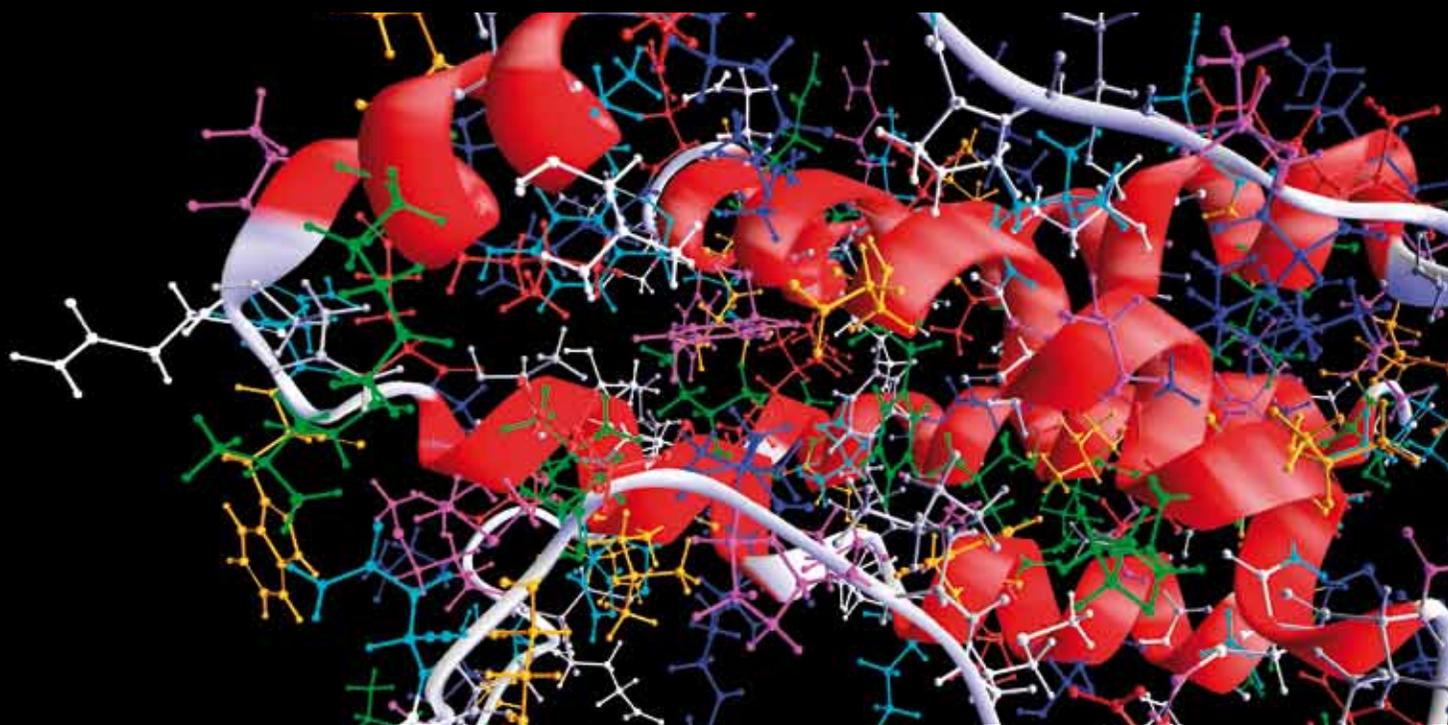


BIOMEDICAL SIGNAL PROCESSING AND MODELING COMPLEXITY OF LIVING SYSTEMS

GUEST EDITORS: CARLO CATTANI, RADU BADEA, SHENGYONG CHEN, AND MARIA CRISAN





Biomedical Signal Processing and Modeling Complexity of Living Systems

Computational and Mathematical Methods in Medicine

Biomedical Signal Processing and Modeling Complexity of Living Systems

Guest Editors: Carlo Cattani, Radu Badea, Shengyong Chen,
and Maria Crisan



Copyright © 2012 Hindawi Publishing Corporation. All rights reserved.

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

Editorial Board

Zvia Agur, Israel
Emil Alexov, USA
Gary C. An, USA
Georgios Archontis, Cyprus
Pascal Au nger, France
Facundo Ballester, Spain
Dimos Baltas, Germany
Chris Bauch, Canada
Maxim Bazhenov, USA
Niko Beerenwinkel, Switzerland
Philip Biggin, UK
Michael Breakspear, Australia
Thierry Busso, France
Carlo Cattani, Italy
Bill Crum, UK
Timothy David, New Zealand
Gustavo Deco, Spain
Carmen Domene, UK
Wim Van Drongelen, USA
Frank Emmert-Streib, UK
Ricardo Femat, Mexico
Alfonso T. García-Sosa, Estonia
Kannan Gunasekaran, USA

Damien R. Hall, Japan
William F. Harris, South Africa
Vassily Hatzimanikatis, USA
Volkhard Helms, Germany
J.-H. S. Hofmeyr, South Africa
Seiya Imoto, Japan
Bleddyn Jones, UK
Lawrence A. Kelley, UK
Lev Klebanov, Czech Republic
Ina Koch, Germany
David Liley, Australia
Quan Long, UK
Yoram Louzoun, Israel
Jianpeng Ma, USA
C.-M. C. Ma, USA
Reinoud Maex, France
Francois Major, Canada
Simeone Marino, USA
Ali Masoudi-Nejad, Iran
Ramit Mehr, Israel
Seth Michelson, USA
Michele Migliore, Italy
Karol Miller, Australia

Ernst Niebur, USA
Kazuhisa Nishizawa, Japan
Martin Nowak, USA
Markus Owen, UK
Hugo Palmans, UK
Lech S. Papiez, USA
Jean Pierre Rospars, France
David James Sherman, France
Sivabal Sivaloganathan, Canada
Stephen W. Smye, UK
Merryn H. Tawhai, New Zealand
Elisabeth Tillier, Canada
Nestor V. Torres, Spain
Anna Tramontano, Italy
Nelson J. Trujillo-Barreto, Cuba
Kutlu O. Ulgen, Turkey
Nagarajan Vaidehi, USA
Edelmira Valero, Spain
Jinliang Wang, UK
Jacek Waniewski, Poland
Guang Wu, China
Henggui Zhang, UK

Contents

Biomedical Signal Processing and Modeling Complexity of Living Systems, Carlo Cattani, Radu Badea, Shengyong Chen, and Maria Crisan
Volume 2012, Article ID 298634, 2 pages

A Fuzzy Model to Interpret Data of Drive Performances from Patients with Sleep Deprivation, Pasquale Sena, Paolo Attianese, Francesca Carbone, Arcangelo Pellegrino, Aldo Pinto, and Francesco Villecco
Volume 2012, Article ID 868410, 5 pages

Theoretical Compartment Modeling of DCE-MRI Data Based on the Transport across Physiological Barriers in the Brain, Laura Fanea, Leontin I. David, Andrei Lebovici, Francesca Carbone, and Silviu A. Sfrangeu
Volume 2012, Article ID 482565, 6 pages

Theorems and Application of Local Activity of CNN with Five State Variables and One Port, Gang Xiong, Xisong Dong, Li Xie, and Thomas Yang
Volume 2012, Article ID 674243, 13 pages

Wavelet Transform Fuzzy Algorithms for Dermoscopic Image Segmentation, Heydy Castillejos, Volodymyr Ponomaryov, Luis Nino-de-Rivera, and Victor Golikov
Volume 2012, Article ID 578721, 11 pages

Software Tool for the Prosthetic Foot Modeling and Stiffness Optimization, Matija Štrbac and Dejan B. Popovic
Volume 2012, Article ID 421796, 8 pages

On the Existence of Wavelet Symmetries in Archaea DNA, Carlo Cattani
Volume 2012, Article ID 673934, 21 pages

The Stochastic Evolution of a Protocell: The Gillespie Algorithm in a Dynamically Varying Volume, T. Carletti and A. Filisetti
Volume 2012, Article ID 423627, 13 pages

Detecting Epileptic Seizure from Scalp EEG Using Lyapunov Spectrum, Truong Quang Dang Khoa, Nguyen Thi Minh Huong, and Vo Van Toi
Volume 2012, Article ID 847686, 11 pages

Iterative Methods for Obtaining Energy-Minimizing Parametric Snakes with Applications to Medical Imaging, Alexandru Ioan Mitrea, Radu Badea, Delia Mitrea, Sergiu Nedevschi, Paulina Mitrea, Dumitru Mircea Ivan, and Octavian Mircia Gurzău
Volume 2012, Article ID 918510, 11 pages

Multicriteria Optimization Model for the Study of the Efficacy of Skin Antiaging Therapy, Maria Crișan, Guido Cappare, Luciana Neamțiu, Ioana Chiorean, Liana Lușța, Diana Crișan, and Radu Badea
Volume 2012, Article ID 103919, 6 pages

Nonlocal Means-Based Denoising for Medical Images, Ke Lu, Ning He, and Liang Li
Volume 2012, Article ID 438617, 7 pages

Solid Dynamic Models for Analysis of Stress and Strain in Human Hearts, Qiu Guan, Miaomiao Lu, Xiaoyan Wang, and Cao Jiang
Volume 2012, Article ID 634534, 6 pages

Characteristics of Evoked Potential Multiple EEG Recordings in Patients with Chronic Pain by Means of Parallel Factor Analysis, Juan Wang, Xiaoli Li, Chengbiao Lu, Logan J. Voss, John P. M. Barnard, and Jamie W. Sleight
Volume 2012, Article ID 279560, 10 pages

