperature at thermal denaturation," *Physical Review E*, vol. 75, no. 6, Article ID 061914, 9 pages, 2007.

[27] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.

[28] C. Cattani, E. Laserra, and I. Bochicchio, "Simplicial approach to fractal structures," *Mathematical Problems in Engineering*, vol. 2012, Article ID 958101, 21 pages, 2012.

[29] P. Herman and A. Eke, "Nonlinear analysis of blood cell flux fluctuations in the rat brain cortex during stepwise hypotension challenge," *Journal of Cerebral Blood Flow & Metabolism*, vol. 26, no. 9, pp. 1189–1197, 2006.

[30] M. Baumert, V. Baier, and A. Voss, "Long-term correlations and fractal dimension of beat-to-beat blood pressure dynamics," *Fluctuation and Noise Letters*, vol. 5, no. 4, pp. L549–L555, 2005.

[31] C. Cattani, "On the existence of wavelet symmetries in Archaea DNA," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 673934, 21 pages, 2012.

[32] S. Y. Ponomarev, V. Putkaradze, and T. C. Bishop, "Relaxation dynamics of nucleosomal DNA," *Physical Chemistry Chemical Physics*, vol. 11, no. 45, pp. 10633–10643, 2009.

[33] J. B. Bassingthwaighte, L. S. Liebovitch, and B. J. West, *Fractal Physiology*, Oxford University Press, 1994.

[34] D. Craciun, A. Isvoran, and N. M. Avram, "Long range correlation of hydrophilicity and flexibility along the calcium binding protein chains," *Physica A*, vol. 388, no. 21, pp. 4609–4618, 2009.

[35] J. Siódmiak, J. J. Uher, I. Santamaría-Holek, N. Kruszewska, and A. Gadomski, "On the protein crystal formation as an interface-controlled process with prototype ion-channeling effect," *Journal of Biological Physics*, vol. 33, no. 4, pp. 313–329, 2007.

[36] S. C. Kou and X. S. Xie, "Generalized langevin equation with fractional gaussian noise: subdiffusion within a single protein molecule," *Physical Review Letters*, vol. 93, no. 18, Article ID 180603, 4 pages, 2004.

[37] H. Sheng, Y.-Q. Chen, and T.-S. Qiu, *Fractional Processes and Fractional Order Signal Processing*, Springer, 2012.

[38] M. Panella, "Advances in biological time series prediction by neural networks," *Biomedical Signal Processing and Control*, vol. 6, no. 2, pp. 112–120, 2011.

[39] Y.-R. Cho and A. Zhang, "Predicting protein function by frequent functional association pattern mining in protein interaction networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 30–36, 2010.

[40] A. Castro, M. A. L. Marques, D. Varsano, F. Sottile, and A. Rubio, "The challenge of predicting optical properties of biomolecules: what can we learn from time-dependent density-functional theory?" *Comptes Rendus Physique*, vol. 10, no. 6, pp. 469–490, 2009.

[41] Q. Lü, H. J. Wu, J. Z. Wu et al., "A parallel ant colonies approach to de novo prediction of protein backbone in CASP8/9," *Science China Information Sciences*. In press.

[42] B. R. Yang, W. Qu, L. J. Wang, and Y. Zhou, "A new intelligent prediction system model-the compound pyramid model," *Science China Information Sciences*, vol. 55, no. 3, pp. 723–736, 2012.

[43] J. L. Suo, X. Y. Ji, and Q. H. Dai, "An overview of computational photography," *Science China Information Sciences*, vol. 55, no. 6, pp. 1229–1248, 2012.

[44] A. Papoulis, "A note on the predictability of band-limited processes," *Proceedings of the IEEE*, vol. 73, no. 8, pp. 1332–1333, 1985.

[45] S. Y. Chen, C. Y. Yao, G. Xiao, Y. S. Ying, and W. L. Wang, "Fault detection and prediction of clocks and timers based on computer audition and probabilistic neural networks," in *Proceedings of the 8th International Workshop on Artificial Neural Networks, IWANN 2005: Computational Intelligence and Bioinspired Systems*, vol. 3512 of *Lecture Notes in Computer Science*, pp. 952–959, June 2005.

[46] R. J. Lyman, W. W. Edmonson, S. McCullough, and M. Rao, "The predictability of continuous-time, bandlimited processes," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 311–316, 2000.

[47] R. J. Lyman and W. W. Edmonson, "Linear prediction of bandlimited processes with flat spectral densities," *IEEE Transactions on Signal Processing*, vol. 49, no. 7, pp. 1564–1569, 2001.

[48] N. Dokuchaev, "The predictability of band-limited, high-frequency and mixed processes in the presence of ideal low-pass filters," *Journal of Physics A*, vol. 41, no. 38, Article ID 382002, 7 pages, 2008.

[49] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, John Wiley & Sons, 1964.

[50] A. N. Kolmogorov, "Interpolation and extrapolation of stationary random sequences," *Izvestiya Akademii Nauk SSSR*, vol. 5, pp. 3–14, 1941.

[51] L. A. Zadeh and J. R. Ragazzini, "An extension of Wiener's theory of prediction," *Journal of Applied Physics*, vol. 21, no. 7, pp. 645–655, 1950.

[52] R. J. Bhansali, "Asymptotic properties of the Wiener-Kolmogorov predictor. I," *Journal of the Royal Statistical Society B*, vol. 36, no. 1, pp. 61–73, 1974.

[53] N. Levinson, "A heuristic exposition of Wiener's mathematical theory of prediction and filtering," *Journal of Mathematical Physics*, vol. 26, pp. 110–119, 1947.

[54] N. Levinson, "The Wiener RMS (root mean squares) error criterion in filter design and prediction," *Journal of Mathematical Physics*, vol. 25, pp. 261–278, 1947.

[55] R. J. Bhansali, "Asymptotic mean-square error of predicting more than one-step ahead using the regression method," *Journal of the Royal Statistical Society C*, vol. 23, no. 1, pp. 35–42, 1974.

[56] J. Makhoul, "Linear prediction: a tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[57] D. L. Zimmerman and N. Cressie, "Mean squared prediction error in the spatial linear model with estimated covariance parameters," *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 1, pp. 27–43, 1992.

[58] D. Huang, "Levinson-type recursive algorithms for least-squares autoregression," *Journal of Time Series Analysis*, vol. 11, no. 4, pp. 295–315, 2008.

[59] R. S. Deo, "Improved forecasting of autoregressive series by weighted least squares approximate REML estimation,"

International Journal of Forecasting, vol. 28, no. 1, pp. 39–43, 2012.

[60] A. Rodríguez and E. Ruiz, "Bootstrap prediction mean squared errors of unobserved states based on the Kalman filter with estimated parameters," *Computational Statistics & Data Analysis*, vol. 56, no. 1, pp. 62–74, 2012.

[61] M. Abt, "Estimating the prediction mean squared error in gaussian stochastic processes with exponential correlation structure," *Scandinavian Journal of Statistics*, vol. 26, no. 4, pp. 563–578, 1999.

[62] R. Kohn and C. F. Ansley, "Prediction mean squared error for state space models with estimated parameters," *Biometrika*, vol. 73, no. 2, pp. 467–473, 1986.

[63] R. T. Baillie, "Asymptotic prediction mean squared error for vector autoregressive models," *Biometrika*, vol. 66, no. 3, pp. 675–678, 1979.

[64] P. Neelamegam, A. Jamaludeen, and A. Rajendran, "Prediction of calcium concentration in human blood serum using an artificial neural network," *Measurement*, vol. 44, no. 2, pp. 312–319, 2011.

[65] E. S. G. Carotti, J. C. De Martin, R. Merletti, and D. Farina, "Compression of multidimensional biomedical signals with spatial and temporal codebook-excited linear prediction," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2604–2610, 2009.

[66] W. Bachta, P. Renaud, L. Cuvillon, E. Laroche, A. Forgione, and J. Gangloff, "Motion prediction for computer-assisted beating heart surgery," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2551–2563, 2009.

[67] H.-H. Lin, C. L. Beck, and M. J. Bloom, "On the use of multivariable piecewise-linear models for predicting human response to anesthesia," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 11, pp. 1876–1887, 2004.

[68] B. S. Atal, "The history of linear prediction," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 154–161, 2006.

[69] 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.

[70] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman & Hall, New York, NY, USA, 1994.

[71] B. B. Mandelbrot, *Multifractals and 1/f Noise*, Springer, 1998.

[72] M. Li and W. Zhao, "On $1/f$ noise," *Mathematical Problems in Engineering*. In press.

[73] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, McGraw-Hill, 1961.

[74] A. Tucker and D. Garway-Heath, "The pseudotemporal bootstrap for predicting glaucoma from cross-sectional visual field data," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 79–85, 2010.

[75] M. Carlini and S. Castellucci, "Modelling the vertical heat exchanger in thermal basin," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '11)*, vol. 6785 of *Lecture Notes in Computer Science*, pp. 277–286, Springer.

[76] M. Carlini, C. Cattani, and A. Tucci, "Optical modelling of square solar concentrator," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '11)*, vol. 6785 of *Lecture Notes in Computer Science*, pp. 287–295, Springer.

[77] R. J. Bhansali and P. S. Kokoszka, "Prediction of long-memory time series: a tutorial review," *Lecture Notes in Physics*, vol. 621, pp. 3–21, 2003.

[78] L. Bisaglia and S. Bordignon, "Mean square prediction error for long-memory processes," *Statistical Papers*, vol. 43, no. 2, pp. 161–175, 2002.

[79] 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.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2012, Article ID 816481, 6 pages
doi:10.1155/2012/816481

Research Article

In Vitro Evaluation of Ferrule Effect and Depth of Post Insertion on Fracture Resistance of Fiber Posts

R. Schiavetti and G. Sannino

Department of Oral Health, University of Rome Tor Vergata, Viale Oxford, 00100 Rome, Italy

Correspondence should be addressed to G. Sannino, gianpaolo.sannino@uniroma2.it

Received 10 October 2012; Accepted 5 November 2012

Academic Editor: Carlo Cattani

Purpose. The analysis of the complex model of fiber post and ferrule is given and studied in this paper. A novel approach and a solution to the evaluation of stress of post and core system within the ferrule effect are proposed. *Methods*. Sixty freshly extracted premolars were selected for the study. The following experimental groups were therefore defined ($n = 10$): (1) 5 mm, (2) 7 mm, (3) 9 mm, (4) ferrule-5 mm, (5) ferrule-7 mm, and (6) ferrule-9 mm. Preshaping drills (C) were used to prepare the root canals at 5, 7, and 9 mm in depth. In specimens of groups 3–6 a circumferential collar of tooth structure of 2 mm in height. Fluorocore 2 core build-up material (I) was used for fiber post luting. With the same material, a buildup of 2 mm in height was created. A controlled compressive load (crosshead speed: 0.75 mm/min) was applied by means of a stainless steel stylus (Ø 1 mm) at the coronal end of the post extruding out of the root. *Results*. In all the tests the level of significance was set at $P < 0.05$. Significantly higher fracture strengths were measured in the presence of a ferrule effect. In groups 1, 2, and 3 (ferrule group), the mean fracture values were, respectively, 163,8 N, 270,9 N, and 254,7 N. These data are higher and statistically significantly different when compared with the three groups 4, 5, and 6 (no-ferrule group), in which the values obtained were, respectively, 40,5 N, 41,7 N, and 44,9 N. *Conclusion*. The ferrule effect in the endodontically treated teeth positively affects the fracture strength of the fiber post. Conversely, post depth insertion did not affect the resistance to fracture.

1. Introduction

A persistent problem in clinical dentistry is represented by the risk fracture of endodontically treated teeth [1]. These teeth are considered to be less resistance, because of the loss of tooth structure during conservative access cavity preparation. The influence of subsequent canal instrumentation and obturation leads to a reduction in the resistance to fracture [2, 3]. To restore these teeth, posts are often required in order to provide anchorage for the core-forming material and coronoradicular stabilization [4, 5]. Cast posts and cores have been used for this purpose for many years, while more recently fiber posts showed to represent a valid alternative. The clinical success of fiber post restorations is mainly related to their biomechanical properties that, being close to those of dentin, reduce stress transmission to the roots [6–9]. The potential of fiber posts to reduce the incidence of nonretrievable root fractures in comparison with cast posts was confirmed in several studies [10–12]. Among the several parameters influencing the success of a post-based rehabilitation, preservation of coronal dental tissue and, particularly, the presence of a ferrule effect have been advocated as favorable conditions to decrease stress transmission to the root [13]. Sorensen and Engelman [14] described the ferrule as the coronal-dentinal extension of the tooth structure occlusal to the shoulder preparation. The ferrule effect in association with cast post and cores has been studied by many investigators [15–17]. Conversely, little information is available if the ferrule is of additional value in providing reinforcement in teeth restored with prefabricated post and composite cores, and the advantages coming from the presence of ferrule in prefabricated post and core are questioned by Al-Hazaimeh and Gutteridge [18].

The main task of this in vitro study is to evaluate the effect of ferrule preparation on fracture resistance of fiber

post, as a function of the presence/absence of a ferrule and as a function of the depth of insertion of the fiber posts.

The formulated null hypothesis was that neither depth of post insertion nor the presence of a 2 mm high ferrule had a significant influence on fracture resistance of a fiber post-retained restoration.

2. Material and Methods

Sixty freshly extracted premolars were selected for the study. Teeth had to be free of cracks, caries, and fractures and were stored at room temperature in saline solution before testing. The anatomic crowns of all teeth were sectioned perpendicularly to the tooth long axis at the cement-enamel junction (CEJ). Roots were endodontically treated using the "step-back" technique [19] to a number 70 size file (A) (see Table 2) and irrigated with 2.5% sodium hypochlorite.

Each canal was obturated using the lateral condensation technique with gutta-percha points (B) and the resin sealer AH Plus Jet (C) (see Table 2). The endodontic access cavities were temporarily filled with a glass ionomer cement (D) (Fuji II, GC corp, Tokyo, Japan). After 24 hours, the coronal seal was removed by means of 240-grit abrasive SiC papers under water cooling. Roots were randomly divided into six experimental groups that differed for the depth of the prepared post space and for the presence or absence of a ferrule effect. The following experimental groups were therefore defined ($n = 10$): (1) 5 mm (Figure 1(a)); (2) 7 mm (Figure 1(b)); (3) 9 mm (Figure 1(c)); (4) ferrule-5 mm (Figure 1(d)); (5) ferrule-7 mm (Figure 1(e)); (6) ferrule-9 mm (Figure 1(f)). Preshaping drills (C) were used to prepare the root canals at 5, 7, and 9 mm in depth. After preparation, it was checked that a 3-mm long gutta-percha apical seal. In specimens of groups 3–6 a circumferential collar of tooth structure of 2 mm in height and 3 mm in width was realized with a diamond bur (Figure 2).

Translucent quartz fiber posts (E) consisting of unidirectional, pretensed fibers bound in a translucent epoxy resin matrix, were used. Each post was tried in into the root canal, and the portion of the post extruding out the root was cut to a standardized length of 4.8 [20]. Prior to cementation, a prehydrolyzed silane coupling agent (F) was applied with a microbrush on the post surface for 30 s. The light cured, self-priming adhesive Prime and Bond NT (G) was applied into the root canal with a microbrush for 20 s and gently air-dried. The excess was removed using paper points. The bonding agent was polymerized with a conventional quartz-tungsten-halogen light (750 mW/cm^2) (H). Fluorocore 2 core build-up material (I) was used for fiber post luting. Base and catalyst (1 : 1) were mixed for 30 s, then the material was applied on the post. The post was seated immediately into the canal and sustained under finger pressure. With the same material, a buildup of 2 mm in height was created. After the first 7-minute autocure period, the material was light-cured for 40 s. After curing, the specimens were prepared as for a prosthetic crown, with a circumferential chamfer reduction of 1,5 mm of maximum thickness, using a chamfer bur of 2 mm in diameter (M). After post cementation, each root was embedded in a self-polymerizing acrylic resin (J) for half of the root length, with the long axis sloped at a 45-degree angle to the base of the resin block. During this procedure, specimens were continuously irrigated with water to avoid overheating due to resin polymerization. Before performing the mechanical test, samples were stored for 24 hours at 37°C and 100% relative humidity.

Each sample was then mounted on a universal testing machine (K). A controlled compressive load (cross-head speed: 0.75 mm/min) was applied by means of a stainless steel stylus (Ø 1 mm) at the coronal end of the post extruding out of the root (Figure 3). A software (L) connected to the loading machine recorded the load at failure of the post-retained restoration measured in Newton (N).

3. Results

Descriptive statistics of fracture strength data are reported in Table 1, along with the significance of between-group differences. As the distribution of fracture strengths was not normal according to the Kolmogorov-Smirnov test, the use of the Two-Way Analysis of Variance to assess the influence of depth, ferrule effect, and between-factor interaction was precluded. Therefore, the Kruskal-Wallis One-Way Analysis of Variance was applied with strength as the dependent variable and experimental group as factor. Consequently, the Dunn's multiple range test was used for post hoc comparisons. In all the tests the level of significance was set at $P < 0.05$. Significantly higher fracture strengths were measured in the presence of a ferrule effect. Neither in the presence or in the absence of a ferrule effect had depth of post insertion a significant influence on fracture strength, as no statistically significant differences emerged either among groups 1–3 or among groups 4–6.

The results obtained from this in vitro study showed a correlation between the presence of the ferrule and increased resistance to fracture. In groups 1, 2, and 3 (with ferrule), the mean fracture values were, respectively, 163,8 N, 270,9 N and 254,7 N. These data are higher and statistically significantly different when compared with the three groups 4, 5, and 6, without ferrule effect, in which the values obtained were, respectively, 40,5 N, 41,7 N, and 44,9 N.

The depth of post insertion did not show to be a parameter affecting the results. In fact, no statistically significant differences were found between groups 1, 2, and 3 as well as between groups 4, 5, and 6.

4. Discussion

Since in the presence of a ferrule, significantly higher fracture strengths were measured, the null hypothesis has to be rejected.

Several factors determine the performances and the success of a rehabilitation clinic in time: types, design, and lengths of post, bonding capacity [21], and ferrule. Large variations exist in regard to the physical and fatigue resistance of resin-fiber posts [22]. The static or dynamic behavior of resin-fiber posts depends on the composition

Table 1: Descriptive statistics of fiber post fracture strength data with the significance of between-group differences.

Number group	Name group	N	Mean	Std. Deviation	Median	25%–75%	Significance $P < 0.05$
1	Ferrule-5 mm	10	163,8	72,5	142,9	132,7–181,1	AB
2	Ferrule-7 mm	10	270,9	105,6	244,9	215,2–350,3	A
3	Ferrule-9 mm	10	254,7	79,1	235,4	193,4–305,6	A
4	No ferrule-5 mm	10	40,5	3,1	40,2	38,4–44,2	C
5	No ferrule-7 mm	10	41,7	5,3	43	36,8–46,2	C
6	No ferrule-9 mm	10	44,9	6,7	44,5	40,5–51,7	BC

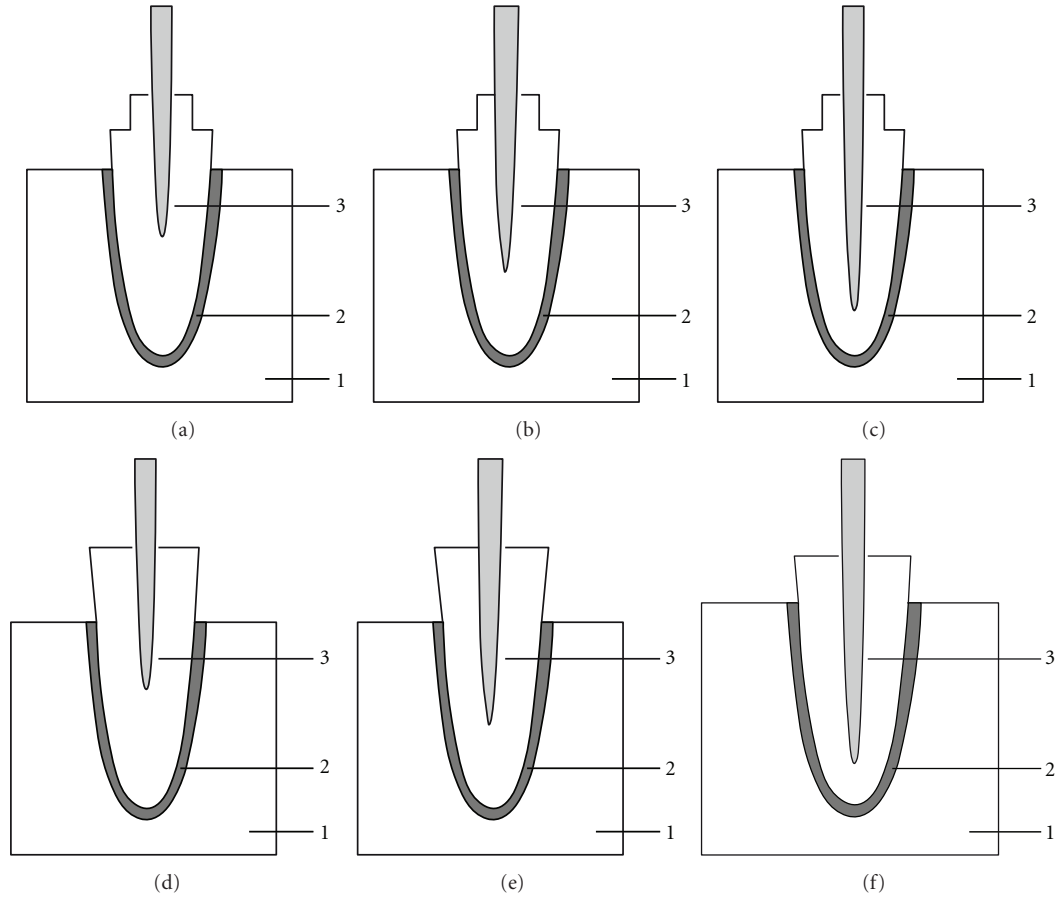


Figure 1: Experimental groups with different post depth (5, 7, and 9 mm) and postspace with (groups a, b, c) and without (groups d, e, f) a ferrule effect.

(fiber type and density) as well as the fabrication process and, in particular, the quality of the resin-fiber interface. In an in vitro study examining physical properties of various posts, it was concluded that the ideal post design comprises a cylindrical coronal portion and a conical apical portion [23]. Much discussed is still the ideal post length, if one part provides greater stability to prosthetic rehabilitation at the same time involves removal of dentin [24] and more because of the existing limitations of adhesive procedures within the root canal [25–27]. It has been demonstrated that the loss of tooth vitality is not accompanied by significant change in tissue moisture or collagen structure [28–30]. The most important changes in tooth biomechanics are attributed to the loss of tissue either at radicular [2, 31] or coronal [31–34] levels, pointing out the importance of a highly conservative approach during endodontic and restorative procedures. The significance of remaining cervical tissue, known as the ferrule, was also well documented [13, 35]. The incorporation of a ferrule is an important factor of tooth preparation when using a post-supported rehabilitation technique [36–38]. The effectiveness of the ferrule has been evaluated with several laboratory tests as fracture resistance, such as [39] impact [40], fatigue [41], and photoelastic analysis [42]. According to these studies the ferrule presence showed values of resistance to fracture much higher and statistically significant differences in groups 1, 2, and 3 than no-ferrule groups (groups 4, 5, 6). Concerning the length of the ferrule, some studies have reported that

Table 2: Classification of instruments used for collecting and measuring data during the tests.

Class	Type
(A)	Flex R File, Union Broach, York, PA, USA
(B)	Dentsply, Maillefer, Tulsa, OK, USA
(C)	DeTrey, Konstanz, Germany
(D)	Fuji II, Gc corp, Tokyo, Japan
(E)	ENDO LIGHT-POST number 3 Bisco, Schaumburg, IL, USA
(F)	Monobond S Ivoclar Vivadent, Schaan, Liechtenstein
(G)	Prime and Bond NT Dentsply DeTrey, Konstanz, Germany
(H)	Optilux 401 Kerr, Danbury, USA
(I)	Fluorocore 2 Dentsply DeTrey, Konstanz, Germany
(J)	ProBase Cold Ivoclar Vivadent, Schaan Fürstentum, Liechtenstein
(K)	Instron Corp, Canton, MA, USA
(L)	Digimax Plus Controls srl, Cernusco s/n, Italy

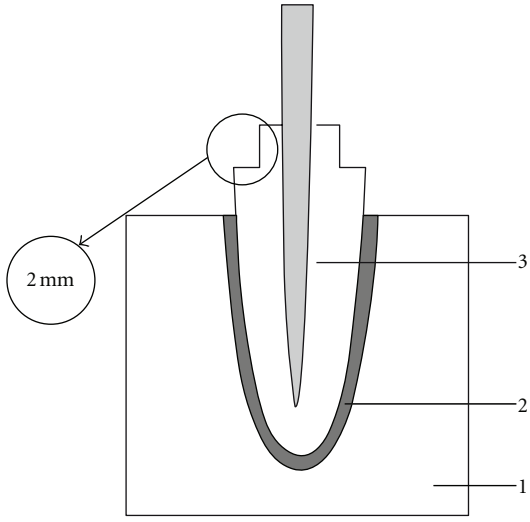


Figure 2: Ferrule effect. A circumferential collar of tooth structure at least 2 mm in height was preserved at the gingival aspect of the preparation.

a tooth should have a minimum amount (2 mm) of coronal structure above the cement-enamel junction (CEJ) to achieve a proper resistance [43, 44].

The results of the present study, in which to assess the mean fracture for each group the force was applied directly on the post head, in order to exclude other variables, have confirmed these observations.

About post insertion depth, it is known that with cast post and core system the post length was an important variable, because reducing post space can permit to save tooth structure positively affecting the tooth fracture resistance. Some authors [45] in a recent study designed to obtain a biofaithful model of the maxillary incisor system and to assess the effect of glass fiber post lengths using Finite Element Analysis showed that the overall system's

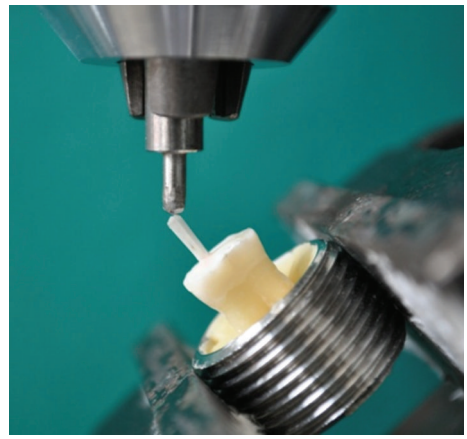


Figure 3: Example of a sample mounted on the loading machine and prepared for the fracture test. The tooth is oriented such as the load applied by means of the metallic stylus would have a 45-degree direction.

strain pattern did not appear to be influenced by post length. This could lead to the conclusion that a post inserted more deeply could be more effective in a fiber post-supported rehabilitation, as the length of the post insertion has a significant effect on retention; the more apically the post is placed in the root, the more retentive is the system [46–48]. This consideration should not be overestimated in clinical practice. The adaptation of the canal shape to the post [49] and the overall length of the root should be in fact taken into consideration, because it has been reported that a residual apical filling of less than 3 mm may result in an unpredictable seal [50, 51].

From the results of the present study, a tendency of the more deeply inserted post to have higher values of resistance to fracture could be anyway observed, particularly in the no-ferrule groups. This might be connected with the use of tapered post, considering that a post inserted more deeply has a wider diameter at the breaking point. The use of a cylindrical shaped post could have minimized this differences, and this could be considered as a limit of the present study, even if Lang et al. [52] showed that if an excessive amount of tooth structure is removed and the natural geometry of the root canal is altered, this will have a destabilizing effect on root-filled teeth. For this reason in clinical practice the use of cylindrical-shaped post have been progressively abandoned and replaced with tapered post.

As general consideration, it should be noted that this in vitro study does not reproduce the exact clinical conditions, where lateral forces should be considered as well as axial forces and fatigue loading, ageing processes, alternate thermal stress, mechanical stress, wear, and water storage. In this in vitro study, in fact, lateral forces were applied with a 45° angle between the post and the loading tip. Moreover, stress applied to the teeth and dental restorations is generally low and repetitive rather than being isolated and loading. However, because of a linear relationship between fatigue and static loading, the compressive static test also

gives valuable information concerning load-bearing capacity [53, 54]. Based on this statement, the results of this in vitro study showed that the ferrule effect positively affects the resistance to fracture of endodontically treated teeth restored with fiber posts. Conversely, post depth of insertion did not affect the resistance to fracture.

5. Conclusion

Within the limitation of this in vitro study, the statistical results showed that the ferrule effect in the endodontically treated teeth positively affects the fracture strength of the fiber post. Conversely, post depth insertion did not affect the resistance to fracture. It could be advisable in the rehabilitation of endodontically treated teeth preserve radicular tissue, reducing the postspace preparation, in order to improve the fracture strength of the post with a ferrule length of at least 2 mm.

References

[1] S. Belli, A. Erdemir, and C. Yildirim, "Reinforcement effect of polyethylene fibre in root-filled teeth: comparison of two restoration techniques," *International Endodontic Journal*, vol. 39, no. 2, pp. 136–142, 2006.

[2] M. Trope and H. L. Ray, "Resistance to fracture of endodontically treated roots," *Oral Surgery Oral Medicine and Oral Pathology*, vol. 73, no. 1, pp. 99–102, 1992.