A Batch Rival Penalized Expectation-Maximization Algorithm for Gaussian Mixture Clustering with Automatic Model Selection, Jiechang Wen, Dan Zhang, Yiu-ming Cheung, Hailin Liu, and Xinge You
Volume 2012, Article ID 425730, 11 pages

The Finite Element Method Applied to a Problem of Blood Flow in Vessels, Gabriela Nuț, Ioana Chiorean, and Maria Crișan
Volume 2012, Article ID 204926, 6 pages

Functional Magnetic Resonance Imaging for Imaging Neural Activity in the Human Brain: The Annual Progress, Shengyong Chen and Xiaoli Li
Volume 2012, Article ID 613465, 9 pages

Abdominal Tumor Characterization and Recognition Using Superior-Order Cooccurrence Matrices, Based on Ultrasound Images, Delia Mitrea, Paulina Mitrea, Sergiu Nedevschi, Radu Badea, Monica Lupșor, Mihai Socaciu, Adela Golea, Claudia Hagi, and Lidia Ciobanu
Volume 2012, Article ID 348135, 17 pages

Approximating Ideal Filters by Systems of Fractional Order, Ming Li
Volume 2012, Article ID 365054, 6 pages

High-Resolution Remotely Sensed Small Target Detection by Imitating Fly Visual Perception Mechanism, Fengchen Huang, Lizhong Xu, Min Li, and Min Tang
Volume 2012, Article ID 789429, 9 pages

Separable Transition Density in the Hybrid Model for Tumor-Immune System Competition, Carlo Cattani and Armando Ciancio
Volume 2012, Article ID 610124, 6 pages

Recent Advances in Morphological Cell Image Analysis, Shengyong Chen, Mingzhu Zhao, Guang Wu, Chunyan Yao, and Jianwei Zhang
Volume 2012, Article ID 101536, 10 pages

Remotely Sensed Image Classification by Complex Network Eigenvalue and Connected Degree,
Mengxi Xu and Chenglin Wei
Volume 2012, Article ID 632703, 9 pages

Influence of Expert-Dependent Variability over the Performance of Noninvasive Fibrosis Assessment in Patients with Chronic Hepatitis C by Means of Texture Analysis, Cristian Vicas, Monica Lupsor, Mihai Socaciu, Sergiu Nedevschi, and Radu Badea
Volume 2012, Article ID 346713, 9 pages

Nested Quantization Index Modulation for Reversible Watermarking and Its Application to Healthcare Information Management Systems, Lu-Ting Ko, Jwu-E. Chen, Yaw-Shih Shieh, Hsi-Chin Hsin, and Tze-Yun Sung
Volume 2012, Article ID 839161, 8 pages

Modeling of Biological Intelligence for SCM System Optimization, Shengyong Chen, Yujun Zheng, Carlo Cattani, and Wanliang Wang
Volume 2012, Article ID 769702, 10 pages

Motion Analysis of Live Objects by Super-Resolution Fluorescence Microscopy, Chunyan Yao, Jianwei Zhang, Guang Wu, and Houxiang Zhang
Volume 2012, Article ID 859398, 8 pages

Modeling and Representation of Human Hearts for Volumetric Measurement, Qiu Guan, Wanliang Wang, and Guang Wu
Volume 2012, Article ID 389463, 6 pages

Medicoeconomic Index for Photo-Induced Skin Cancers, Ioana Chiorean, Liana Lupşa, Luciana Neamţiu, and Maria Crişan
Volume 2012, Article ID 414683, 3 pages

Target Contour Recovering for Tracking People in Complex Environments, Jianhua Zhang, Sheng Liu, Y. F. Li, and Jianwei Zhang
Volume 2012, Article ID 506908, 10 pages

An Adaptive Coding Pass Scanning Algorithm for Optimal Rate Control in Biomedical Images, Hsi-Chin Hsin, Tze-Yun Sung, and Yaw-Shih Shieh
Volume 2012, Article ID 935914, 7 pages

Asymptotic Identity in Min-Plus Algebra: A Report on CPNS, Ming Li and Wei Zhao
Volume 2012, Article ID 154038, 11 pages

Editorial

Biomedical Signal Processing and Modeling Complexity of Living Systems

Carlo Cattani,¹ Radu Badea,² Shengyong Chen,³ and Maria Crisan⁴

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

²Department of Clinical Imaging Ultrasound, "Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca, Romania

³College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

⁴Department of Histology, "Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca, Romania

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

Received 22 November 2012; Accepted 22 November 2012

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

Living systems are often maintained by information flows and, as such, they present interesting mathematical problems for example, in the modeling and analysis of spatial structures, self-organization, environmental interaction, behavior, and development. Biomedical signals extract information from the complex phenomena being measured, which are typically a time series having both a regular and random components. Solutions attempt to map general principles, which are used to model how the living systems work. Many researchers have been studying these problems because of their interesting mathematical features and because of their scientific importance. The focus of this special issue is the mathematical analysis and modeling of time series in living systems and biomedical signals. It is mostly interested in the related new development of both theoretical study and practical implementation, either with modeling, complexity, statistics, or signal transformation in living systems.

The papers selected for this special issue represent a good panel in recent challenges. The topics of the research papers and review papers are connected with the living systems and biomedical signals, including modeling dynamical complexity in living systems, for example, network dynamics, mass action, dynamical systems theory, methods for analysis and characterization of dynamical complexity, biomedical signal analysis such as mathematical pattern analysis of biological signals, generative mechanisms of biological signal patterning, implementation of signal analysis algorithms, linking biological structure to biological signal generation, and intracellular signal processing, as well as related models and applications, such as systems theory, biological organization, and biomedical information processing.

This special issue contains 31 papers. In the category of modeling dynamical complexity, L. Sena et al. present a fuzzy model to interpret data of drive performances from patients with sleep deprivation. S. Chen et al. review the modeling of biological intelligence for supply chain management system optimization. M. Li presents approximating ideal filters by systems of fractional order. L. T. Ko et al. present nested quantization index modulation for reversible watermarking and its application to healthcare information management systems. C. Cattani studies on the existence of wavelet symmetries in Archaea DNA. A. Ciancio and C. Cattani present separable transition density in the hybrid model for tumor-immune competition. Q. Guan et al. present solid dynamic models for analysis of stress and strain in human hearts. G. Xiong et al. present theorems and application of local activity of CNN with five state variables and one port. L. Fanea et al. present theoretical compartment modeling of DCE-MRI data based on the transport across physiological barriers in the brain. M. Crisan et al. present a multicriteria optimization model for the study of the efficacy of skin antiaging therapy.

In the category of methods for analysis of dynamical complexity, L. Xu et al. present high resolution remotely sensed small target detection by imitating fly visual perception mechanism. J. Zhang et al. present target contour recovering for tracking people in complex environments. M. Li and W. Zhao present a report on CPNs for asymptotic identity in min-plus algebra. M. Xu and C. Wei present remotely sensed image classification by complex network eigenvalue and connected-degree. J. Wen et al. present a batch rival penalized expectation-maximization algorithm

for Gaussian mixture clustering with automatic model selection. M. Štrbac and D. Popovic show a software tool for the prosthetic foot modeling and stiffness optimization. T. Carletti and A. Filisetti present the stochastic evolution of a protocell: the Gillespie algorithm in a dynamically varying volume.

In the category of biomedical signal analysis, S. Chen and X. Li review the annual progress of functional magnetic resonance imaging for imaging neural activity in the human brain. H. C. Hsin et al. present an adaptive coding pass scanning algorithm for optimal rate control in biomedical images. I. Chiorean et al. present a medicoeconomic index for photo-induced skin cancers. C. Vicas et al. present the influence of expert dependent variability over the performance of noninvasive fibrosis assessment at patients with chronic hepatitis C, by the means of texture analysis. G. Nut et al. present a finite element method applied to a problem of blood flow in vessels. S. Chen et al. summarize recent advances in morphological cell image analysis. C. Yao et al. present motion analysis of live objects by super-resolution fluorescence microscopy. H. Castillejos et al. present wavelet transform fuzzy algorithms for dermoscopic image segmentation. D. A. Mitrea et al. present abdominal tumor characterization and recognition using superior order cooccurrence matrices, based on ultrasound images. Q. Guan et al. present a method for modeling and representation of human hearts for volumetric measurement. X. Li et al. present characteristics of evoked potential multiple EEG recordings in patients with chronic pain by means of parallel factor analysis. A. I. Mitrea et al. present iterative methods for obtaining energy-minimizing parametric snakes, with application to medical imaging. K. Lu et al. present nonlocal means based denoising for medical images. K. T. Q. Dang et al. present detecting epileptic seizure from scalp EEG using Lyapunov spectrum.

Of 67 submissions, 31 papers are selected in this special issue. Of course, the topics and papers are not an exhaustive representation of the area of biomedical signal processing and modeling complexity of living systems. It can be seen that although some solutions and models become available, most problems remain open and research is highly active in this field. In the near future, we expect more contributions that will address all of the key aspects raised above. Nonetheless, the special issue represents the recent concerns in the community and we have the pleasure of sharing them with the readers.

Acknowledgments

We would like to thank the authors for their excellent contributions and the reviewers for their help in improving the papers.