[3] E. S. Reeh, H. H. Messer, and W. H. Douglas, "Reduction in tooth stiffness as a result of endodontic and restorative procedures," *Journal of Endodontics*, vol. 15, no. 11, pp. 512–516, 1989.

[4] O. Pontius and J. W. Hutter, "Survival rate and fracture strength of incisors restored with different post and core systems and endodontically treated incisors without coronoradicular reinforcement," *Journal of Endodontics*, vol. 28, no. 10, pp. 710–715, 2002.

[5] F. H. O. Mitsui, G. M. Marchi, L. A. F. Pimento, and P. M. Ferraresi, "In vitro study of fracture resistance of bovine roots using different intraradicular post systems," *Quintessence International*, vol. 35, no. 8, pp. 612–616, 2004.

[6] M. Hayashi, Y. Takahashi, S. Imazato, and S. Ebisu, "Fracture resistance of pulpless teeth restored with post-cores and crowns," *Dental Materials*, vol. 22, no. 5, pp. 477–485, 2006.

[7] M. Ferrari, M. C. Cagidiaco, C. Goracci et al., "Long-term retrospective study of the clinical performance of fiber posts," *The American Journal of Dentistry*, vol. 20, no. 5, pp. 287–291, 2007.

[8] M. C. Cagidiaco, C. Goracci, F. Garcia-Godoy, and M. Ferrari, "Clinical studies of fiber posts: a literature review," *International Journal of Prosthodontics*, vol. 21, no. 4, pp. 328–336, 2008.

[9] M. Ferrari, A. Vichi, F. Mannocci, and P. M. Mason, "Retrospective study of the clinical performance of fiber posts," *The American Journal of Dentistry*, vol. 13, no. 2, pp. 9b–13b, 2000.

[10] M. Ferrari, M. C. Cagidiaco, S. Grandini, M. De Sanctis, and C. Goracci, "Post placement affects survival of endodontically treated premolars," *Journal of Dental Research*, vol. 86, no. 8, pp. 729–734, 2007.

[11] G. Heydecke, F. Butz, and J. R. Strub, "Fracture strength and survival rate of endodontically treated maxillary incisors with approximal cavities after restoration with different post and core systems: an in-vitro study," *Journal of Dentistry*, vol. 29, no. 6, pp. 427–433, 2001.

[12] B. Akkayan and T. Gülmez, "Resistance to fracture of endodontically treated teeth restored with different post systems," *Journal of Prosthetic Dentistry*, vol. 87, no. 4, pp. 431–437, 2002.

[13] A. Martínez-Insua, L. da Silva, B. Rilo, and U. Santana, "Comparison of the fracture resistances of pulpless teeth restored with a cast post and core or carbon-fiber post with a composite core," *The Journal of Prosthetic Dentistry*, vol. 80, no. 5, pp. 527–532, 1998.

[14] J. A. Sorensen and M. J. Engelman, "Ferrule design and fracture resistance of endodontically treated teeth," *The Journal of Prosthetic Dentistry*, vol. 63, no. 5, pp. 529–536, 1990.

[15] W. J. Libman and J. I. Nicholls, "Load fatigue of teeth restored with cast posts and cores and complete crowns," *The International Journal of Prosthodontics*, vol. 8, no. 2, pp. 155–161, 1995.

[16] W. A. Saupe, A. H. Gluskin, and R. A. Radke, "A comparative study of fracture resistance between morphologic dowel and cores and a resin-reinforced dowel system in the intraradicular restoration of structurally compromised roots," *Quintessence International*, vol. 27, no. 7, pp. 483–491, 1996.

[17] R. W. Loney, W. E. Kotowicz, and G. C. Mcdowell, "Three-dimensional photoelastic stress analysis of the ferrule effect in cast post and cores," *The Journal of Prosthetic Dentistry*, vol. 63, no. 5, pp. 506–512, 1990.

[18] N. Al-Hazaimeh and D. L. Gutteridge, "An in vitro study into the effect of the ferrule preparation on the fracture resistance of crowned teeth incorporating prefabricated post and composite core restorations," *International Endodontic Journal*, vol. 34, no. 1, pp. 40–46, 2001.

[19] C. Dobó-Nagy, T. Serbán, J. Szabó, G. Nagy, and M. Madléna, "A comparison of the shaping characteristics of two nickel-titanium endodontic hand instruments," *International Endodontic Journal*, vol. 35, no. 3, pp. 283–288, 2002.

[20] E. Asmussen, A. Peutzfeldt, and T. Heitmann, "Stiffness, elastic limit, and strength of newer types of endodontic posts," *Journal of Dentistry*, vol. 27, no. 4, pp. 275–278, 1999.

[21] A. D. Kececi, B. Ureyen Kaya, and N. Adanir, "Micro push-out bond strengths of four fiber-reinforced composite post systems and 2 luting materials," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology*, vol. 105, no. 1, pp. 121–128, 2008.

[22] S. Grandini, C. Goracci, F. Monticelli, F. R. Tay, and M. Ferrari, "Fatigue resistance and structural characteristics of fiber posts: three-point bending test and SEM evaluation," *Dental Materials*, vol. 21, no. 2, pp. 75–82, 2005.

[23] H. Lambjerg-Hansen and E. Asmussen, "Mechanical properties of endodontic posts," *Journal of Oral Rehabilitation*, vol. 24, no. 12, pp. 882–887, 1997.

[24] A. H. L. Tjan and S. B. Whang, "Resistance to root fracture of dowel channels with various thicknesses of buccal dentin walls," *The Journal of Prosthetic Dentistry*, vol. 53, no. 4, pp. 496–500, 1985.

[25] D. Dietschi, S. Ardu, A. Rossier-Gerber, and I. Krejci, "Adaptation of adhesive post and cores to dentin after in vitro occlusal loading: evaluation of post material influence," *Journal of Adhesive Dentistry*, vol. 8, no. 6, pp. 409–419, 2006.

[26] S. Bouillaguet, S. Troesch, J. C. Wataha, I. Krejci, J. M. Meyer, and D. H. Pashley, "Microtensile bond strength between adhesive cements and root canal dentin," *Dental Materials*, vol. 19, no. 3, pp. 199–205, 2003.

[27] F. Mannocci, M. Sherriff, M. Ferrari, and T. F. Watson, "Microtensile bond strength and confocal microscopy of dental adhesives bonded to root canal dentin," *The American Journal of Dentistry*, vol. 14, no. 4, pp. 200–204, 2001.

[28] A. R. Helfer, S. Melnick, and H. Schilder, "Determination of the moisture content of vital and pulpless teeth," *Oral Surgery, Oral Medicine, Oral Pathology*, vol. 34, no. 4, pp. 661–670, 1972.

[29] J. L. Gutmann, "The dentin-root complex: anatomic and biologic considerations in restoring endodontically treated teeth," *The Journal of Prosthetic Dentistry*, vol. 67, no. 4, pp. 458–467, 1992.

[30] E. M. Rivera and M. Yamauchi, "Site comparisons of dentine collagen cross-links from extracted human teeth," *Archives of Oral Biology*, vol. 38, no. 7, pp. 541–546, 1993.

[31] E. S. Reeh, H. H. Messer, and W. H. Douglas, "Reduction in tooth stiffness as a result of endodontic and restorative procedures," *Journal of Endodontics*, vol. 15, no. 11, pp. 512–516, 1989.

[32] W. H. Douglas, "Methods to improve fracture resistance of teeth," in *Proceedings of the International Symposium on Posterior Composite Resin Dental Restorative Materials*, G. Vanherle and D. C. Smith, Eds., pp. 433–441, Peter Szulc Publishing, Utrecht, The Netherlands, 1985.

[33] J. Linn and H. H. Messer, "Effect of restorative procedures on the strength of endodontically treated molars," *Journal of Endodontics*, vol. 20, no. 10, pp. 479–485, 1994.

[34] P. Pantvisai and H. H. Messer, "Cuspal deflection in molars in relation to endodontic and restorative procedures," *Journal of Endodontics*, vol. 21, no. 2, pp. 57–61, 1995.

[35] P. R. Cathro, N. P. Chandler, and J. A. Hood, "Impact resistance of crowned endodontically treated central incisors with internal composite cores," *Endodontics and Dental Traumatology*, vol. 12, no. 3, pp. 124–128, 1996.

[36] H. Rosen, "Operative procedures on mutilated endodontically treated teeth," *The Journal of Prosthetic Dentistry*, vol. 11, no. 5, pp. 973–986, 1961.

[37] A. G. Gegauff, "Effect of crown lengthening and ferrule placement on static load failure of cemented cast post-cores and crowns," *Journal of Prosthetic Dentistry*, vol. 84, no. 2, pp. 169–179, 2000.

[38] J. R. Pereira, F. de Ornelas, P. C. Conti, and A. L. do Valle, "Effect of a crown ferrule on the fracture resistance of endodontically treated teeth restored with prefabricated posts," *Journal of Prosthetic Dentistry*, vol. 95, no. 1, pp. 50–54, 2006.

[39] J. R. Pereira, T. M. Neto, V. d. C. Porto, L. F. Pegoraro, and A. L. do Valle, "Influence of the remaining coronal structure on the resistance of teeth with intraradicular retainer," *Brazilian Dental Journal*, vol. 16, no. 3, pp. 197–201, 2005.

[40] P. R. Cathro, N. P. Chandler, and J. A. Hood, "Impact resistance of crowned endodontically treated central incisors with internal composite cores," *Endodontics and Dental Traumatology*, vol. 12, no. 3, pp. 124–128, 1996.

[41] F. Isidor, K. Brøndum, and G. Ravnholt, "The influence of post length and crown ferrule length on the resistance to cyclic loading of bovine teeth with prefabricated titanium posts," *International Journal of Prosthodontics*, vol. 12, no. 1, pp. 79–82, 1999.

[42] R. W. Loney, W. E. Kotowicz, and G. C. Mcdowell, "Three-dimensional photoelastic stress analysis of the ferrule effect in cast post and cores," *The Journal of Prosthetic Dentistry*, vol. 63, no. 5, pp. 506–512, 1990.

[43] K. C. Trabert and J. P. Cooney, "The endodontically treated tooth: restorative concepts and techniques," *Dental Clinics of North America*, vol. 28, no. 4, pp. 923–951, 1984.

[44] G. W. Wagnild and K. L. Mueller, "Restoration of the endodontically treated tooth," in *Pathways of the Pulp*, S. Cohen and R. C. Burns, Eds., pp. 765–795, Elsevier Saunders, St. Louis, Mo, USA, 8th edition, 2001.

[45] M. Ferrari, R. Sorrentino, F. Zarone, D. Apicella, R. Aversa, and A. Apicella, "Non-linear viscoelastic finite element analysis of the effect of the length of glass fiber posts on the biomechanical behaviour of directly restored incisors and surrounding alveolar bone," *Dental Materials Journal*, vol. 27, no. 4, pp. 485–498, 2008.

[46] J. P. Standlee, A. A. Caputo, and E. C. Hanson, "Retention of endodontic dowels: effects of cement, dowel length, diameter, and design," *The Journal of Prosthetic Dentistry*, vol. 39, no. 4, pp. 400–405, 1978.

[47] J. Nissan, Y. Dmitry, and D. Assif, "The use of reinforced composite resin cement as compensation for reduced post length," *Journal of Prosthetic Dentistry*, vol. 86, no. 3, pp. 304–308, 2001.

[48] I. Nergiz, P. Schmage, M. Özcan, and U. Platzer, "Effect of length and diameter of tapered posts on the retention," *Journal of Oral Rehabilitation*, vol. 29, no. 1, pp. 28–34, 2002.

[49] M. K. Wu, A. R'oris, D. Barkis, and P. R. Wesselink, "Prevalence and extent of long oval canals in the apical third," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics*, vol. 89, no. 6, pp. 739–743, 2000.

[50] L. Abramovitz, R. Lev, Z. Fuss, and Z. Metzger, "The unpredictability of seal after post space preparation: a fluid transport study," *Journal of Endodontics*, vol. 27, no. 4, pp. 292–295, 2001.

[51] M. K. Wu, Y. Pehlivan, E. G. Kontakiotis, and P. R. Wesselink, "Microleakage along apical root fillings and cemented posts," *The Journal of Prosthetic Dentistry*, vol. 79, no. 3, pp. 264–269, 1998.

[52] H. Lang, Y. Korkmaz, K. Schneider, and W. H. M. Raab, "Impact of endodontic treatments on the rigidity of the root," *Journal of Dental Research*, vol. 85, no. 4, pp. 364–368, 2006.

[53] S. Garoushi, L. V. J. Lassila, A. Tezvergil, and P. K. Vallittu, "Static and fatigue compression test for particulate filler composite resin with fiber-reinforced composite substructure," *Dental Materials*, vol. 23, no. 1, pp. 17–23, 2007.

[54] M. Naumann, G. Sterzenbach, and P. Pröschel, "Evaluation of load testing of postendodontic restorations in vitro: linear compressive loading, gradual cycling loading and chewing simulation," *Journal of Biomedical Materials Research B*, vol. 74, no. 2, pp. 829–834, 2005.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2012, Article ID 651564, 9 pages
doi:10.1155/2012/651564

Research Article

Optimization and Implementation of Scaling-Free CORDIC-Based Direct Digital Frequency Synthesizer for Body Care Area Network Systems

Ying-Shen Juang,[1] Lu-Ting Ko,[2] Jwu-E. Chen,[2] Tze-Yun Sung,[3] and Hsi-Chin Hsin[4]

[1] *Department of Business Administration, Chung Hua University, Hsinchu City 300-12, Taiwan*
[2] *Department of Electrical Engineering, National Central University, Chungli City 320-01, Taiwan*
[3] *Department of Microelectronics Engineering, Chung Hua University, Hsinchu City 300-12, Taiwan*
[4] *Department of Computer Science and Information Engineering, National United University, Miaoli 360-03, Taiwan*

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

Received 11 August 2012; Accepted 15 September 2012

Academic Editor: Sheng-yong Chen

Coordinate rotation digital computer (CORDIC) is an efficient algorithm for computations of trigonometric functions. Scaling-free-CORDIC is one of the famous CORDIC implementations with advantages of speed and area. In this paper, a novel direct digital frequency synthesizer (DDFS) based on scaling-free CORDIC is presented. The proposed multiplier-less architecture with small ROM and pipeline data path has advantages of high data rate, high precision, high performance, and less hardware cost. The design procedure with performance and hardware analysis for optimization has also been given. It is verified by Matlab simulations and then implemented with field programmable gate array (FPGA) by Verilog. The spurious-free dynamic range (SFDR) is over 86.85 dBc, and the signal-to-noise ratio (SNR) is more than 81.12 dB. The scaling-free CORDIC-based architecture is suitable for VLSI implementations for the DDFS applications in terms of hardware cost, power consumption, SNR, and SFDR. The proposed DDFS is very suitable for medical instruments and body care area network systems.