*Carlo Cattani
Radu Badea
Shengyong Chen
Maria Crisan*

Research Article

A Fuzzy Model to Interpret Data of Drive Performances from Patients with Sleep Deprivation

Pasquale Sena,¹ Paolo Attianese,¹ Francesca Carbone,¹ Arcangelo Pellegrino,² Aldo Pinto,¹ and Francesco Villecco²

¹Department of Pharmaceutical and Biomedical Sciences, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy

²Department of Industrial Engineering, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy

Correspondence should be addressed to Francesco Villecco, fvillecco@unisa.it

Received 30 September 2011; Revised 27 April 2012; Accepted 28 April 2012

Academic Editor: Carlo Cattani

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

The search for safe vehicles is increasing with both diffusion of high traffic density over the world and availability of new technologies providing sophisticated tools previously impossible to realize. Design and development of the necessary devices may be based on simulation tests that reduce cost allowing trials in many directions. A proper choice of the arrangement of the drive simulators, as much as of the parameters to be monitored, is of basic importance as they can address the design of devices somehow responsible for the drivers safety or, even their lives. This system setup, consisting of a free car simulator equipped with a monitoring system, collects in a nonintrusive way data of the car lateral position within the road lane and of its first derivative. Based on these measured parameters, the system is able to detect symptoms of drowsiness and sleepiness. The analysis is realized by a fuzzy inferential process that provides an immediate warning signal as soon as drowsiness is detected with a high level of certainty. Enhancement of reliability and minimisation of the false alarm rate are obtained by operating continuous comparison between learned driver typical modalities of operation on the control command of the vehicle the pattern recorded.

1. Introduction

Development of new systems aimed at increasing the equipments for active safety of motor vehicles is draining more and more attention in both academies and industries, involving scientists from many fields ranging from engineering to medicine. Actually, attention is focused on systems or devices able to account for all possible risks either deriving from an unfavourable external environment or from improper drivers psychophysical conditions. These latter are the real challenge to win, as the current approach possible is to intervene as soon as there is evidence of any driver difficulty that, however, often comes out when safety is already compromised.

Microsleeps are among the most critical occurrences. Data from the National Statistics Institute (ISTAT) on this subject (as of 20.11.2008 for the year 2007) say that accidents due to physical or mental conditions of the driver correspond to approximately 3.8% of the total. Among these, the fraction of accidents officially related to classic microsleeps is 0.26%

[1]. However, the analysis is misleading as it underestimates the real effect of sleepiness.

It is a matter of fact that the real nature of traffic accidents is multifactorial. This hinders a correct evaluation of the contribution of sleepiness as much as of microsleeps. When considering the role of sleepiness, this is often masked from other more evident factors such as excessive speed, or weather conditions or state of the vehicle, becoming thus a cofactor whereas somnolence should be considered the final common pathway of the other factors listed above.

To date, this interaction is not deeply studied although recent studies outline, with increasing relevance, characteristics that can be considered typical of situations with excessive sleepiness, and point out that drowsiness may be, in fact, an important risk factor in the occurrence of accidents or may contribute, even to a less extent, to the increase of the mortality index in car accidents [2–5]. It has to be noted that the percentage of traffic accidents attributed to sleepiness considerably vary from country to country, due to differences in instruments and methods of data collection

used by the local authorities. Other factors of variability may lead to the underestimation of the phenomenon, such as type of population studied, different definitions of sleepiness, objective difficulty to assess the degree of sleepiness of the driver at the accident site. Actually, most of the available data come from inferential evidence.

Accidents caused by microsleeps have some peculiarities, that may help in a classification [6]. In fact, these accidents mainly occur at night time, early morning, or middle afternoon and, more frequently, during monotonous driving conditions such as in highways or when the driver is alone in the vehicle. All these accidents show similar dynamics, with a single vehicle involved leaving the road regardless of weather or traffic conditions, without any trial of the driver to avoid the accident.

Thus, microsleeps just are the emerging tip of a progressive phenomenon characterized by a gradual reduction in the level of supervision, which in turn induces a continuous decrease in psychomotor performances and, consequently a reduction of reaction times, a diminished risk perception and a decrease of attention which may have a value in determining the occurrence of a traffic accident. This specially occurs when driving is monotonous and repetitive, favouring the emergence of a latent sleepiness which is often not perceived by the driver.

This partial drowsiness, can be also induced or facilitated by a variety of abnormal conditions, dysfunctions and diseases, either common or rather widespread in the general population. Pregnancy, jetlag, excessive exercise, or intake of abuse of psychoactive drugs such as alcohol and marijuana, can affect alertness causing excessive sleepiness, impaired reaction times, decreased ability to recognize environmental signals, or impaired judgment. The same happens with sleep disorders, such as insomnia, restless legs syndrome, or the sleep apnea. This deterioration of driving skills can also be promoted by specific pathological conditions such as hypothyroidism, migraine, depression, and certain infectious diseases, such as the common flu.

In addition to drugs that act specifically on the brain (sedatives, hypnotics, antidepressants, and neuroleptics), some antihistamines, including those of more recent introduction, can have side effects like sedation but different individual responses. The drowsiness has also been reported with drugs widely used as the nonsteroidal anti-inflammatory, for example, diclofenac or indomethacin, and antihypertensives, especially beta blockers.

All these conditions leading to a decrease of attention and concentration during driving and reduced ability in evaluating the risks (e.g., from excessive speed) confirm the multifactorial nature as much as the complexity of the system driver-vehicle environment and outlines the benefits driving safety that could derive from an approach addressed to evaluate the physiopathological conditions of the driver rather than to wait for his failure.

2. The Need for Prevention

As above noted, the strategy of many automobile industries for active prevention of accidents from sleepiness is to install

inside the vehicle purposely designed devices able to verify the state of vigilance of the driver and to warn him in case of risks.

People driving with excessive sleepiness make great efforts to stay awoken so that the phases of the sleep onset at drive are not comparable to those at home. In fact, the physiological events preceding the sleep show different durations and sequences in the two cases, making data currently collected hard to interpretate.

There is evidence that others parameters could be considered, as frequency and amplitude of the sudden turns of the steering wheel, as they appear able to offer a good indication of the level of sleepiness, although changes in the vehicle direction can sometimes depend on ability and experience of the driver. Measure of frequency of blinking seems also to be promising in terms of effectiveness and reliability.

A risk holds when using the parameters above to define a safety threshold: the related devices could become themselves a source of further risk, as a driver who feels like falling asleep might decide to keep on driving, relying on the safety devices intervention.

Furthermore, most of the approaches currently used in the development and design of driver monitoring devices for detection of conditions of drowsiness are based on the identification of the “dozing” or of a condition of marked drowsiness so that intervention are possible only at the end of the progressive decay phase which, in fact, precedes the onset of sleep.

This paper aims at pointing out the possibility of developing system based on the idea that the progressive decaying of drivers ability to appropriately react to external stimuli rather than its final effect should be monitored, making thus more effective any safety device that in this way could evidence the risk situation largely in time.

3. Analysis of the Multifactorial System

Driving a vehicle, is a very complex task that requires a high degree of the neuromotor coordination and the development of a number of appropriate, well-balanced abilities.

The driving ability is primarily attributable to a strong propensity of the driver to develop proper automate mechanisms of control of the vehicle and to establish a continuous monitoring through a number of proprioceptive feedbacks. Driving is thus a complex multifactorial and multiparametric phenomenon that obeys to the principles of neurophysiology, which state that the central nervous system (CNS) does not deal with single muscles, but with a general coordinated movement. Thus, the simplest motor response (e.g., the activation of arts to control the vehicle) actually corresponds to a complex motor act. Moreover, activation and space time organization of the motor response to control steering wheel, pedals, and gear shift depend on the continuous flux of exteroceptive information: visual, auditory, haptic, and vestibular. Driving can thus be defined a built-in complex motor acts, that activates integrate systems and central and affective-emotional components, through peripheral proprioceptive afferents. Any alteration of this complex system due to an inefficiency in a few simple

districts, will affect the final result, that is, the safe control of the vehicle.

Following the considerations above, any system capable of nonintrusively monitoring the performance of the driver will need a detailed model of the vehicle-driver couple.

In particular, the vehicle knowledge has to be associated to the knowledge of the environment the vehicle operates in, being the latter the most difficult part. It is enough to consider, for example, the uncertainty introduced by the presence of other vehicles that move and can suddenly change direction to outline the difficulty of finding probabilistic mathematical functions allowing an easy formulation of the model. These difficulties are also caused by the vehicle interaction with the environment, as is the case of the vehicle dynamics changing for a road surface wet, or of the measurements that may vary with light and geometric or physical properties of the objects.

Therefore, the knowledge basis to define the model must necessarily rely on a simultaneous analysis of all the different parameters, involving both the driver-and the vehicle-related.

However there is a lack of standardization of the measured variables that makes hard even to decide the parameter to be monitored.

This system should be based on a series of continuous measurements of variables serving as indicators of the driver performance (e.g., steering wheel movements), and on their processing to decide the permission to drive.

4. The Experimental Setup

A free car simulator for noncommercial use named Racer (Dolphinity) running on a gaming PC was used to collect data of driving performances. The simulator offers to driver a scenery with a ground level first person perspective view of a highway with no other cars.

The “quick race” mode of the simulator was used, permitting logging of all data of interest. Subjects moved through the environment in a FIAT Punto GT, controlled using a compatible game controller (Thrustmaster t500 rs) consisted of a steering wheel of 30 cm of diameter, 3 pedals equipped with digital encoders, and a gear shift with H pattern.

A modification was employed which ensured that data regarding vehicle parameters and action of the driver on the control commands were continuously acquired and stored.

A separate PC, linked to the PC running the physical engine of the simulator, read in real time the data logged and processed the data to work out the final score.

5. Development of an Experimental Monitoring Device

The idea is to develop an on-board detection system capable of monitoring the driver and formulate an algorithm able to analyze data and decide when the driving ability is materially impaired by drowsiness.

The ideal monitoring system should integrate specific measures of the drive performance with direct possibly

noninvasive measurements of the psycho-physiological state of the driver together with measures of reaction time of secondary tasks could be added to evaluate the status of the driver decide consent or denial to drive.

In fact, to an early detection of the condition of drowsiness, the variables must be analyzed in real time and in a combined way because no individual measure appears to be sufficient for a reliable detection of somnolence.