1. Introduction

Direct digital frequency synthesizer (DDFS) has been widely used in the modern communication systems. DDFS is preferable to the classical phase-locked-loop- (PLL-) based synthesizer in terms of switching speed, frequency resolution, and phase noise, which are beneficial to the high-performance communication systems. Figure 1 depicts the conventional DDFS architecture [1], which consists of a phase accumulator, a sine/cosine generator, a digital-to-analog converter (DAC), and a low-pass filter (LPF). As noted, two inputs: the reference clock and the frequency control word (FCW) are used; the phase accumulator integrates FCW to produce an angle in the interval of $[0, 2\pi)$, and the sine/cosine generator computes the sinusoidal values. In practice, the sine/cosine generator is implemented digitally, and thus followed by digital-to-analog conversion and low-pass filtering for analogue outputs. Such systems can be applied in many fields, especially in industrial, biological, and medical applications [2–4].

The simplest way to implement the sine/cosine generator is to use ROM lookup table (LUT). However, a large ROM is needed [5]. Several efficient compression techniques have been proposed to reduce the lookup table size [5–10]. The quadrant compression technique can compress the lookup table and then reduce the ROM size by 75% [6]. The Sunderland architecture splits the ROM into two smaller memories [7], and the Nicholas architecture improves the Sunderland architecture to achieve a higher ROM-compression ratio (32 : 1) [8]. The ROM size can be further reduced by using the polynomial approximations [11–18] or CORDIC algorithm [19–27]. In the polynomial approximations-based DDFSs, the interval of $[0, \pi/4]$ is divided into subintervals, and sine/cosine functions are evaluated in each subinterval.

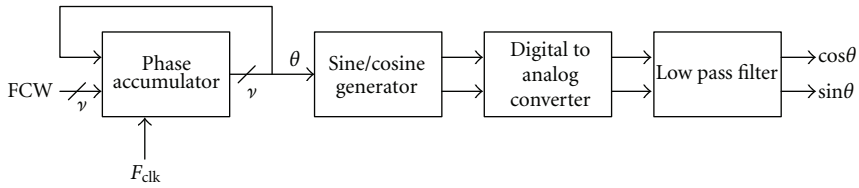


Figure 1: The conventional DDFS architecture.

The polynomial approximations-based DDFS requires a ROM to store the coefficients of the polynomials and the polynomial evaluation hardware with multipliers. In the circular mode of CORDIC, which is an iterative algorithm to compute sine/cosine functions, an initial vector is rotated with a predetermined sequence of subangles such that the summation of the rotations approaches the desired angle [28, 29]. CORDIC has been widely used for the sine/cosine generator of DDFS [19–27]. Compared to the lookup table-based DDFS, the CORDIC-based DDFS has the advantage of avoiding the exponential growth of hardware complexity while the output word size increases [30–33].

In Figure 1, the word length of the phase accumulator is v bits; thus, the period of the output signal is as follows:

$$T_o = \frac{2^v T_s}{\mathrm{FCW}}, \tag{1}$$

where FCW is the phase increment and T_s denotes the sampling period. It is noted that the output frequency can be written by

$$F_o = \frac{1}{T_0} = \frac{F_s}{2^v} \cdot \mathrm{FCW}. \tag{2}$$

According to the equation above, the minimum change of output frequency is given by

$$\Delta F_{o,\min} = \frac{F_s}{2^v}(\mathrm{FCW}+1) - \frac{F_s}{2^v}\mathrm{FCW} = \frac{F_s}{2^v}. \tag{3}$$

Thus, the frequency resolution of DDFS is dependent on the word length of the phase accumulator as follows:

$$\Delta F_o \geq \frac{F_s}{2^v}. \tag{4}$$

The bandwidth of DDFS is defined as the difference between the highest and the lowest output frequencies. The highest frequency is determined by either the maximum clock rate or the speed of logic circuitries; the lowest frequency is dependent on FCW. Spurious-free dynamic range (SFDR) is defined as the ratio of the amplitude of the desired frequency component to that of the largest undesired one at the output of DDFS, which is often represented in $\mathrm{dB_c}$ as follows:

$$\mathrm{SFDR} = 20\log\left(\frac{A_p}{A_s}\right) = 20\log\left(A_p\right) - 20\log(A_s), \tag{5}$$

where A_p is the amplitude of the desired frequency component and A_s is the amplitude of the largest undesired one.

In this paper, a novel DDFS architecture based on the scaling-free CORDIC algorithm [34] with ROM mapping is presented. The rest of the paper is organized as follows. In Section 2, CORDIC is reviewed briefly. In Section 3, the proposed DDFS architecture is presented. In Section 4, the hardware implementation of DDFS is given. Conclusion can be found in Section 5.

2. The CORDIC Algorithm

CORDIC is an efficient algorithm that evaluates various elementary functions including sine and cosine functions. As hardware implementation might only require simple adders and shifters, CORDIC has been widely used in the high speed applications.

2.1. The CORDIC Algorithm in the Circular Coordinate System. A rotation of angle θ in the circular coordinate system can be obtained by performing a sequence of micro-rotations in the iterative manner. Specifically, a vector can be successively rotated by the use of a sequence of pre-determined step-angles: $\alpha(i) = \tan^{-1}(2^{-i})$. This methodology can be applied to generate various elementary functions, in which only simple adders and shifters are required. The conventional CORDIC algorithm in the circular coordinate system is as follows [28, 29]:

$$x(i+1) = x(i) - \sigma(i)2^{-i}y(i), \tag{6}$$

$$y(i+1) = y(i) + \sigma(i)2^{-j}x(i), \tag{7}$$

$$z(i+1) = z(i) - \sigma(i)\alpha(i), \tag{8}$$

$$\alpha(i) = \tan^{-1}2^{-i}, \tag{9}$$

where $\sigma(i) \in \{-1, +1\}$ denotes the direction of the ith micro-rotation, $\sigma_i = \mathrm{sign}(z(i))$ with $z(i) \to 0$ in the vector rotation mode [34], $\sigma_i = -\mathrm{sign}(x(i)) \cdot \mathrm{sign}(y(i))$ with $y(i) \to 0$ in the angle accumulated mode [34], the corresponding scale factor $k(i)$ is equal to $\sqrt{1+\sigma^2(i)2^{-2i}}$, and $i = 0, 1, \ldots, n-1$. The product of the scale factors after n micro-rotations is given by

$$K_1 = \prod_{i=0}^{n-1} k(i) = \prod_{i=0}^{n-1} \sqrt{1+2^{-2i}}. \tag{10}$$

In the vector rotation mode, $\sin\theta$ and $\cos\theta$ can be obtained with the initial value: $(x(0), y(0)) = (1/K_1, 0)$. More

specifically, x_{out} and y_{out} are computed from the initial value: $(x_{\text{in}}, y_{\text{in}}) = (x(0), y(0))$ as follows:

$$\begin{bmatrix} x_{\text{out}} \\ y_{\text{out}} \end{bmatrix} = K_1 \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_{\text{in}} \\ y_{\text{in}} \end{bmatrix}. \tag{11}$$

2.2. Scaling-Free CORDIC Algorithm in the Circular Coordinate System. Based on the following approximations of sine and cosine functions:

$$\begin{aligned} \sin\alpha(i) &\cong \alpha(i) = 2^{-i}, \\ \cos\alpha(i) &\cong 1 - \frac{\alpha^2(i)}{2} = 1 - 2^{-(2i+1)}, \end{aligned} \tag{12}$$

the scaling-free CORDIC algorithm is thus obtained by using (6), (7), and the above. In which, the iterative rotation is as follows:

$$\begin{aligned} \begin{bmatrix} x(i+1) \\ y(i+1) \end{bmatrix} &= \begin{bmatrix} 1 - 2^{-(2i+1)} & 2^{-i} \\ -2^{-i} & 1 - 2^{-(2i+1)} \end{bmatrix} \begin{bmatrix} x(i) \\ y(i) \end{bmatrix}, \\ z(i+1) &= z(i) - 2^{-i}. \end{aligned} \tag{13}$$

For the word length of w bits, it is noted that the implementation of scaling-free CORDIC algorithm utilizes four shifters and four adders for each micro-rotation in the first $w/2$-microrotations; it reduces two shifters and two adders for each microrotation in the last $w/2$-micro-rotations [24, 34, 35].

3. Design and Optimization of the Scaling-Free CORDIC-Based DDFS Architecture

In this section, the architecture together with performance analysis of the proposed DDFS is presented. It is a combination of the scaling-free-CORDIC algorithm and LUT; this hybrid approach takes advantage of both CORDIC and LUT to achieve high precision and high data rate, respectively. The proposed DDFS architecture consists of phase accumulator, radian converter, sine/cosine generator, and output stage.

3.1. Phase Accumulator. Figure 2 shows the phase accumulator, which consists of a 32-bit adder to accumulate the phase angle by FCW recursively. At time n, the output of phase accumulator is $\phi = (n \cdot \text{FCW})/2^{32}$ and the sine/cosine generator produces $\sin((n \cdot \text{FCW})/2^{32})$ and $\cos((n \cdot \text{FCW})/2^{32})$. The *load* control signal is used for FCW to be loaded into the register, and the *reset* signal is to initialize the content of the phase accumulator to zero.

3.2. Radian Converter. In order to convert the output of the phase accumulator into its binary representation in radians, the following strategy has been adopted. Specifically, an efficient ROM reduction scheme based on the symmetry property of sinusoidal wave can be obtained by simple logic operations to reconstruct the sinusoidal wave from its first quadrant part only. In which, the first two MSBs of an angle

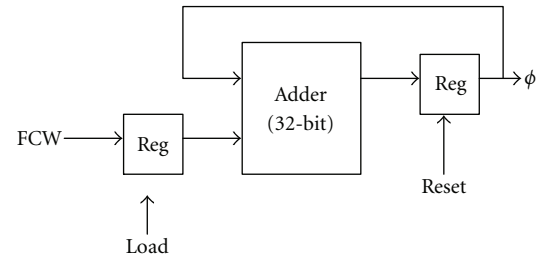


Figure 2: The phase accumulator in DDFS.

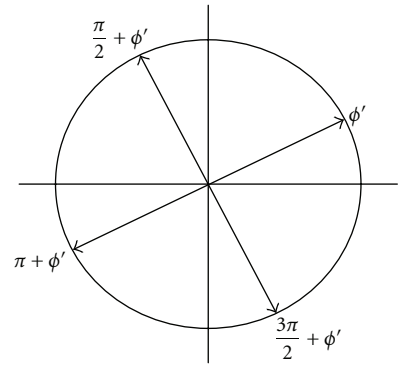


Figure 3: Symmetry-based map of an angle in either the second, third, or fourth quadrant to the corresponding angle in the first quadrant.

indicate the quadrant of the angle in the circular coordinate and the third MSB indicates the half portion of the quadrant; thus, the first three MSBs of an angle are used to control the interchange/negation operation in the output stage. As shown in Figure 3, the corresponding angles of ϕ' in the second, third, and fourth quadrants can be mapped into the first quadrant by setting the first two MSBs to zero. The radian of ϕ' is therefore obtained by $\theta = (\pi/4)\phi'$, which can be implemented by using simple shifters and adders array shown in Figure 4. Note that the third MSB of any radian value in the upper half of a quadrant is 1, and the sine/cosine of an angle γ in the upper half of a quadrant can be obtained from the corresponding angle in the lower half as shown in Figure 5. More specifically, as $\cos\gamma = \sin((\pi/2) - \gamma)$ and $\sin\gamma = \cos((\pi/2) - \gamma)$, the normalized angle can be obtained by replacing θ with $\theta' = 0.5 - \theta$ while the third MSB is 1. In case the third MSB is 0, there is no need to perform the replacement as $\theta' = \theta$.

3.3. Sine/Cosine Generator. As the core of the DDFS architecture, the sine/cosine generator produces sinusoidal waves based on the output of the radian converter. Without loss of generality, let the output resolution be of 16 bits, for the sine/cosine generator consisting of a cascade of w processors, each of which performs the sub-rotation by a fixed angle of 2^{-i} radian as follows:

$$\begin{aligned} x(i+1) &= \left(1 - \sigma(i)2^{-(2i+1)}\right)x(i) + \sigma(i)2^{-i}y(i), \\ y(i+1) &= \left(1 - \sigma(i)2^{-(2i+1)}\right)y(i) - \sigma(i)2^{-i}x(i). \end{aligned} \tag{14}$$

Table 1: The hardware costs in 16-bit DDFS with respect to the number of the replaced CORDIC stages (m: the number of the replaced CORDIC stages, 16-bit adder: 200 gates, 16-bit shift: 90 gates, and 1-bit ROM: 1 gate).

m	0	1	2	3	4	5	6	7
CORDIC processor requirement:								
CORDIC processor-A	7	5	4	3	2	1	0	0
CORDIC processor-B	9	9	9	9	9	9	9	8
Hardware cost:								
16-bit Adders	46	38	34	30	26	22	18	16
16-bit Shifters	46	38	34	30	26	22	18	16
ROM size (bits)	4×16	8×16	14×16	26×16	50×16	102×16	194×16	386×16
Total gate counts	13404	11148	10084	9116	8340	8012	8324	10816

Table 2: Control signals of the output stage.

MSB's of ϕ	ϕ	$xinv$	$yinv$	$swap$	$\cos 2\pi\phi$	$\sin 2\pi\phi$
0 0 0	$0 < 2\pi\phi < \frac{\pi}{4}$	0	0	0	$\cos\theta$	$\sin\theta$
0 0 1	$\frac{\pi}{4} < 2\pi\phi < \frac{\pi}{2}$	0	0	1	$\sin\theta$	$\cos\theta$
0 1 0	$\frac{\pi}{2} < 2\pi\phi < \frac{3\pi}{4}$	0	1	1	$-\sin\theta$	$\cos\theta$
0 1 1	$\frac{3\pi}{4} < 2\pi\phi < \pi$	1	0	0	$-\cos\theta$	$\sin\theta$
1 0 0	$-\pi < 2\pi\phi < -\frac{3\pi}{4}$	1	1	0	$-\cos\theta$	$-\sin\theta$
1 0 1	$-\frac{3\pi}{4} < 2\pi\phi < -\frac{\pi}{2}$	1	1	1	$-\sin\theta$	$-\cos\theta$
1 1 0	$-\frac{\pi}{2} < 2\pi\phi < -\frac{\pi}{4}$	1	0	1	$\sin\theta$	$-\cos\theta$
1 1 1	$-\frac{\pi}{4} < 2\pi\phi < 0$	0	1	0	$\cos\theta$	$-\sin\theta$

Table 3: Comparisons of the proposed DDFS with other related works.

DDFS	Kang and Swartzlander, 2006 [23]	Sharma et al., 2009 [26]	Jafari et al., 2005 [17]	Ashrafi and Adhami, 2007 [18]	Yi et al., 2006 [6]	De Caro et al., 2009 [27]	This work, Juang et al., 2012
Process (μm)	0.13	—	0.5	0.35	0.35	0.25	0.18
Core area (mm^2)	0.35	—	—	—	—	0.51	0.204
Maximum sampling rate (MHz)	1018	230	106	210	100	385	500
Power consumption (mW)	0.343	—	—	112	0.81	0.4	0.302
SFDR (dB_c)	90	54	—	72.2	80	90	86.85
SNR (dB)	—	—	—	67	—	70	81.12
Output resolution (bit)	17	10	14	12	16	13	16
Tuning latency (clock)	—	—	33	—	—	—	11

For $8 \le i < 16$

$$x(i+1) = x(i) + \sigma(i)2^{-i}y(i),$$
$$y(i+1) = y(i) - \sigma(i)2^{-i}x(i), \quad (15)$$

where $\sigma(i) \in \{1, 0\}$ representing the positive or zero subrotation, respectively. Figure 6 depicts the CORDIC processor-A for the first 7 microrotations, which consists of four 16-bit adders and four 16-bit shifters. The CORDIC processor-B with two 16-bit adders and two 16-bit shifters for the last 9 microrotations is shown in Figure 7.

The first m CORDIC stages can be replaced by simple LUT to reduce the data path at the cost of hardware complexity increasing exponentially. Table 1 depicts the hardware costs in 16-bit DDFS with respect to the number of the replaced CORDIC-stages, where each 16-bit adder, 16-bit

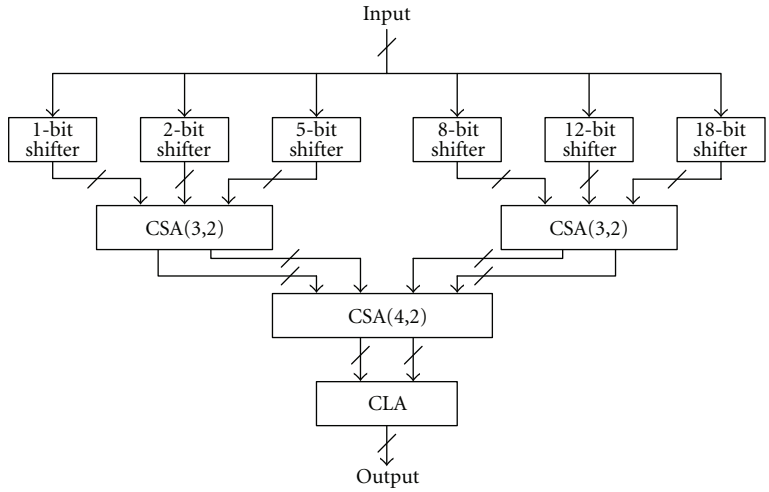


Figure 4: The constant ($\pi/4$) multiplier.

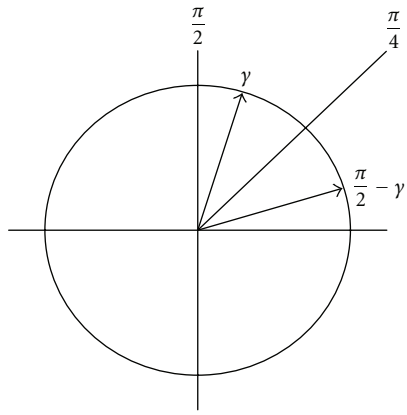


Figure 5: $\pi/4$-mirror map of an angle γ above $\pi/4$ to the corresponding angle $\pi/2 - \gamma$ below $\pi/4$.

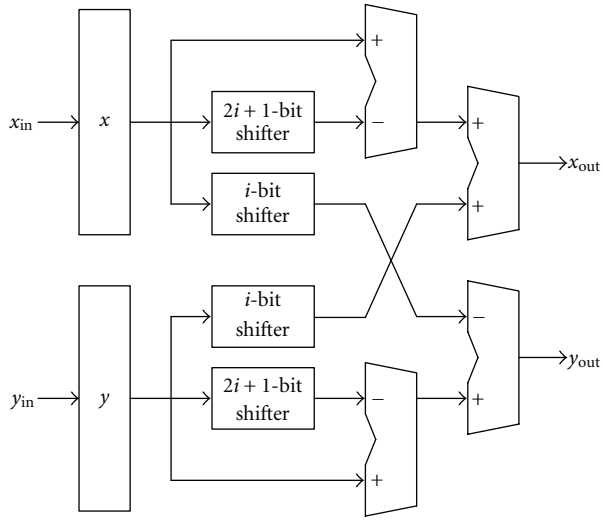


Figure 6: The CORDIC processor-A.

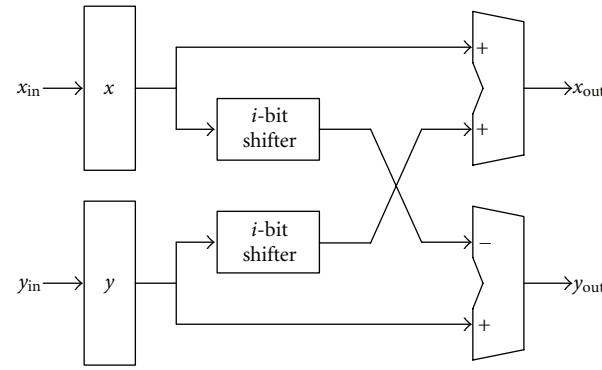


Figure 7: The CORDIC processor-B.

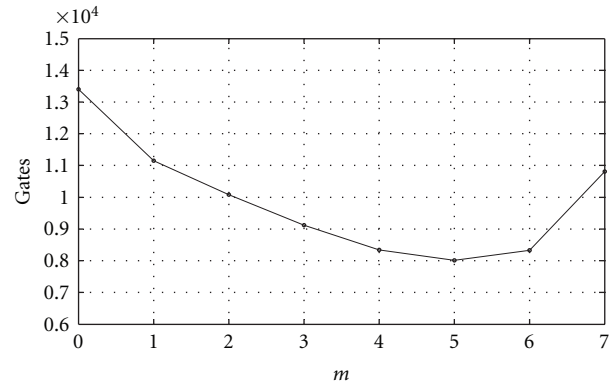


Figure 8: Hardware requirements with respect to the replaced CORDIC stages.

shifter, and 1-bit memory require 200 gates, 90 gates, and 1 gate [36], respectively. Figure 8 shows the hardware requirements with respect to the number of the replaced CORDIC-stages [24]. Figure 9 shows the SFDR/SNRs with respect to

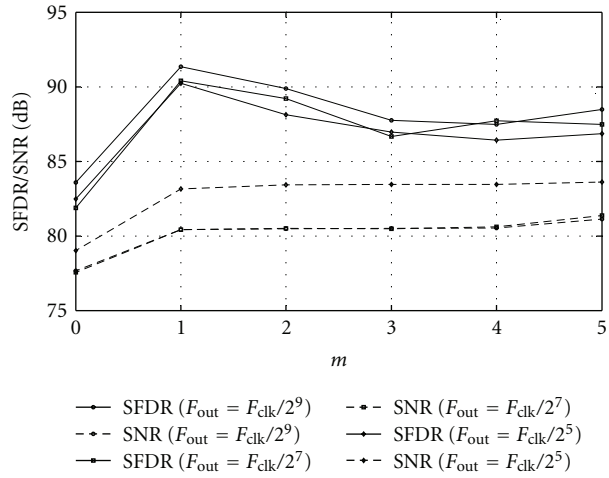


Figure 9: SFDR/SNRs with respect to the replaced CORDIC-stages.

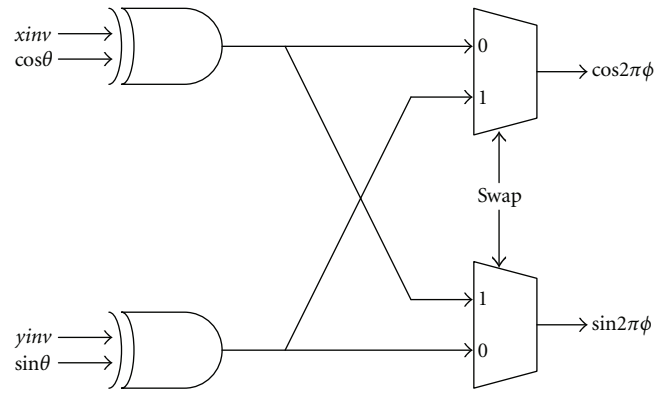


Figure 10: The output stage.

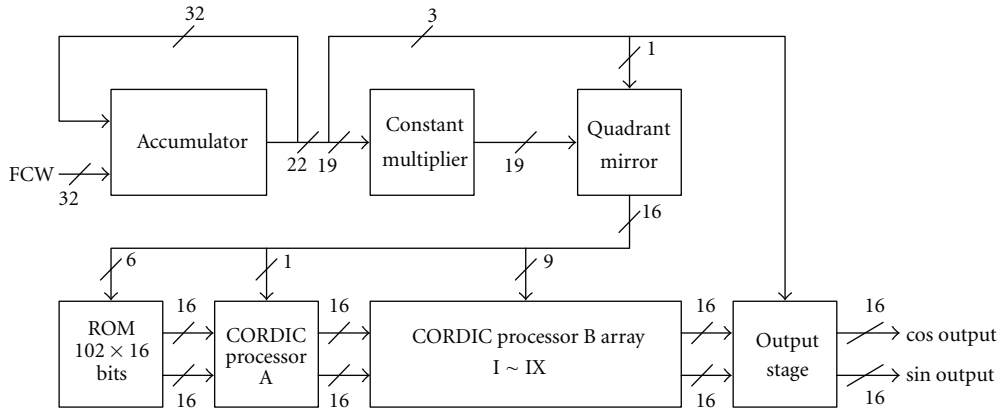


Figure 11: The proposed DDFS architecture.

the replaced CORDIC-stages [25]. As one can expect, based on the above figures, there is a tradeoff between hardware complexity and performance in the design of DDFS.

3.4. Output Stage. Figure 10 shows the architecture of output stage, which maps the computed $\sin\theta$ and $\cos\theta$ to the desired $\sin\phi$ and $\cos\phi$. As mentioned previously, the above mapping can be accomplished by simple negation and/or interchange operations. The three control signals: *xinv*, *yinv*, and *swap* derived from the first three MSBs of ϕ are shown in Table 2. *xinv* and *yinv* are for the negation operation of the output and *swap* for the interchange operation.

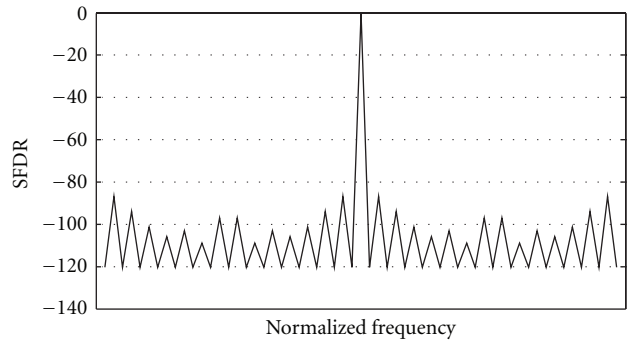


Figure 12: SFDR of the proposed DDFS architecture at output frequency $F_{clk}/2^5$.

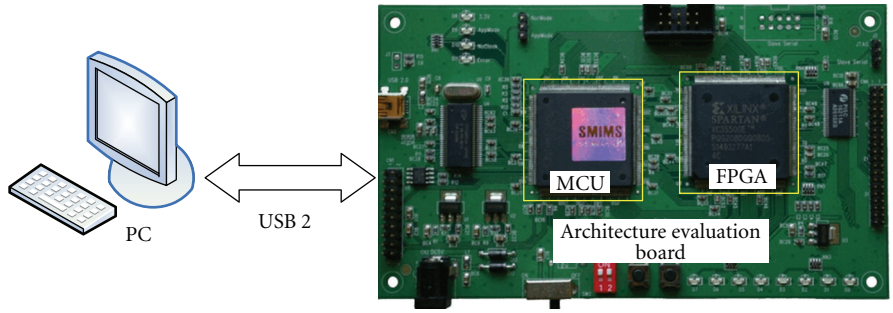


Figure 13: Block diagram and circuit board of the architecture development and verification platform.

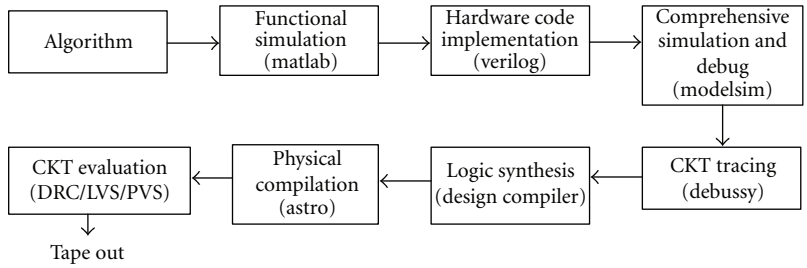


Figure 14: Cell-based design flow.

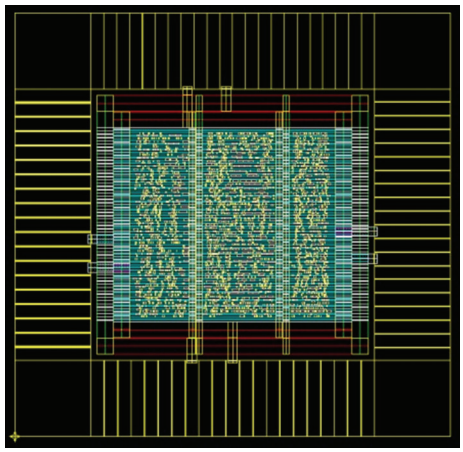


Figure 15: Layout view of the proposed scaling-free-CORDIC-based DDFS.

4. Hardware Implementation of the Scaling-Free CORDIC-Based DDFS

In this section, the proposed low-power and high-performance DDFS architecture ($m = 5$) is presented. Figure 11 depicts the system block diagram; SFDR of the proposed DDFS architecture at output frequency $F_{clk}/2^5$ is shown in Figure 12. As one can see, the SFDR of the proposed architecture is more than 86.85 dBc.

The platform for architecture development and verification has also been designed as well as implemented to evaluate the development cost [37–40]. The proposed DDFS architecture has been implemented on the Xilinx FPGA emulation board [41]. 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 13 depicts block diagram and circuit board of the architecture development and evaluation platform. In which, the microcontroller read data and commands from PC and writes the results back to PC via USB 2.0 bus; the Xilinx Spartan-3 FPGA implements the proposed DDFS architecture. The hardware code in Verilog runs on PC with the ModelSim simulation tool [42] and Xilinx ISE smart compiler [43]. 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 DDFS architecture is able to improve the power consumption and computation speed significantly. Moreover, all the control signals are internally generated on-chip. The proposed DDFS provides both high performance and less hardware.