The choice of variables in this work has been made on the basing on particular considerations. The monitoring variable to evaluate the driven condition must necessarily be easy to measure, of practical application, and of an adequate sensitivity. They must also be repeatable and must be able to describe the process on the basis of objective observable parameters. Therefore, a system has been developed that during a simulated driving session records, by a logging device data of the relative position of the car center line and of the lane edge, and compares the driver decisions to the ideal trajectories and options by a controller. Parameters such as the movement of the pedals and of the steering wheel are simultaneously measured and associated to data recorded.

6. The Algorithm

The approach used to develop an algorithms with a high detection accuracy has been to operate mathematical optimization techniques using multiple regressions and linear discriminant analyses. One consider that, although the algorithm is formulated to provide the best performance when all the all the input signals are validated, one or more of the input signals could be missing (e.g., a reliable estimate of the road limits). Therefore, the system has been set to discriminate the valid signals and to ensure that the detection algorithm can always be able to operate.

The algorithms optimization for the detection of somnolence requires a measure of the limit threshold, or in any case of a specific marker that can establish the definition of drowsiness. This may be based on physiological variables, on measures of performance, or on subjective parameters, that have to be sufficiently repeatable in a given set of experiments so that the detection algorithm can be “trained” to determine the value of the alarm threshold definition. In particular, this latter must be adjustable to allow a turning of the system sensitivity according to the different driving conditions.

Note that a baselining is necessary for the algorithm to obtain measures for the individual driver comparable with all data collected.

Furthermore, the algorithm, to be truly effective in identifying the condition of drowsiness, should indicate when the somnolence level has exceeded a prespecified threshold, and this cannot be done through the evaluation of a single parameter, but by a linear combination of the different available measures.

7. The Analysis System

The definition of the input parameters used to measure the driving performance is independent of the identification of variables that were actually related to sleepiness. In the

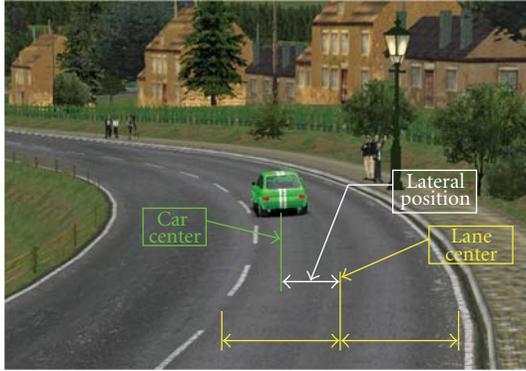


FIGURE 1: Means of lateral position.

literature there are several studies that report the changes in driving performance due to sleep deprivation through various levels of propensity to sleep [7, 8].

Results show a continuing decline of the driver performance with the progression of the state of drowsiness and the driver quality continuously reduces.

In all of these studies, the lateral control is used as a key parameter for the assessment of the driving quality. In particular, the ability to control the horizontal movement of the vehicle and to keep it properly and safely within the lane, has proved to be an excellent indicator of the driving performance.

The effectiveness of the lateral control operated by the driver can be expressed by calculating the standard deviation of the lateral position (Figure 1), defined as the horizontal position of the vehicle within the lane, determined with respect to a specific point of the vehicle and the road.

In most cases, the lateral position is determined by calculating the distance between the central axis of the vehicle and the center of the lane. When these two axes are aligned, the vehicle is located in an ideal position. When the vehicle is moving within the lane, either left or right, the deviation from the lane center, is measured by the so-called deviation of lateral position.

In addition to the lateral position (LP) is acquired and processed also the first derivative (LP') that describes the changes in lateral velocity and is defined such that positive values correspond to lateral movement of the car toward the left side of the lane limit, while negative values correspond to movements toward the right limit.

The magnitude of the drift to one side and its relationship with the actions made on the steering wheel to compensate for, can provide useful information on both the driving quality and style, defined as the tendency of drivers to over- or under-estimate the lateral drift and the greater or lesser propensity to operate correctly on the steering wheel.

7.1. The Fuzzification of the Inputs. Both lateral position and the drift are expressed as positive values if the vehicle has moved or drifted to right edge and negative if to the left edge of the roadway. Both variables are fuzzified using 5 fuzzy sets that describe the range values of -2 to $+2$ meters for the

lateral position and -20 and $+20$ centimeters for the drift as follows:

- (i) $Sx+$: very left;
- (ii) Sx : left;
- (iii) C : central;
- (iv) Dx : a right;
- (v) $Dx+$: very right.

The fuzzy sets are reported in Figure 2.

7.2. The Inferential Matrix. Using an inferential matrix, that includes as inputs lateral position and its first derivative, is generated a judgment (\mathcal{J}) on the quality of the vehicle lateral control operated by the driver, in the specific time interval delimited by the time step of signals acquisition.

This matrix considers “safe” the associations of inputs that bring the vehicle to converge towards the center of the lane, whereas are evaluated “unsafe” those that lead the vehicle to diverge from the center of the lane and, then, can lead to go off track.

The inference appear as follows:

| | $Sx+$ | Sx | C | Dx | $Dx+$ |
|-------|-------|-------|-------|-------|-------|
| $Sx+$ | $Us+$ | Us | NSS | S | $S+$ |
| Sx | Us | NSS | S | $S+$ | S |
| C | NSS | S | $S+$ | S | NSS |
| Dx | S | $S+$ | S | NSS | Us |
| $Dx+$ | $S+$ | S | NSS | Us | $Us+$ |

If LP is X And LP' is Y Then \mathcal{J} is Z

The possible outputs of the system are 5:

- (i) $Us+$: very unsafe;
- (ii) Us : unsafe;
- (iii) NSS : not so safe;
- (iv) S : safe;
- (v) $S+$: very safe.

7.3. The Defuzzification. According to the inferential rules activated in the matrix, a crisp value is obtained corresponding to the final score attributed to the driver at that given time of the driving.

This value is obtained by applying the method of the center of the areas to the fuzzy sets activated, according to membership functions similar to those used for the fuzzification of the inputs and will assume value in the range 0–100.

The fuzzy sets for the judgment output are reported in Figure 3.

8. The Evaluation of the Driving Style

Crisp values of judgments generated are progressively stored in a dynamic database. An external computational unit,

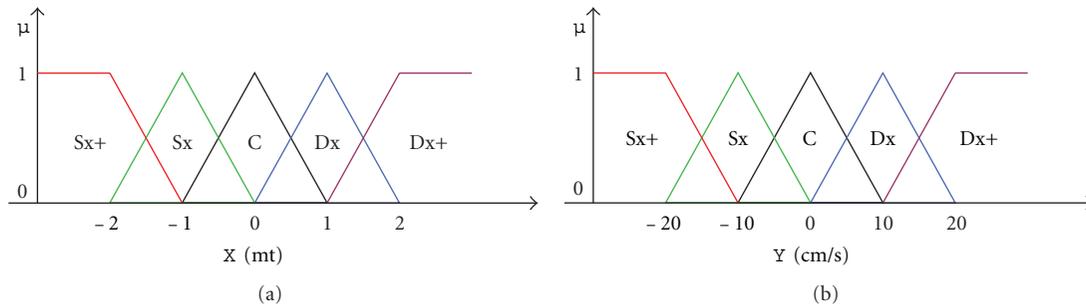


FIGURE 2: Membership functions of inputs.

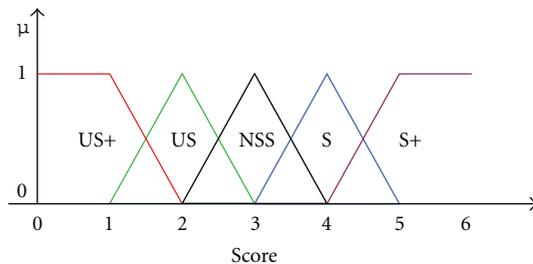


FIGURE 3: Membership functions of outputs.

linked to the database, calculates in real time the normalized statistics distribution of the judgment value, that is used as a reference to evaluate the performance of the driver in normal condition.

9. The Detection of the Driver Impairment

This unit performs a statistical comparison between the distribution stored in the database containing the reference for normal condition and the distribution recorded in the last 10 minutes of driving to detect any significant difference.

The system also works out the trend of variation of the judgment value, pointing out the trends describing a progressive deterioration of driving quality, expressed as a continuous reduction of the score of judgment.

10. Conclusion

The system that we developed here provides a valuable tool for exploring the possibility of implementing an on-board device to detect driver drowsiness and improve driving safety.

Our analysis algorithm allowed the system to carefully control the typical patterns of actions made by a specific driver in normal conditions and to detect significant differences in those patterns due to incoming of drowsiness conditions.

The system could benefit of advantages associated with the use of a specific model of steering control whose design is closely guided by human data, instead of a static threshold to quantify the limits of driver safety performance under sleepiness conditions.

The system seems to be very promising for further use as an effective safety device and appears also able to allow understanding of the driver behavior and the relationship holding between the driver's sensory cues and his consequent actions.

The use of baseline recording of "natural data points" reduces the influence on data analyses of variability in driver.

Acknowledgments

The work has been carried out in the framework of the Project PON01_00744 DriveIn2 "DRIVER Monitoring: Technologies, Methodologies, and IN-vehicle INnovative systems for a safe and eco-compatible driving," funded by the Italian Ministry for University and Research (MIUR).

References

- [1] A.C.I. and ISTAT, "Incidenti Stradali 2008," 2008.
- [2] E. R. Braver, J. F. Pantula, S. P. Baker et al., "The sleep of long-haul truck drivers," *The New England Journal of Medicine*, vol. 338, no. 6, pp. 389–391, 1998.
- [3] J. A. Horne and L. A. Reyner, "Sleep related vehicle accidents," *British Medical Journal*, vol. 310, no. 6979, pp. 565–567, 1995.
- [4] D. Fell, "The road to fatigue: circumstances leading to fatigue accidents," in *Fatigue and Driving: Driver Impairment, Driver Fatigue and Driving Simulation*, L. Hartley, Ed., Taylor & Francis, London, UK, 1995.
- [5] P. Philip, P. Sagaspe, J. Taillard et al., "Fatigue, sleep restriction, and performance in automobile drivers: a controlled study in a natural environment," *Sleep*, vol. 26, no. 3, pp. 277–280, 2003.
- [6] J. C. Stutts, J. W. Wilkins, J. S. Osberg, and B. V. Vaughn, "Driver risk factors for sleep-related crashes," *Accident Analysis and Prevention*, vol. 35, no. 3, pp. 321–331, 2003.
- [7] J. A. Horne and L. A. Reyner, "Driver sleepiness," *Journal of Sleep Research, Supplement*, vol. 4, supplement 2, pp. 23–29, 1995.
- [8] G. Gunzelmann, L. R. Moore, D. D. Salvucci, and K. A. Glucka, "Sleep loss and driver performance: quantitative predictions with zero free parameters," *Cognitive Systems Research*, vol. 12, no. 2, pp. 154–163, 2011.

Research Article

Theoretical Compartment Modeling of DCE-MRI Data Based on the Transport across Physiological Barriers in the Brain

Laura Fanea,^{1,2} Leontin I. David,¹ Andrei Lebovici,²
Francesca Carbone,³ and Silviu A. Sfrangeu²

¹ Department of Biomedical Physics, Physics Faculty, Babes-Bolyai University, No. 1 M. Kogalniceanu Street, 400084 Cluj-Napoca, Romania

² Department of Radiology, Cluj County Emergency Hospital, No. 3-5, Clinicilor Street, 400006 Cluj-Napoca, Romania

³ Faculty of Pharmacy, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy

Correspondence should be addressed to Laura Fanea, laura.fanea@yahoo.com

Received 29 September 2011; Accepted 18 February 2012

Academic Editor: Maria Crisan

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

Neurological disorders represent major causes of lost years of healthy life and mortality worldwide. Development of their quantitative interdisciplinary *in vivo* evaluation is required. Compartment modeling (CM) of brain data acquired *in vivo* using magnetic resonance imaging techniques with clinically available contrast agents can be performed to quantitatively assess brain perfusion. Transport of ¹H spins in water molecules across physiological compartmental brain barriers in three different pools was mathematically modeled and theoretically evaluated in this paper and the corresponding theoretical compartment modeling of dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) data was analyzed. The pools considered were blood, tissue, and cerebrospinal fluid (CSF). The blood and CSF data were mathematically modeled assuming continuous flow of the ¹H spins in these pools. Tissue data was modeled using three CMs. Results in this paper show that transport across physiological brain barriers such as the blood to brain barrier, the extracellular space to the intracellular space barrier, or the blood to CSF barrier can be evaluated quantitatively. Statistical evaluations of this quantitative information may be performed to assess tissue perfusion, barriers' integrity, and CSF flow *in vivo* in the normal or disease-affected brain or to assess response to therapy.

1. Introduction

Neurological disorders are diseases of the central and peripheral nervous system affecting approximately one billion people worldwide irrespective of age, sex, education, or income. The most frequent neurological disorders at present are migraines, followed by neurological problems caused by cerebrovascular disease, epilepsy, Alzheimer's disease, and so forth [1]. Hydrocephalus is especially painful, representing the leading cause for brain surgery in children in the United States of America [2]. Neuropsychiatric disorders and injuries, in particular, represent major causes of lost years of healthy life and are significantly underestimated when measured by mortality alone [1].

Impact of neurological disorders is expected to increase, for example, the number of people affected by Alzheimer's disease and other dementia conditions is expected to double

every twenty years. Almost seven million people die every year as a result of neurological disorders, the most affected being people in the lower middle income category. Investigation of neurological disorders increases in importance due to prolonged ageing also [1].

Neuroscientifically relevant observations were recorded since 4000 BC and they contributed to the development of neurosciences [3]. During the previous two centuries, brain research belonged to many different areas that differed in methodology and targets: the morphological, the physiological, and the psychological. The latter used to consider the brain as a black box where only the input and output were known but not at all the neuronal components and the way they interact with each other. More recently, scientific and technological research, from molecular to behavioural levels, have been carried out but they have not been developed in a really interdisciplinary way [1].

Kinetics of an agent through physiological barriers, including physiological brain barriers, can be quantitatively evaluated *in vivo*, using exogenous or endogenous magnetic resonance imaging (MRI) techniques and compartment modeling (CM) of the MRI data [4–6]. Theoretical description of two and three CMs of endogenous (without contrast agents for MRI, the ASL techniques) and exogenous (with extra- or intracellular contrast agents for MRI) perfusion MRI techniques have been described [4–6]. The signal-to-noise ratio of the data acquired using endogenous perfusion MRI techniques is still much lower than that of the data obtained with exogenous perfusion MRI techniques. Intracellular contrast agents present at the moment are only used in clinical trials. Perfusion MRI techniques using extracellular contrast agents (the DCE-MRI techniques) are less invasive and they are routinely used in clinical MRI, including cerebral MRI.

Mathematical modeling of data acquired using exogenous perfusion MRI techniques with intracellular contrast agents (the ssCE techniques) which might become mathematically very complex since there will be more mechanisms of relaxation time modifications in each compartment of a voxel (i.e., volumetric image element) to be considered. This will lead to more complex equations to be solved and would introduce more parameters to be estimated when fitting the MRI data to the mathematical model developed.

Mathematical modeling of the brain data acquired *in vivo* using exogenous perfusion MRI techniques with extracellular contrast agents is easier since the signal-to-noise ratio of the MRI data is the highest and the contrast agent does not penetrate in the extracellular space (EES) through the intact blood-to-brain barrier (BBB). Increase in the signal-to-noise ratio of the DCE-MRI has recently been obtained using a slow infusion technique of Gd-based extracellular clinically available contrast agents [7]. This infusion technique allowed a much easier mathematical modeling of the mouse brain DCE-MRI data. Slow infusion techniques can also be used clinically. Yankeelov et al. [8] used a slow infusion technique of Gd-based extracellular contrast agents to quantitatively evaluate DCE-MRI data of the breast.

The mathematical modeling of the perfusion MRI data in general and DCE-MRI data in particular, based on the physiological compartmentalization of voxels, gives more complex and realistic information on the kinetics of agents through barriers between physiological compartments. For brain and brain conditions, this might allow quantitative assessment of the BBB, blood-to-CSF barrier (B-CSF-B), or that of the output to input flow (IOF) imbalance of ^1H spins in water molecules through the CSF spaces.

Theoretical description of the physiological compartmentalization of tissue and CSF voxels and the mathematical modeling of the blood, tissue, and CSF data acquired using DCE-MRI and a slow infusion technique of Gd-based contrast agents is presented in this paper. Kinetics of ^1H spins in water molecules present in two different phases (with unaffected and shortened spin-lattice relaxation time, T_1) in the three pools assumed (blood, tissue, and CSF) was quantitatively evaluated in this study. The relative volumes of the physiological compartments of voxels situated in the

tissue and CSF regions can also be estimated using this physiological CM of the DCE-MRI brain data.

Results in this study show that quantitative information can be extracted from brain data acquired *in vivo* using DCE-MRI techniques. Integrity of physiological barriers in the brain, IOF imbalances in CSF spaces, and relative volumes of physiological compartments can be assessed using this quantitative information.

Future possible statistical experimental studies using these CMs might provide quantitative information on the BBB and B-CSF-B integrity and/or shrinkage of the brain tissue in the diseased brain, for example. Potential of these CMs for the *in vivo* quantitative monitoring of neurological disorders, therapies, or normal brain function can also be assessed.

2. Mathematical Model

The ^1H nuclei relevant for DCE-MRI images in the blood, tissue, and CSF pools exist in two different phases: one with unaffected longitudinal relaxation time T_1 and one with T_1 shortened. The relaxation time T_1 is shortened due to the spin-lattice interactions between the ^1H nuclei in water molecules with the paramagnetic Gd ions in the exogenous clinically available contrast agents for MRI. Transport of the ^1H nuclei in these two aqueous phases across the barriers between the compartments of the pools evaluated was mathematically modeled.

2.1. Transport in the Blood Pool. Infusion of the agent in the blood pool is continuous and, therefore, the ^1H spins in the two phases flow continuously in this pool [9–11] at a concentration rate C_{IN} and leave the pool at a rate k_{OUT} , their kinetics being described by the following arterial input function, AIF:

$$\text{AIF}(t) = C_{\text{IN}} t e^{-tk_{\text{OUT}}}. \quad (1)$$

The AIF in (1) refers to the time points before and during injection. No more contrast agents will enter the blood compartment after the end of the slow infusion of the contrast agent, only the elimination mechanisms will take place during this period of time.

Concentration of the ^1H spins in water molecules in the mixed phases in the blood, C_{BLOOD} , depends on the hematocrit, Hct, level and is given by the arterial input function multiplied by the $(1-\text{Hct})$ factor.