The chip has been synthesized by using the TSMC 0.18 μm 1P6M CMOS cell libraries [44]. The physical circuit has been synthesized by the Astro tool. The circuit has been evaluated by DRC, LVS, and PVS [45]. Figure 14 shows the cell-based design flow.

Figure 15 shows layout view of the proposed scaling-free CORDIC-based DDFS. The core size obtained by the Synopsys design analyzer is $452 \times 452\,\mu\text{m}^2$. The power consumption obtained by the PrimePower is 0.302 mW with clock rate of 500 MHz at 1.8 V. The tuning latency is 11 clock cycles. All of the control signals are internally generated on-chip. The chip provides both high throughput and low gate count.

5. Conclusion

In this paper, we present a novel DDFS architecture-based on the scaling-free CORDIC algorithm with small ROM and pipeline data path. Circuit emulation shows that the proposed high performance architecture has the advantages of high precision, high data rate, and simple hardware. For 16-bit DDFS, the SFDR of the proposed architecture is more than 86.85 dBc. As shown in Table 3, the proposed DDFS is superior to the previous works in terms of SFDR, SNR, output resolution, and tuning latency [6, 17, 18, 26, 27]. According to the high performance of the proposed DDFS, it is very suited for medical instruments and body care network systems [46–49]. The proposed DDFS with the use of the portable Verilog is a reusable IP, which can be implemented in various processes with tradeoffs of performance, area, and power consumption.

Acknowledgment

The National Science Council of Taiwan under Grants NSC100-2628-E-239-002-MY2 and NSC100-2410-H-216-003 supported this work.

References

[1] J. Tierney, C. Rader, and B. Gold, "A digital frequency synthesizer," *IEEE Transactions on Audio and Electroacoustics*, vol. 19, no. 1, pp. 48–57, 1971.

[2] S. Chen, M. Zhao, G. Wu, C. Yao, and J. Zhang, "Recent advances in morphological cell image analysis," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 101536, 10 pages, 2012.

[3] Z. Teng, A. J. Degnan, U. Sadat et al., "Characterization of healing following atherosclerotic carotid plaque rupture in acutely symptomatic patients: an exploratory study using in vivo cardiovascular magnetic resonance," *Journal of Cardiovascular Magnetic Resonance*, vol. 13, article 64, 2011.

[4] S. Chen and X. Li, "Functional magnetic resonance imaging for imaging neural activity in the human brain: the annual progress," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 613465, 9 pages, 2012.

[5] J. Vankka, "Methods of mapping from phase to sine amplitude in direct digital synthesis," in *Proceedings of the 50th IEEE International Frequency Control Symposium*, pp. 942–950, June 1996.

[6] S. C. Yi, K. T. Lee, J. J. Chen, and C. H. Lin, "A low-power efficient direct digital frequency synthesizer based on new two-level lookup table," in *Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE '06)*, pp. 963–966, May 2006.

[7] D. A. Sunderland, R. A. Strauch, S. S. Wharfield, H. T. Peterson, and C. R. Cole, "CMOS/SOS frequency synthesizer LSI circuit for spread spectrum communications," *IEEE Journal of Solid-State Circuits*, vol. 19, no. 4, pp. 497–506, 1984.

[8] H. T. Nicholas, H. Samueli, and B. Kim, "Optimization of direct digital frequency synthesizer performance in the presence of finite word length effects," in *Proceedings of the 42nd Annual Frequency Control Symposium*, pp. 357–363, June 1988.

[9] L. A. Weaver and R. J. Kerr, "High resolution phase to sine amplitude conversion," U.S. Patent 4 905 177, 1990.

[10] A. Bonfanti, D. De Caro, A. D. Grasso, S. Pennisi, C. Samori, and A. G. M. Strollo, "A 2.5-GHz DDFS-PLL with 1.8-MHz bandwidth in 0.35-μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 6, pp. 1403–1413, 2008.

[11] A. Bellaouar, M. S. O'brecht, A. M. Fahim, and M. I. Elmasry, "Low-power direct digital frequency synthesis for wireless communications," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 3, pp. 385–390, 2000.

[12] A. Bellaouar, M. S. O'Brecht, and M. I. Elmasry, "Low-power direct digital frequency synthesizer architecture," U.S. Patent 5 999 581, 1999.

[13] M. M. El Said and M. I. Elmasry, "An improved ROM compression technique for direct digital frequency synthesizers," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 437–440, May 2002.

[14] G. C. Gielis, R. van de Plassche, and J. van Valburg, "A 540-MHz 10-b polar-to-Cartesian converter," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 11, pp. 1645–1650, 1991.

[15] D. De Caro, E. Napoli, and A. G. M. Strollo, "Direct digital frequency synthesizers with polynomial hyperfolding technique," *IEEE Transactions on Circuits and Systems II*, vol. 51, no. 7, pp. 337–344, 2004.

[16] Y. H. Chen and Y. A. Chau, "A direct digital frequency synthesizer based on a new form of polynomial approximations," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 436–440, 2010.

[17] H. Jafari, A. Ayatollahi, and S. Mirzakuchaki, "A low power, high SFDR, ROM-less direct digital frequency synthesizer," in *Proceedings of the IEEE Conference on Electron Devices and Solid-State Circuits (EDSSC '05)*, pp. 829–832, December 2005.

[18] A. Ashrafi and R. Adhami, "Theoretical upperbound of the spurious-free dynamic range in direct digital frequency synthesizers realized by polynomial interpolation methods," *IEEE*

Transactions on Circuits and Systems I, vol. 54, no. 10, pp. 2252–2261, 2007.

[19] S. Nahm, K. Han, and W. Sung, "CORDIC-based digital quadrature mixer: comparison with a ROM-based architecture," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '98)*, pp. 385–388, June 1998.

[20] A. Madisetti, A. Y. Kwentus, and A. N. Willson, "100-MHz, 16-b, direct digital frequency synthesizer with a 100-dBc spurious-free dynamic range," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 8, pp. 1034–1043, 1999.

[21] A. Madisetti and A. Y. Kwentus, "Method and apparatus for direct digital frequency synthesizer," U.S. Patent 5 737 253, 1998.

[22] E. Grayver and B. Daneshrad, "Direct digital frequency synthesis using a modified CORDIC," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '98)*, vol. 5, pp. 241–244, June 1998.

[23] C. Y. Kang and E. E. Swartzlander Jr., "Digit-pipelined direct digital frequency synthesis based on differential CORDIC," *IEEE Transactions on Circuits and Systems I*, vol. 53, no. 5, pp. 1035–1044, 2006.

[24] T. Y. Sung and H. C. Hsin, "Design and simulation of reusable IP CORDIC core for special-purpose processors," *IET Computers and Digital Techniques*, vol. 1, no. 5, pp. 581–589, 2007.

[25] T. Y. Sung, L. T. Ko, and H. C. Hsin, "Low-power and high-SFDR direct digital frequency synthesizer based on hybrid CORDIC algorithm," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '09)*, pp. 249–252, May 2009.

[26] S. Sharma, P. N. Ravichandran, S. Kulkarni, M. Vanitha, and P. Lakshminarsimahan, "Implementation of Para-CORDIC algorithm and it's applications in satellite communication," in *Proceedings of the International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom '09)*, pp. 266–270, October 2009.

[27] D. De Caro, N. Petra, and A. G. M. Strollo, "Digital synthesizer/mixer with hybrid CORDIC—multiplier architecture: error analysis and optimization," *IEEE Transactions on Circuits and Systems I*, vol. 56, no. 2, pp. 364–373, 2009.

[28] J. Volder, "The CORDIC trigonometric computing technique," *IRE Transactions on Electronic Computers*, vol. 8, no. 3, pp. 330–334, 1959.

[29] J. S. Walther, "A unified algorithm for elementary functions," in *Proceedings of the Joint Computer Conference*, pp. 379–385, 1971.

[30] S. Chen, W. Huang, C. Cattani, and G. Altieri, "Traffic dynamics on complex networks: a survey," *Mathematical Problems in Engineering*, vol. 2012, Article ID 732698, 23 pages, 2012.

[31] W. Huang and S. Y. Chen, "Epidemic metapopulation model with traffic routing in scale-free networks," *Journal of Statistical Mechanics*, vol. 2011, no. 12, Article ID P12004, 19 pages, 2011.

[32] H. Shi, W. Wang, N. M. Kwok, and S. Y. Chen, "Game theory for wireless sensor networks: a survey," *Sensors*, vol. 12, no. 7, pp. 9055–9097, 2012.

[33] C. Cattani, S. Y. Chen, and G. Aldashev, "Information and modeling in complexity," *Mathematical Problems in Engineering*, vol. 2012, Article ID 868413, 4 pages, 2012.

[34] Y. H. Hu, "CORDIC-based VLSI architectures for digital signal processing," *IEEE Signal Processing Magazine*, vol. 9, no. 3, pp. 16–35, 1992.

[35] K. Maharatna, A. S. Dhar, and S. Banerjee, "A VLSI array architecture for realization of DFT, DHT, DCT and DST," *Signal Processing*, vol. 81, no. 9, pp. 1813–1822, 2001.

[36] T. Y. Sung, "Memory-efficient and high-speed split-radix FFT/IFFT processor based on pipelined CORDIC rotations," *IEE Proceedings*, vol. 153, no. 4, pp. 405–410, 2006.

[37] C. Cattani, "On the existence of wavelet symmetries in archaea DNA," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 673934, 16 pages, 2012.

[38] M. Li, "Approximating ideal filters by systems of fractional order," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 365054, 6 pages, 2012.

[39] S. Chen, Y. Zheng, C. Cattani, and W. Wang, "Modeling of biological intelligence for SCM system optimization," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 769702, 30 pages, 2012.

[40] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 2009, pp. 207–217, 2009.

[41] SMIMS Technology Corp., 2010, http://www.smims.com/.

[42] ModelSim—Simulation and debug, 2010, http://model.com/content/modelsim-pe-simulation-and-debug.

[43] Xilinx FPGA products, 2010, http://www.xilinx.com/products/.

[44] Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu City, Taiwan, and National Chip Implementation Center (CIC), National Science Council, Hsinchu City, Taiwan, TSMC 0.18 CMOS Design Libraries and Technical Data, v.1.6, 2010.

[45] Cadence Design Systems, 2010, http://www.cadence.com/products/pages/default.aspx.

[46] D. Prutchi and M. Norris, *Design and Development of Medical Electronic Instrumentation: A Practical Perspective of the Design, Construction, and Test of Medical Devices*, John Wiley & Sons, 2005.

[47] N. Li, J. Guo, H. S. Nie, W. Yi, H. J. Liu, and H. Xu, "Design of embedded bio-impedance analyzer based on digital auto balancing bridge method," *Applied Mechanics and Materials*, vol. 136, pp. 396–401, 2011.

[48] K. H. Lin, W. H. Chiu, and J. D. Tseng, "Low-complexity architecture of carrier frequency offset estimation and compensation for body area network systems," *Computer and Mathematics with Applications*, vol. 64, no. 5, pp. 1400–1408, 2012.

[49] J. Guo and P. Dong, "Design of dual phase signals generator based on AD9833," *Lecture in Electrical Engineering*, vol. 139, pp. 7–13, 2012.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2012, Article ID 648320, 7 pages
doi:10.1155/2012/648320

Research Article

A Rate-Distortion-Based Merging Algorithm for Compressed Image Segmentation

Ying-Shen Juang,[1] Hsi-Chin Hsin,[2] Tze-Yun Sung,[3] Yaw-Shih Shieh,[3] and Carlo Cattani[4]

[1] *Department of Business Administration, Chung Hua University, Hsinchu City 30012, Taiwan*
[2] *Department of Computer Science and Information Engineering, National United University, Miaoli 36003, Taiwan*
[3] *Department of Electronics Engineering, Chung Hua University, Hsinchu City 30012, Taiwan*
[4] *Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy*

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

Received 6 August 2012; Accepted 5 September 2012

Academic Editor: Sheng-yong Chen

Original images are often compressed for the communication applications. In order to avoid the burden of decompressing computations, it is thus desirable to segment images in the compressed domain directly. This paper presents a simple rate-distortion-based scheme to segment images in the JPEG2000 domain. It is based on a binary arithmetic code table used in the JPEG2000 standard, which is available at both encoder and decoder; thus, there is no need to transmit the segmentation result. Experimental results on the Berkeley image database show that the proposed algorithm is preferable in terms of the running time and the quantitative measures: probabilistic Rand index (PRI) and boundary displacement error (BDE).

1. Introduction

Data segmentation is important in many applications [1–6]. Early research work on image segmentation is mainly at a single scale, especially for medical images [7–9]. In the human visual system (HVS), the perceived image is decomposed into a set of band-pass subimages by means of filtering with simple visual cortical cells, which can be well modeled by Gabor filters with suitable spatial frequencies and orientations [10]. Other state-of-the-art multiscale techniques are based on wavelet transform (WT), which provides an efficient multiresolution representation in accord with the property of HVS [11]. Specifically, the higher-detail information of an image is projected onto a shorter basis function with higher spatial resolution. Various WT-based features and algorithms were proposed in the literature for image segmentation at multiple scales [12–14].

For the communication applications, original images are compressed in order to make good use of memory space and channel bandwidth. Thus, it is desirable to segment a compressed image directly. The Joint Photographic Expert Group (JPEG) standard adopts discrete cosine transform for subband image coding. In order to improve the compression performance of JPEG with more coding advantages, for example, embedded coding and progressive transmission, the JPEG2000 standard adopts WT as the underlying transform algorithm. Specifically, embedding coding is to code an image into a single code stream, from which the decoded image at any bit rate can be obtained. The embedded code stream of an image is organized in decreasing order of significance for progressive transmission over band-limited channels. This property is particularly desirable for the Internet streaming and database browsing applications [15–17]. Zargari proposed an efficient method for JPEG2000 image retrieval in the compressed domain [18]. Pi proposed a simple scheme to estimate the probability mass function (PMF) of wavelet subbands by counting the number of 1-bits and used the global PMF as features to retrieve similar images from a large database [19]. For image segmentation, however, the local PMF is needed. In [20], we proposed a simple method to compute the local PMF of wavelet coefficients based on the MQ table. It can be applied to a JPEG2000 code stream directly, and the local PMF can be used as features to segment a JPEG2000 image in the compressed domain.

Motivated by the idea behind the postcompression rate distortion (PCRD) algorithm [15], we propose a simple algorithm called the rate-distortion-based merging (RDM) algorithm for JPEG2000 image segmentation. It can be applied to a JPEG2000 code stream instead of the decoded image. As a result, the burden of decoding computation can be saved. In addition, the RDM algorithm is based on the MQ table, which is available at both encoder and decoder; thus, no overhead transmission is added from a segmentation viewpoint. The remainder of the paper proceeds as follows. In Section 2, the JPEG2000 standard is reviewed briefly. In Section 3, the MQ-table-based rate distortion slope (MQRDS) is proposed to examine the significance of wavelet segments; based on which, the RDM algorithm is thus proposed to merge wavelet segments with similar characteristics. Experimental results on the Berkeley color image database are given in Section 4. Conclusions can be found in Section 5.