2.2. Tissue Perfusion. Each voxel corresponding to white or gray matter regions was compartmentalized as shown in Figure 1. The four physiological compartments of a tissue voxel are blood (accessible to ^1H spins in water molecules), extracellular space (EES—accessible to ^1H spins in water molecules), intracellular space (IES—accessible to ^1H spins in water molecules), and a space not accessible to water (NOW—not accessible to ^1H spins in water molecules). The bidirectional transport of the ^1H spins across the blood to EES and the EES to IES barriers are represented with arrows. Rates of ^1H spins with unaffected T_1 can be neglected (~ 0)

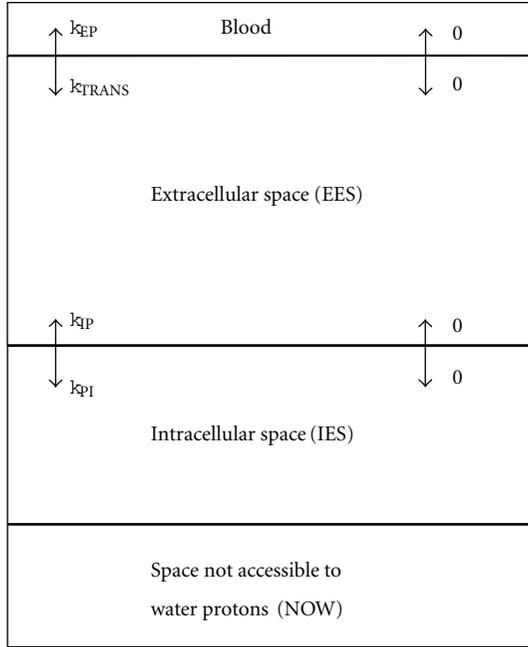


FIGURE 1: Physiological compartmentalization of a tissue voxel in the region of white or gray matter.

compared to that of ^1H spins with short T_1 . The fast and slow exchange of the ^1H spins in water molecules between these compartments is bidirectional. These exchanges together with accumulation of the ^1H spins in the two phases over time in a compartment produce dynamic changes of the amplitude of the corresponding nuclear magnetic resonance (NMR) signals.

The four tissue compartments in Figure 1 are the blood, the EES, the intracellular space (IES), and a space not accessible to ^1H nuclei in water molecules (NOW). The relative volumes of these compartments: v_{BLOOD} , v_{EES} , v_{IES} , and v_{NOW} fulfill the condition:

$$v_{\text{BLOOD}} + v_{\text{EES}} + v_{\text{IES}} + v_{\text{NOW}} = 1. \quad (2)$$

Transport of the ^1H spins in water molecules across barriers is bidirectional. Its mechanisms are similar to action potential transport across cellular membranes due to the concentration gradient of the ^1H spins in a phase in a compartment.

The concentration of the ^1H spins producing changes of the NMR signal in a tissue, C_{TISSUE} , depends on the concentration of the ^1H nuclei producing changes of the NMR signals in each of these compartments: C_{BLOOD} , C_{EES} and C_{IES}

$$C_{\text{TISSUE}}(t) = v_{\text{BLOOD}}C_{\text{BLOOD}}(t) + v_{\text{EES}}C_{\text{EES}}(t) + v_{\text{IES}}C_{\text{IES}}(t). \quad (3)$$

The fast and slow transport of the ^1H spins in the two phases across the blood to EES and the EES to IES barriers also contributes to the NMR signal intensity changes over time.

2.2.1. Bidirectional Transport across the Blood to EES Barrier. Concentration of the ^1H spins producing changes of the MRI

signal intensity in a voxel due to its transport from the blood into the EES compartment can be determined by solving the Kety-Schmidt equation (4) for each phase of the ^1H nuclei in water molecules.

Gd-based extracellular contrast agents do not penetrate into the EES space while the BBB is intact and the ^1H spins with shortened T_1 are moving much slower compared with those with unaffected T_1 [4, 5]. Contribution of the ^1H spins with unaffected T_1 can, therefore, be neglected and the Kety-Schmidt equation that needs to be solved is given below in (4).

The rate of the concentration variation of the ^1H spins in water molecules in the EES compartment, " dC_{EES}/dt " increases due to the transport of the ^1H nuclei with shortened T_1 from the blood into the EES compartment at a rate k_{TRANS} and decreases due to these spins reentering the blood compartment at a rate k_{EP} :

$$\frac{dC_{\text{EES}}(t)}{dt} + k_{\text{EP}}C_{\text{EES}}(t) = k_{\text{TRANS}}C_{\text{BLOOD}}(t). \quad (4)$$

All initial concentrations (at $t = 0$) are zero.

Equation (4) is a linear nonhomogeneous first order differential equation whose general solution is given in the Appendix.

2.2.2. Bidirectional Transport across the EES to IES Barrier.

Extracellular contrast agents do not enter the IES compartment and contribution of ^1H spins with unaffected T_1 to the signal intensity change can again be neglected on the assumption made in the previous subsection. The rate of the concentration variation of the ^1H spins in water molecules in the IES space, dC_{IES}/dt , increases due to the transport of the ^1H nuclei with shortened T_1 from the EES into the IES compartment at a rate k_{PI} and decreases due to these spins reentering the EES compartment at a rate k_{IP} :

$$\frac{dC_{\text{IES}}(t)}{dt} + k_{\text{IP}}C_{\text{IES}}(t) = k_{\text{PI}}C_{\text{EES}}(t), \quad (5)$$

with $C_{\text{EES}}(t)$ obtained by solving (4). Solution to this equation is also given in the Appendix.

2.3. Flow in the CSF Spaces.

CSF is formed by three different important mechanisms and flows several regions in the brain [12]. Only the main mechanism of CSF formation, that of the blood filtration, was considered in this study. Kinetics of the ^1H spins producing changes of the MRI signal intensity in CSF spaces will, therefore, be different compared to that in the tissue. Each voxel in a CSF region can be compartmentalized in two spaces [13] as described in Figure 2. The two physiological compartments of a CSF voxel are blood and CSF, both accessible to ^1H nuclei in water molecules. ^1H nuclei in water molecules in the blood are filtrated in a CSF region at an overall rate k'_{IN} and evacuate the space at an overall rate k'_{OUT} . The fluid flowing through, accumulated in the space, represents the CSF. The arrows represent the unidirectional transport of ^1H nuclei in water molecules across the blood to CSF barrier.

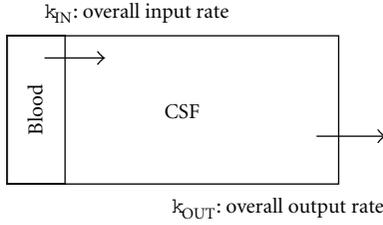


FIGURE 2: Physiological compartmentalization of a CSF voxel.

There are two pools containing ^1H nuclei in water molecules in a CSF voxel: the blood and the CSF pool. The relative volumes of these two compartments are: v_{BLOOD} , the relative volume of the blood pool, and v_{CSF} , the relative volume of the CSF pool:

$$v_{\text{BLOOD}} + v_{\text{CSF}} = 1. \quad (6)$$

Flow of the mixed phases ^1H spins in water molecules in a CSF space is continuous and it will be defined by a CSF input function (CSFIF), similar to the AIF. The spins enter a CSF space at a rate k'_{IN} and leave the space at a rate k'_{OUT}

$$\text{CSFIF}(t) = k'_{\text{IN}} t e^{-k'_{\text{OUT}} t}. \quad (7)$$

The disturbance of the input and output flow mechanisms through the B-CSF-B barrier together with the predominance of the disturbance of a mechanism can be quantitatively assessed by the output to input flow, IOF. The output to input flow, IOF, represents the mean relative normalized rate of ^1H nuclei in water molecules in a CSF space. The mean input/output rates of ^1H nuclei in water molecules flowing in a CSF space are calculated for each disease stage (e.g., control, mild, and severe). Normalization of each mean input/output rate in a CSF space at a disease stage is performed against the corresponding control (normal brain) input/output mean rate in that CSF space:

$$\text{IOF} = \frac{\bar{k}'_{\text{IN}} \Big|_{\text{NORMALIZED}}}{\bar{k}'_{\text{OUT}} \Big|_{\text{NORMALIZED}}}. \quad (8)$$

3. Results and Discussion

Similar blood concentration curves and AIFs were obtained for the mouse [7] and rat [13] brain DCE-MRI data. The overall input and output rates in the blood pool of the rat brains evaluated showed that the most rapid evacuation of the contrast agent from the blood pool took place at the mild stages of communicating hydrocephalus (C-HC), compared to the normal and severely C-HC-affected rat brains [13]. The values of the overall input rates in the blood pool ranged from 0.0427 to 0.0956 mM/min, while that of the overall output rates in the blood pool ranged from 0.0281 and 0.0344 1/min [13].

Transport rates between the blood compartment and the EES compartment in the tissue pool ranging between 0 and 0.0005 1/min were calculated using this compartment

modeling applied to DCE-MRI images of normal and hydrocephalic rat brains [13]. The changes of the NMR signal intensities in the brain regions characterized by null k_{TRANS} values were produced only by the ^1H nuclei in water molecules in the blood compartment in the tissue pool evaluated. The zero values of these rates also show that resolution and signal-to-noise ratio of the DCE-MRI technique evaluated need to be increased to detect transport rate values as small as 0.0001 1/min. All k_{TRANS} values calculated showed that, as opposed to normal pressure hydrocephalus [14], the BBB is intact in C-HC.

For fractional volumes of the EES compartment of 20%, the fractional volumes calculated for the blood compartment in the tissue pool (cortex and thalamus—left and right hemispheres) ranged between 0.05 and 7.98%, while the fractional volumes of the IES compartment ranged between 72 and 77% [13].