2. Review of the JPEG2000 Standard

The core module of the JPEG2000 standard is the embedded block coding with optimized truncation (EBCOT) algorithm [15], which adopts wavelet transform (WT) as the underlying method to decompose an image into multiresolution subbands. WT has many desirable properties, for example, the self-similarity of wavelet coefficients across subbands of the same orientation, the joint space-spatial frequency localization with orientation selectivity, and the energy clustering within each subband [11]. The fundamental idea behind EBCOT is to take advantage of the energy clustering property of wavelet coefficients. EBCOT is a two-tier algorithm; tier-1 consists of bit plane coding (BPC) followed by arithmetic coding (AC); tier-2 is primarily for optimal rate control. Three coding passes, namely, the significance propagation (SP) pass, the magnitude refinement (MR) pass, and the clean-up (CU) pass, are involved with four primitive coding operations, namely, the significance coding operation, the sign coding operation, the magnitude refinement coding operation, and the clean-up coding operation. For a wavelet coefficient that is currently insignificant, if any of the 8 neighboring coefficients are already significant, it is coded in the SP pass using the significance coding operation; otherwise, it is coded in the CU pass using the clean-up coding operation. If this coefficient becomes significant, its sign is then coded using the sign coding operation. The magnitude of the significant wavelet coefficients that have been found in the previous coding passes is updated using the magnitude refinement coding operation in the MR pass. The resulting code streams of coding passes can be compressed further by using a context-based arithmetic coder known as the MQ coder. JPEG2000 defines 18 context labels for the MQ coder and stores their respective probability models in the MQ table. Specifically, 10 context labels are used for the significance coding operation and the clean-up coding operation, 5 context labels are used for the sign coding operation, and 3 context labels are used for the magnitude refinement coding operation.

In JPEG2000, a large image can be partitioned into nonoverlapped subimages called tiles for computational simplicity. WT is then applied to the tiles of an image for subband decompositions; and each wavelet subband is further divided into small blocks called code blocks. The code blocks of an image are independently coded from the most significant bit plane (MSB) to the least significant bit plane (LSB). Based on the true rate-distortion slope (RDS) of code blocks, JPEG2000 concatenates the significant code streams with large RDS using the post compression rate distortion (PCRD) algorithm for optimal rate control. More specifically, let $\{B_i\}$ be a set of code blocks in the whole image. The code stream of B_i can be terminated at the end of a coding pass, say n_i, with the bit rate denoted by $R_i^{n_i}$; all the end points of coding passes are possible truncation points. The distortion incurred by discarding the coding passes after n_i is denoted by $D_i^{n_i}$. PCRD selects the optimal truncation points to minimize the overall distortion: $D = \sum_i D_i^{n_i}$ subject to the rate constraint: $R = \sum_i R_i^{n_i} \le R_c$, where R_c is a given bitrate. It is noted that the coding passes with nonincreasing RDS are candidates for the optimal truncation points. Motivated by the idea of the above, a new technique is proposed to segment JPEG2000 images in the JPEG2000 domain; the detail is given in the following section.

3. Image Segmentation in the JPEG2000 Domain

This section presents a simple merging algorithm for JPEG2000 image segmentation. It merges wavelet segments with similar characteristics based on the change of the estimated RDS in the JPEG2000 domain. Thus, the proposed algorithm can be applied to a JPEG2000 code stream without decompressing complexity.

3.1. MQ Table-Based Probability Mass Function. In JPEG2000, the wavelet coefficients of an image are quantized with bit planes, and binary wavelet variables are almost independent across bit planes. The probability mass function (PMF) known as the wavelet histogram [19] can be approximated by

$$P(|c| = x) = \prod_{j=0}^{n-1} P_j\left(x_j\right),$$
$$x = \sum_{j=0}^{n-1} x_j \cdot 2^j; \quad x_j \in \{0, 1\}, \tag{1}$$

where x is the magnitude of a wavelet coefficient, c, $P_j(\circ)$ is the PMF of the binary wavelet variable, x_j, on the jth bit plane, and n is the number of bit planes. For image segmentation, the local PFM is needed. We had proposed a simple method to estimate the local PMF based on the MQ table [20]. Specifically, the probability of 1-bit, $P_j(x_j = 1)$, is given by

$$P_j\left(x_j = 1\right) = \begin{cases} Qe\_\text{Value} & \text{if MPS} = 0, \\ 1 - Qe\_\text{Value} & \text{if MPS} = 1, \end{cases} \tag{2}$$

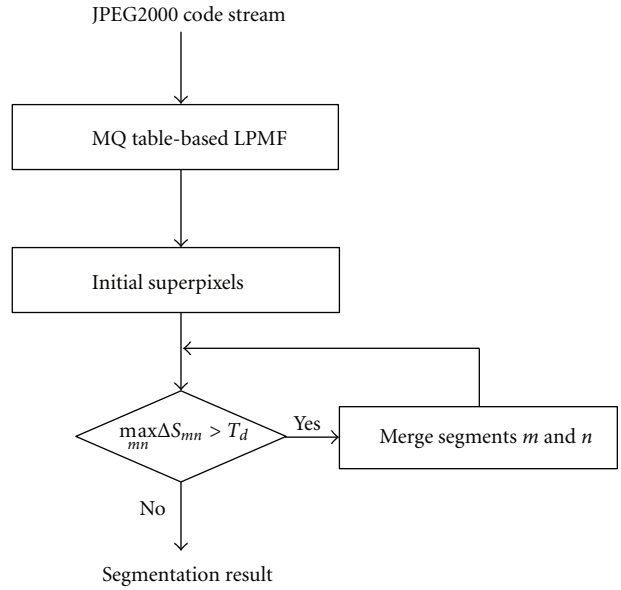


Figure 1: Flowchart of the RDM algorithm.

where $Qe_$Value is the probability of the less probable symbol (LPS), which is stored in the MQ table and MPS denotes the more probable symbol. The set $\{P_j(x_j = 1);\ j = 0, \ldots, n-1\}$ obtained from the MQ table can be used to compute the local PMF. As the MQ table is also available at decoder, no overhead transmission is needed for the computation of PMF. In addition, JPEG2000 defines only 18 context labels to model the binary wavelet variables; thus, the computation of PMF is simple.

3.2. MQ Table-Based Rate Distortion Slope and Merging Algorithm. Motivated by the post compression rate distortion (PCRD) algorithm [15], we propose the MQ table-based rate distortion slope (MQRDS) for image segmentation in the JPEG2000 domain as follows:

$$S_m = \frac{E[D_m]}{E[L_m]}, \tag{3}$$

where D_m is the distortion of wavelet segment: m defined as

$$D_m = \sum_{i=1}^{N_m} x_{m,i}^2, \tag{4}$$

$x_{m,i}$ is a wavelet coefficient at location: i in wavelet segment, m represented by

$$x_{m,i} = \sum_{j=0}^{n-1} x_{m,i,j} \cdot 2^j; \quad x_{m,i,j} \in \{0, 1\}. \tag{5}$$

The estimate of D_m can be computed by

$$\begin{aligned} E[D_m] &= \sum_{i=1}^{N_m} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} E\left[x_{m,i,j} \cdot x_{m,i,k}\right] \cdot 2^{j+k} \\ &\cong \sum_{i=1}^{N_m} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} E\left[x_{m,i,j}\right] \cdot E[x_{m,i,k}] \cdot 2^{j+k}, \end{aligned} \tag{6}$$

in which $E[x_{m,i,j}]$ can be obtained from the binary arithmetic code table known as the MQ table as follows:

$$E\left[x_{m,i,j}\right] = P_{m,i,j}\left(x_{m,i,j} = 1\right). \tag{7}$$

The estimate of code length $E[L_m]$ can be efficiently obtained by using [2]

$$E[L_m] = (D + N_m) \cdot E[R_m] - N_m \log_2\left(\frac{N_m}{N}\right) \tag{8}$$

$$E[R_m] = \sum_{j=0}^{n-1} H\left(x_{m,j}\right), \tag{9}$$

$$\begin{aligned} H\left(x_{m,j}\right) = &- P_{m,j}\left(x_{m,j} = 1\right) \cdot \log_2\left(P_{m,j}\left(x_{m,j} = 1\right)\right) \\ &- P_{m,j}\left(x_{m,j} = 0\right) \cdot \log_2\left(P_{m,j}\left(x_{m,j} = 0\right)\right), \end{aligned} \tag{10}$$

$$P_{m,j}\left(x_{m,j}\right) = \frac{1}{N_m} \sum_{i=1}^{N_m} \cdot P_{m,i,j}\left(x_{m,i,j}\right), \tag{11}$$

where j denotes the bit plane index, $x_{m,i,j}$ is the binary variable of $x_{m,i}$ on bit plane j, which are independent across bit planes, n is the number of bit planes, D is the feature space dimension, N_m is the number of wavelet coefficients in segment m, $N = \sum_{m=1}^{K} N_m$ is the total number of wavelet coefficients, and $H(\circ)$ is an entropy operation. After merging two wavelet segments, say m and n, the change of MQRDS is given by

$$\begin{aligned} &\Delta S_{mn} \\ &= \frac{[S_{mn} - ((N_m/(N_m + N_n))S_m + (N_n/(N_m + N_n))S_n)]}{S_{mn}}, \end{aligned} \tag{12}$$

where S_m and S_n are the MQRDS of wavelet segments, m and n, with sizes N_m and N_n, respectively, and S_{mn} is the MQRDS of the merged wavelet segment. As one can see, the change of MQRDS is likely to be increased significantly for wavelet segments with similar characteristics. Thus, we propose a simple algorithm called the rate-distortion-based merging (RDM) algorithm for JPEG2000 image segmentation, which is presented in the steps below.

The RDM Algorithm

Step 1. Given a JPEG2000 code stream, compute the MQ table-based local PMF of wavelet coefficients using (2).

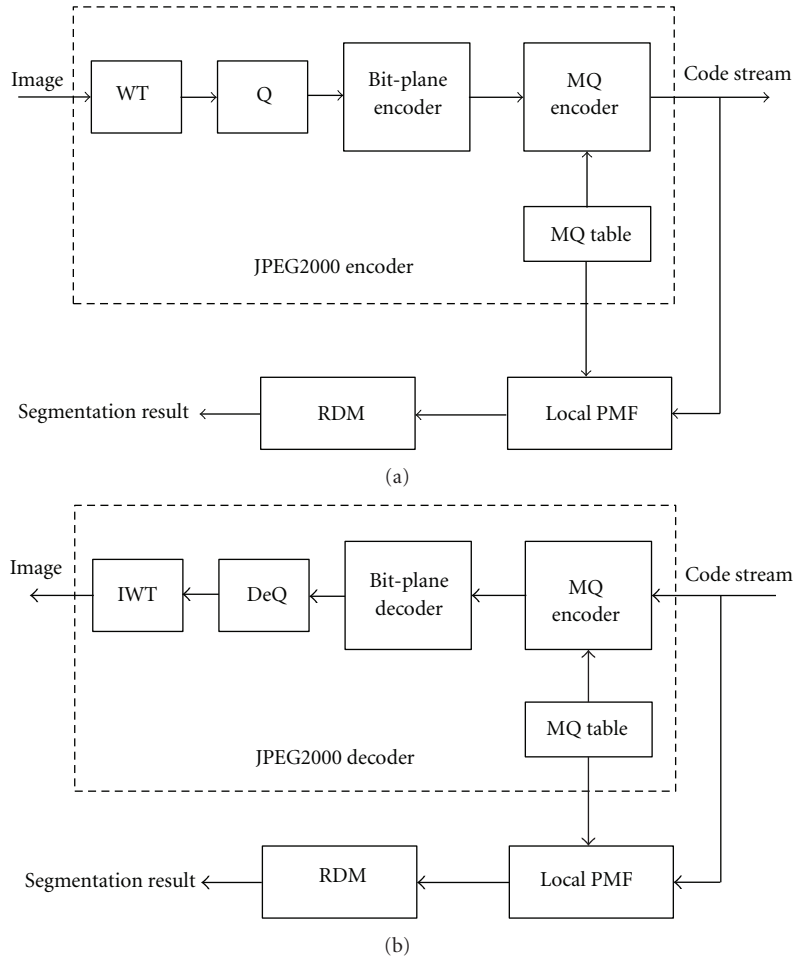


Figure 2: Image segmentation using RDM in the JPEG2000 domain; (a) encoder; (b) decoder.

Step 2. As mentioned in [2], a set of oversegmented regions known as superpixels is in general needed for any merging algorithms; this low-level initial segmentation can be obtained by coarsely clustering the local PMF as features.

Step 3. For all pairs of superpixels, compute their respective changes of MQRDS using (12), and merge the one with maximum change of MQRDS.

Step 4. Continue the merging process in step 3 until the change of MQRDS is insignificant.

In order to reduce the computation time, the following equation can be used to approximate (6):

$$E[D_m] \cong N_m \cdot \left[\sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \left(\frac{1}{N_m} \sum_{i=1}^{N_m} P_{m,i,j}\left(x_{m,i,j} = 1\right) \right) \cdot \left(\frac{1}{N_m} \sum_{i=1}^{N_m} P_{m,i,k}\left(x_{m,i,k} = 1\right) \right) \cdot 2^{j+k} \right]. \tag{13}$$

Moreover, the cross terms of the previous equation are not significant and can be discarded for computational simplicity. Figure 1 depicts flowchart of the RDM algorithm. It is noted that the MQ table defined in JPEG2000 is finite, thus (10) can be obtained by look-up table (LUT); this sure reduces the computation time further. As shown in Figure 2, RDM can be applied to a JPEG2000 code stream directly; this is one of the advantages of RDM.