The IOF values calculated for the CSF pool clearly showed that these are the most sensitive indicators of the mechanisms of the dynamic disturbances in C-HC. The flow disturbances were more pronounced in the aqueduct (IOF = 2.49) and in the IVth ventricle (IOF = 1.96) in the mild stage of the C-HC. The IOF values calculated (IOF > 1) in the lateral ventricle (mild and severe stages of C-HC), aqueduct (mild and severe C-HC stages) and the IVth ventricle (mild C-HC stage) were larger than the corresponding IOF values calculated (IOF = 1) in the normal brain. The most pronounced increases of the IOF values were detected for the mild C-HC stages in all CSF spaces analyzed. In the IVth ventricle, the IOF value increased (IOF = 1.96) for the mild stage of C-HC and decreased below the corresponding IOF (1/IOF = 1.7) for the severe stage of C-HC. The IOF values calculated in the lateral ventricle and aqueduct increased relatively to the corresponding IOF values of the normal brain, and although they decreased in the severe C-HC stage relative to the mild stage, they remained larger than the corresponding IOF values calculated for the normal brain.

The increases of the IOF values can be correlated with decreases of the overall output rates relative to the corresponding input rates in a CSF space at a disease stage. More ^1H spins with decreased T_1 will accumulate at the level of a CSF space compared to the normal brain (control stage), producing volume dilatations of these spaces compared to the corresponding spaces of the normal brain. These dilatations reduced in the lateral ventricle and aqueduct, IOF decreased with disease severity from 1.87 to 1.55 and from 2.49 to 1.4. In the IVth ventricle, IOF increased in the mild C-HC stage (IOF = 1.96) and decreased in the severe stage (1/IOF = 1.7). More ^1H spins with decreased T_1 will accumulate in the CSF spaces evaluated in the mild C-HC stage (the aqueduct and the lateral and the IVth ventricles) and in the severe C-HC stage (aqueduct and lateral ventricle) compared to the normal brain (control stage). Many more ^1H spins with decreased T_1 will evacuate the IVth ventricle in the severe stage of C-HC compared to the normal brain (control stage). This massive evacuation at the level of the IVth ventricle in the severe C-HC stage indicates the disruption of the B-CSF-B at this level of the CSF space in the severe C-HC stage. The calculated IOF value at the level

of the IVth ventricle in the severe C-HC stage indicates the disruption of the normal evacuation mechanisms of the ^1H spins with decreased T_1 at this level of the CSF and for this C-HC stage. The evacuation, in this situation, is made as if no B-CSF-B exists anymore.

All IOF values calculated in the C-HC brains can physiologically be correlated with the disruption of the B-CSF-B in C-HC at the level of the aqueduct and the lateral and the IVth ventricles. These calculated IOF values indicate the predominance of the disturbed CSF flow mechanism. The disturbance of the input mechanism is predominant if $\text{IOF} > 1$, while that of the output mechanism is predominant if $\text{IOF} < 1$.

Increases in the fractional blood volumes in the rat brain cortex of up to 160 times relative to the control values were detected using this CM of the DCE-MRI rat brain data. These changes might be physiologically correlated to the shrinkage of the brain tissue in the C-HC.

In the CSF pool, values of the relative volume of the CSF compartment less than 100% were estimated [13]. The blood compartment was detected in the lateral and the IVth ventricles and the values of its fractional volume ranged from 9.4 to 16.5% [13]. The largest values of the fractional volume of the blood compartment calculated in the normal brain, show the presence of the blood compartment in the small CSF spaces (a few pixels), due to the partial volume and noise effects affecting the MRI data in these regions. These data show that more accurate information could be extracted from the mathematically modeled DCE-MRI data acquired with higher spatial resolution and signal-to-noise ratio.

4. Conclusions

Compartment modeling of the DCE-MRI data provides quantitative information on the permeability of the BBB and B-CSF-B. Transfer rates as small as 0.0001 1/min were estimated by mathematically modeling the DCE-MRI data in the rat brain tissue pool [13]. The transfer rates ranged from 0 to 0.0005 1/min, showing no BBB breakdown in the rat brain affected by C-HC [13].

Tissue perfusion and kinetics of ^1H nuclei in water molecules in different phases through different tissue and CSF compartments of the normal and disease-affected brain can be quantitatively evaluated *in vivo* using CM of the DCE-MRI data [13].

Statistical experimental analyses of these CMs for refined stages of disease severity are required for accurate description of the normal and/or of brain disease mechanisms in general and in C-HC in particular.

Even without any statistical analysis, the calculated IOF values for each CSF space analyzed clearly show that this is the most sensitive indicator of the CSF flow disturbances through the CSF spaces in C-HC. The predominance of the input or output flow disturbance in a CSF space at a disease stage can also be established based on the IOF values calculated.

CM of the DCE-MRI data may represent an important clinical imaging analysis method. It can provide quantitative information that can be used to assess physiology of the normal brain, mechanisms of brain diseases, including hydrocephalus, or responses to therapies.

Appendices

A. Tissue Concentration Modeling

Concentration of ^1H spins in water molecules in the tissue can be estimated quantitatively by fitting the DCE-MRI data to the tissue concentration curve obtained by solving (4) and (5). Solution of (4) is given below

$$C_{\text{EES}}(t) = \frac{C_{\text{IN}}k_{\text{TRANS}}}{(k_{\text{EP}} - k_{\text{OUT}})^2} \times \left\{ [(k_{\text{EP}} - k_{\text{OUT}})t - 1]e^{-k_{\text{OUT}}t} + e^{-k_{\text{EP}}t} \right\}. \quad (\text{A.1})$$

The contrast agents in DCE-MRI are extracellular and only transport of ^1H nuclei in water molecules takes place between the EES and the IES compartments. Based on these assumptions, concentration of ^1H spins in the IES and EES compartments will be nearly the same

$$C_{\text{IES}}(t) \cong C_{\text{EES}}(t). \quad (\text{A.2})$$

The tissue concentration curves are then estimated by placing the C_{EES} and C_{IES} expressions above in (3).

B. Concentration Curves

Kinetics of ^1H spins in water molecules in the three pools evaluated can be obtained by fitting the concentration curves to (1), (7), (A.1). The concentration curves can be obtained by normalizing the MRI signal intensity, $\text{SI}(t)$, over time to the signal intensity acquired before injecting the contrast agent, $\text{SI}(0)$. The signal intensity at any of the time points considered is acquired using gradient echo with variable flip angle based pulse sequences with as short as possible echo times [13]:

$$\frac{\text{SI}(0)}{\text{SI}(t)} = \frac{(1 - e^{-T_R/T_{10}})(1 - e^{-T_R/T_1} \cos\text{FA})}{(1 - e^{-T_R/T_1})(1 - e^{-T_R/T_{10}} \cos\text{FA})}. \quad (\text{B.1})$$

Concentration is direct proportional to T_1 [8, 9] and, if T_{10} is known, the concentration curves are obtained from (B.1). T_{10} and T_1 represent the spin-lattice relaxation times of ^1H nuclei in water molecules in a region evaluated (i.e., blood, tissue, or CSF) before and after infusion of the contrast agent.

FA and T_R are pulse sequence parameters: the flip angle and the repetition time

$$T_1 = \frac{T_R}{\ln \left[\frac{\cos FA - (SI(0)/SI(t)) \times (1 - e^{-T_R/T_{10}} \cos FA)/(1 - e^{-T_R/T_{10}})}{1 - (SI(0)/SI(t)) \times (1 - e^{-T_R/T_{10}} \cos FA)/(1 - e^{-T_R/T_{10}})} \right]} \quad (\text{B.2})$$

Initial T_{10} values of ^1H nuclei in water molecules in the blood, tissue and CSF regions can be calculated *in vivo* [15]. Calculation of T_1 is not necessary. The concentration curves in the blood data can be estimated by fitting the MRI data in the sagittal sinus [16] to (B.2).

Acknowledgments

The first author (L. Fanea) is grateful to The Child Brain Foundation and to the financial support under EC Grant POSDRU/89/1.5/S/60189.