4. Experimental Results

In the first experiment, the potential of the MQ table-based local PMF (LPMF) is shown by segmenting images with Brodatz textures. As noted, the essential characteristics of textures are mainly contained in the middle-high-frequency wavelet subbands; thus, we applied a simple clustering algorithm known as K-means to the LPMF of wavelet coefficients to generate an initial segmentation. The number of superpixels was set to 30, which was then finely merged using the RDM algorithms. Figure 3(a) shows the test image with two Brodatz textures, namely, wood and grass. The segmentation result and error image with white pixels representing misclassifications are shown in Figure 3(b) and Figure 3(c), respectively. Figure 3(d) shows the percentages of errors at various rates of bits per pixel (bpp). It is noted

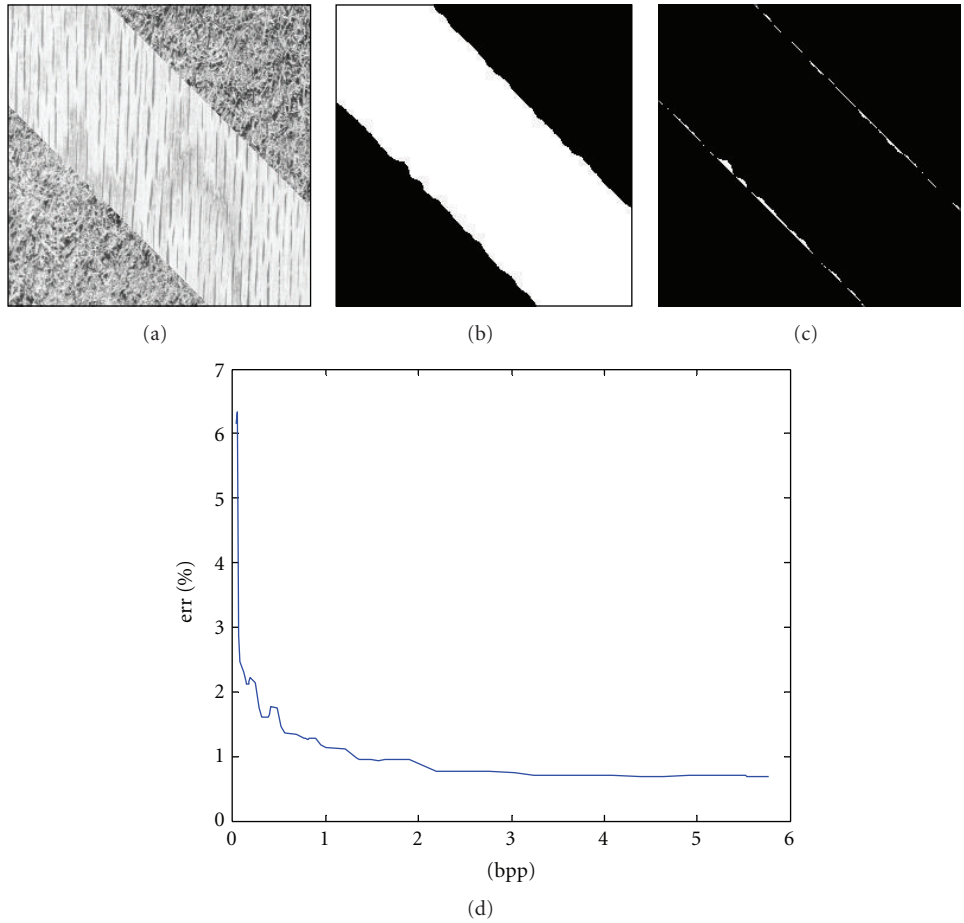


Figure 3: (a) Test image; (b) the segmentation result and (c) error image at 1 bpp; (d) error rates in percentage at various bpp rates.

that the segmentation results even at low-middle bpp rates are still satisfactory. Hence, a small portion of JPEG2000 code stream is sufficient for the segmentation task.

The RDM algorithm has also been extensively evaluated on the Berkeley image database [21]. We adopted the Waveseg algorithm [14] to compute the initial superpixels of a natural color image. In order to avoid decoding a JPEG2000 code stream, the Waveseg algorithm was applied to the estimated wavelet coefficients instead of the decoded wavelet coefficients. More specifically, the estimated wavelet coefficient of x_i using the MQ table-based LPMF is as follows.

$$\begin{aligned} E[x_i] &= \sum_{j=0}^{n-1} E\left[x_{i,j}\right] \cdot 2^j \\ &= \sum_{j=0}^{n-1} P_{i,j}\left(x_{i,j} = 1\right) \cdot 2^j, \end{aligned} \tag{14}$$

where $P_{i,j}(x_{i,j} = 1)$ is the probability of 1-bit on the jth bit plane, which can be obtained from the MQ table. The resulting superpixels were then merged by RDM with threshold, T_d, set to 0.1. We compared the RDM algorithm with two other state-of-the-art algorithms known as Mean-shift [22] and CTM [2]. In Mean-shift, the parameters, h_s and h_r, were set to 13 and 19, respectively; in CTM, the threshold γ was set to 0.1, as suggested in [2]. The original images shown at the top of Figure 4 are natural images contained in the Berkeley database, namely, Pyramids, Landscape, Horses, and Giraffes. Their respective segmentation results using RDM, CTM, and Mean-shift are shown in the second, third, and fourth rows. Visual inspection shows that RDM and Mean-shift have similar performances for the first three images; the performances of RDM and CTM are similar to detect the giraffes shown in the fourth image.

In addition to visual inspection [23, 24], two commonly used measures, namely, the probabilistic Rand index (PRI) and the boundary displacement error (BDE) [25], were adopted for quantitative comparisons. Table 1 gives the average PRI performance on the Berkeley database. PRI ranges from 0 to 1, and higher is better. BDE measures the average displacement error of boundaries between segmented images, which is nonnegative, and lower is better. The average BDE performance is given in Table 2. It is noted that RDM outperforms CTM and Mean-shift in terms of the PRI and BDE measures.

The running times on a PC are given in Table 3. It shows that RDM is faster than CTM and Mean-shift due largely to the simple computations of (8) and (13). Moreover, RDM

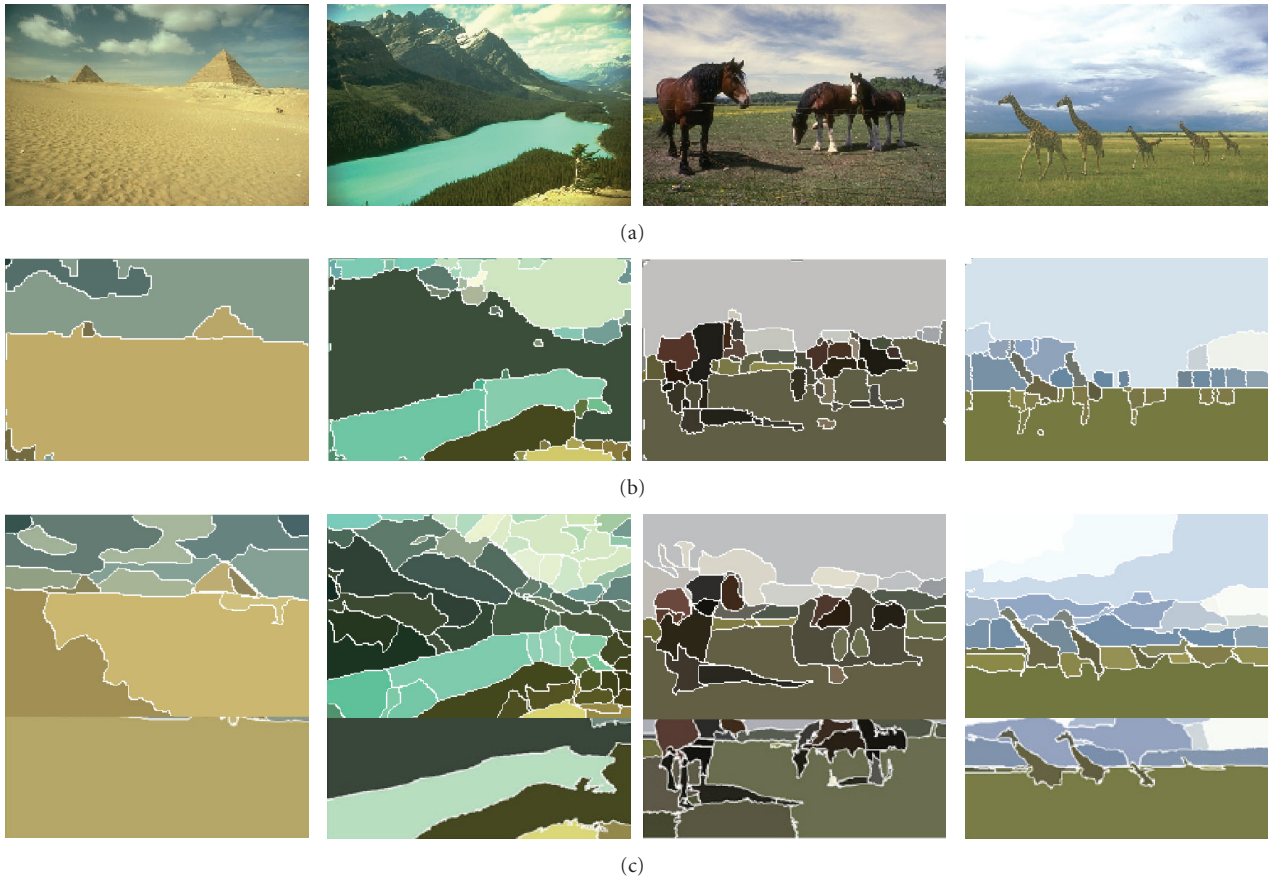


Figure 4: (a) Original images; (b) segmentation using RDM; (c) segmentation using CTM; (d) segmentation using Mean-shift.

Table 1: Average PRI on the Berkeley database.

RDM	CTM	Mean-shift
0.771	0.762	0.755

Table 2: Average BDE on the berkeley database.

RDM	CTM	Mean-shift
8.7	9.4	9.7

Table 3: Execution times.

	Pyramids	Landscape	Horses	Giraffes
RDM	8.9 s	8.7 s	10.7 s	6.8 s
Mean-shift	18.3 s	27.5 s	20.7 s	18.9 s
CTM	35.3 s	17.2 s	57.6 s	13.5 s

can be applied to a JPEG2000 code stream directly while most algorithms such as Mean-shift and CTM are primarily applied to the original or decoded image and it takes more time to decode a compressed image.

5. Conclusions

The MQ table defined in the JPEG2000 standard provides useful information that can be used to compute the local probability mass function (LPMF) of wavelet coefficients. A simple LPMF-based scheme has been proposed to estimate the rate distortion slope (RDS) of a wavelet segment. It is noted that the RDS is increased significantly after merging a pair of wavelet segments with similar characteristics into a single segment. Similar ideas of the above can be used to improve the rate control performance of JPEG2000 [26–28]. In this paper, we propose the rate-distortion-based merging (RDM) algorithm to segment images in the framework of JPEG2000. RDM has been evaluated on images with Brodatz textures and the Berkeley color image database. Experimental results show that the segmentation performance even at low-middle bpp rates is rather promising. For natural images with high-detail contents, RDM is preferable in terms of the average PRI and BDE measures. In addition, the total running time of RDM, which includes the computation of superpixels and the merging process, is faster than Mean-shift and CTM.

As RDM is based on the MQ table, which is available at both encoder and decoder, no overhead transmission is needed to compute the LPMF of wavelet coefficients. RDM can be applied to a JPEG2000 code stream directly; thus,

the burden of decompressing computation can be avoided, and memory space that is required to store the decompressed image is no longer necessary from the segmentation point of view.

Acknowledgments

The authors are grateful to the maintainers of the Berkeley image database. The National Science Council of Taiwan, under Grants NSC100-2628-E-239-002-MY2 and NSC100-2410-H-216-003, supported this work.

References

[1] Y. Xia, D. Feng, and R. Zhao, "Adaptive segmentation of textured images by using the coupled Markov random field Model," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3559–3566, 2006.

[2] A. Y. Yang, J. Wright, Y. Ma, and S. Shankar Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 212–225, 2008.

[3] N. A. M. Isa, S. A. Salamah, and U. K. Ngah, "Adaptive fuzzy moving K-means clustering algorithm for image segmentation," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 4, pp. 2145–2153, 2009.

[4] S. Xiang, C. Pan, F. Nie, and C. Zhang, "Turbopixel segmentation using eigen-images," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 3024–3034, 2010.

[5] M. Li and W. Zhao, "Quantitatively investigating locally weak stationarity of modified multifractional gaussian noise," *Physica A*, vol. 391, no. 24, pp. 6268–6278, 2012.

[6] M. Li and W. Zhao, "Variance bound of ACF estimation of one block of fGn with LRD," *Mathematical Problems in Engineering*, vol. 2010, Article ID 560429, 14 pages, 2010.

[7] S. Chen and X. Li, "Functional magnetic resonance imaging for imaging neural activity in the human brain: the annual progress," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 613465, 9 pages, 2012.

[8] Z. Teng, J. He, A. J. Degnan et al., "Critical mechanical conditions around neovessels in carotid atherosclerotic plaque may promote intraplaque hemorrhage," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 223, no. 2, pp. 321–326, 2012.

[9] 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.

[10] D. E. Ilea and P. F. Whelan, "CTex—an adaptive unsupervised segmentation algorithm based on color-texture coherence," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1926–1939, 2008.

[11] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, Calif, USA, 1999.

[12] M. K. Bashar, N. Ohnishi, and K. Agusa, "A new texture representation approach based on local feature saliency," *Pattern Recognition and Image Analysis*, vol. 17, no. 1, pp. 11–24, 2007.

[13] C. M. Pun and M. C. Lee, "Extraction of shift invariant wavelet features for classification of images with different sizes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1228–1233, 2004.

[14] C. R. Jung, "Unsupervised multiscale segmentation of color images," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 523–533, 2007.

[15] T. Acharya and P. S. Tsai, *JPEG2000 Standard for Image Compression: Concepts, Algorithms and VLSI Architectures*, John Wiley & Sons, New York, NY, USA, 2005.

[16] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.

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

[18] F. Zargari, A. Mosleh, and M. Ghanbari, "A fast and efficient compressed domain JPEG2000 image retrieval method," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1886–1893, 2008.

[19] M. H. Pi, C. S. Tong, S. K. Choy, and H. Zhang, "A fast and effective model for wavelet subband histograms and its application in texture image retrieval," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3078–3088, 2006.

[20] H. C. Hsin, "Texture segmentation in the joint photographic expert group 2000 domain," *IET Image Processing*, vol. 5, no. 6, pp. 554–559, 2011.

[21] http://www.eecs.berkeley.edu/~yang/software/lossy_segmentation/.

[22] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[23] H. C. Hsin, T.-Y. Sung, Y.-S. Shieh, and C. Cattani, "MQ Coder based image feature and segmentation in the compressed domain," *Mathematical Problems in Engineering*, vol. 2012, Article ID 490840, 14 pages, 2012.

[24] S. Chen, M. Zhao, G. Wu, C. Yao, and J. Zhang, "Recent advances in morphological cell image analysis," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 101536, 10 pages, 2012.

[25] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 929–944, 2007.

[26] H. C. Hsin and T. Y. Sung, "Context-based rate distortion estimation and its application to wavelet image coding," *WSEAS Transactions on Information Science and Applications*, vol. 6, no. 6, pp. 988–993, 2009.

[27] H.-C. Hsin and T.-Y. Sung, "Image segmentation in the JPEG2000 domain," in *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR'11)*, pp. 24–28, 2011.

[28] H.-C. Hsin, T.-Y. Sung, Y.-S. Shieh, and C. Cattani, "Adaptive binary arithmetic coder-based image feature and segmentation in the compressed domain," *Mathematical Problems in Engineering*, vol. 2012, Article ID 490840, 14 pages, 2012.