References

- [1] J. A. Aarli, T. Dua, A. Janca, and A. Muscetta, *Neurological Disorders Public Health Challenges*, World Health Organization Press, Geneva, Switzerland, 2006.
- [2] J. M. Drake, J. R. W. Kestle, R. Milner et al., "Randomized trial of cerebrospinal fluid shunt valve design in pediatric hydrocephalus," *Neurosurgery*, vol. 43, no. 2, pp. 294–303, 1998.
- [3] D. Semple, R. Smith, J. Burns et al., "Flexicover," in *Oxford Handbook of Psychiatry*, pp. 20–28, Oxford University Press, New York, NY, USA, 2nd edition, 2005.
- [4] P. S. Tofts, G. Brix, D. L. Buckley et al., "Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusible tracer: standardized quantities and symbols," *Journal of Magnetic Resonance Imaging*, vol. 10, no. 3, pp. 223–232, 1999.
- [5] X. Li, W. D. Rooney, and C. S. Springer Jr., "A unified magnetic resonance imaging pharmacokinetic theory: intravascular and extracellular contrast reagents," *Magnetic Resonance in Medicine*, vol. 54, no. 6, pp. 1351–1359, 2005.
- [6] L. M. Parkes, "Quantification of cerebral perfusion using arterial spin labeling: two-compartment models," *Journal of Magnetic Resonance Imaging*, vol. 22, no. 6, pp. 732–736, 2005.
- [7] M. M. Pike, W. D. Rooney, X. Li, C. N. Stoops, G. Y. Gillespie, and C. S. Springer Jr., "High-field MRI detection of magnetist permeation into normal mouse brain parenchymal and ventricular spaces," in *Proceedings of the 16th Annual International Society for Magnetic Resonance in Medicine Meeting*, Toronto, Canada, May 2008.
- [8] T. E. Yankeelov, M. Lepage, A. Chakravarthy et al., "Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer: initial results," *Magnetic Resonance Imaging*, vol. 25, no. 1, pp. 1–13, 2007.
- [9] D. Z. Ting, P. S. Hagan, and S. I. Chan, "Nuclear magnetic resonance studies of cation transport across vesicle bilayer membranes," *Biophysical Journal*, vol. 34, no. 2, pp. 189–215, 1981.
- [10] X. Fan, M. Medved, G. S. Karczmar et al., "Diagnosis of suspicious breast lesions using an empirical mathematical model for dynamic contrast-enhanced MRI," *Magnetic Resonance Imaging*, vol. 25, no. 5, pp. 593–603, 2007.
- [11] T. E. Yankeelov, J. J. Luci, L. M. DeBusk, P. C. Lin, and J. C. Gore, "Incorporating the effects of transcytolemmal water exchange in a reference region model for DCE-MRI analysis: theory, simulations, and experimental results," *Magnetic Resonance in Medicine*, vol. 59, no. 2, pp. 326–335, 2008.
- [12] J. G. Veening and H. P. Barendregt, "The regulation of brain states by neuroactive substances distributed via the cerebrospinal fluid; a review," *Cerebrospinal Fluid Research*, vol. 7, no. 1, pp. 1–16, 2010.
- [13] L. Fanea, S. Rashid, J. P. McAllister Jr. et al., "Dynamic contrast enhanced MRI for investigation of the blood-brain barrier in an experimental model of communicating hydrocephalus," in *Proceedings of the 17th Annual International Society for Magnetic Resonance in Medicine Meeting*, Honolulu, Hawaii, USA, May 2009.
- [14] D. Greitz, "Radiological assessment of hydrocephalus: new theories and implications for therapy," *Neurosurgical Review*, vol. 27, no. 3, pp. 145–165, 2004.
- [15] R. A. de Graaf, P. B. Brown, S. McIntyre, T. W. Nixon, K. L. Behar, and D. L. Rothman, "High magnetic field water and metabolite proton T1 and T2 relaxation in rat brain *in vivo*," *Magnetic Resonance in Medicine*, vol. 56, no. 2, pp. 386–394, 2006.
- [16] J. R. Ewing, R. A. Knight, T. N. Nagaraja et al., "Patlak plots of Gd-DTPA MRI data yield blood-brain transfer constants concordant with those of ^{14}C -sucrose in areas of blood-brain opening," *Magnetic Resonance in Medicine*, vol. 50, no. 2, pp. 283–292, 2003.

Research Article

Theorems and Application of Local Activity of CNN with Five State Variables and One Port

Gang Xiong,¹ Xisong Dong,¹ Li Xie,² and Thomas Yang³

¹ State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

² Department of Information Science and Electronics Engineering, Zhejiang University, Hangzhou 310027, China

³ The Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA

Correspondence should be addressed to Xisong Dong, dongcomic04@163.com

Received 22 September 2011; Accepted 15 November 2011

Academic Editor: Sheng-yong Chen

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

Coupled nonlinear dynamical systems have been widely studied recently. However, the dynamical properties of these systems are difficult to deal with. The local activity of cellular neural network (CNN) has provided a powerful tool for studying the emergence of complex patterns in a homogeneous lattice, which is composed of coupled cells. In this paper, the analytical criteria for the local activity in reaction-diffusion CNN with five state variables and one port are presented, which consists of four theorems, including a series of inequalities involving CNN parameters. These theorems can be used for calculating the bifurcation diagram to determine or analyze the emergence of complex dynamic patterns, such as chaos. As a case study, a reaction-diffusion CNN of hepatitis B Virus (HBV) mutation-selection model is analyzed and simulated, the bifurcation diagram is calculated. Using the diagram, numerical simulations of this CNN model provide reasonable explanations of complex mutant phenomena during therapy. Therefore, it is demonstrated that the local activity of CNN provides a practical tool for the complex dynamics study of some coupled nonlinear systems.

1. Introduction

Coupled nonlinear dynamical systems have been widely studied in recent years. However, the dynamical properties of these systems are difficult to deal with. Although the research on emergence and complexity has gained much attention during the past decades, the determination, prediction, and control of the complex patterns generated from high-dimensional coupled nonlinear systems are still far from perfect. Nature abounds with complex patterns and structures emerging from homogeneous media, and the local activity is the origin of these complexities [1, 2]. The cellular neural network (CNN), firstly introduced by Chua and Yang [3] as an implementable alternative to fully connected Hopfield neural network, has been widely studied for image processing, robotic, biological versions, and higher brain functions, and so on [3]. Many of the coupled nonlinear systems can be modeled and studied via the CNN paradigm

[4]. The local activity proposed by Chua asserts that a wide spectrum of complex behaviors may exist if the cell parameters of the corresponding CNN are chosen in or nearby the edge of chaos [2, 4]. There have been quite a few new methods developed for complex systems [5–8], and local activity has attracted the attention of many researchers. Now, local activity has been successfully applied to the research of complex patterns generated from several CNNs in physical, biological, and chemical domains, such as Fitzhugh-Nagumo equation [9], Brusselator equation [10], Gierer-Meinhardt equation [11], Oregonator equation [12], Hodgkin-Huxley equation [13], Van Der Pol equation [14], the biochemical model [15], coupled excitable cell model [16], tumor growth and immune model [17], Lorenz model [18], advanced image processing [19], Rossler equation [20], images analysis [21, 22], data prediction [23], neutron transport equation [24], vision safety [25], retinomorph model [26], and theory research [27–30], and so forth.

Although Chua presents the main theorem of local activity at a cell equilibrium point [1, 2], it is actually difficult to “test” directly the complex patterns of the high-dimensional coupled nonlinear systems, since the theorem contains no recipe for finding whether a variable actually exists or not. It is necessary to develop some mathematical criteria according to the numbers of the variables and ports; that is the topic addressed in this paper.

The remaining of this paper is organized as follows. The local activity of CNN is introduced in Section 2. A set of theorems for testing the local activity of reaction-diffusion CNN with five state variables and one port are set up in Section 3. As an application of the theorems, a coupled reaction-diffusion CNN of hepatitis B Virus (HBV) mutation-selection model is introduced, aiming at describing HBV mutation in the therapeutic process. The bifurcation diagrams of this CNN are developed and some numerical simulations are presented in Section 4. Concluding remarks are given in Section 5.

2. Local Activity Theory of CNN

The CNN architecture is composed of a two-dimensional $M \times N$ array of cells. Each cell is denoted by $C(i, j)$, where $i = 1, 2, \dots, M$, $j = 1, 2, \dots, N$. The dynamics of each cell is given by the equation:

$$\dot{x}_{ij} = -x_{ij} + \sum_{k=-r}^r \sum_{l=-r}^r a_{kl} y_{i+k, j+l} + \sum_{k=-r}^r \sum_{l=-r}^r b_{kl} u_{i+k, j+l} + z_{ij}, \quad (1)$$

where x_{ij} , y_{ij} , u_{ij} are the state, output, and input variables of the cell, respectively. $a_{k,l}$, $b_{k,l}$, z_{ij} are the elements of the A-template, the B-template, and threshold, respectively. r is the radius of influence sphere. The output y_{ij} is the piecewise linear function given by

$$y_{i,j} = \frac{1}{2} \left(\left| x_{i,j} + 1 \right| - \left| x_{i,j} - 1 \right| \right), \quad (2)$$

$$i = 1, 2, \dots, M; \quad j = 1, 2, \dots, N.$$

Clearly, CNN with different template elements may have different functions.

A vast majority of active homogeneous media that are known to exhibit complexity are modeled by a reaction-diffusion partial differential equation (PDE):

$$\frac{\partial x_i}{\partial t} = f_i(X) + D_i \left(\frac{\partial^2 x_i}{\partial x^2} + \frac{\partial^2 x_i}{\partial y^2} + \frac{\partial^2 x_i}{\partial z^2} \right), \quad i = 1, 2, \dots, n, \quad (3)$$

where $X = (x_1, x_2, \dots, x_n)$ is state variables, (x, y, z) is spatial coordinates, $f_i(x_1, x_2, \dots, x_n)$ is a coupled nonlinear vector function called the kinetic term, and D_1, D_2, \dots, D_n are constants called diffusion coefficients. Replacing the Laplace in above formulation by its discrete version yields

$$\frac{\partial^2 x_i}{\partial x^2} + \frac{\partial^2 x_i}{\partial y^2} + \frac{\partial^2 x_i}{\partial z^2} \longrightarrow \nabla^2 X_{\alpha,\beta,\gamma}, \quad (4)$$

where

$$\begin{aligned} & \left(\nabla^2 X_{\alpha,\beta,\gamma} \right)_i \\ &= x_i(\alpha - 1, \beta, \gamma) + x_i(\alpha + 1, \beta, \gamma) + x_i(\alpha, \beta - 1, \gamma) \\ & \quad + x_i(\alpha, \beta + 1, \gamma) + x_i(\alpha, \beta, \gamma - 1) \\ & \quad + x_i(\alpha, \beta, \gamma + 1) - 6x_i(\alpha, \beta, \gamma). \end{aligned} \quad (5)$$

Chua et al. have introduced reaction-diffusion CNN equations:

$$\dot{X}_{\alpha,\beta,\gamma} = f(X_{\alpha,\beta,\gamma}) + D \nabla^2 X_{\alpha,\beta,\gamma}, \quad (6)$$

where $D = \text{diag}(D_1, D_2, \dots, D_n)$, $X_{\alpha,\beta,\gamma}$ denotes the state variable located at a point in three-dimensional space with spatial coordinates. Chua refers to the process of transforming a PDE into a reaction-diffusion CNN [2].

From Chua and his collaborators' point, PDEs are merely mathematical abstractions of nature, and the concept of a continuum is in fact an idealization of reality. Even the collection of all electrons in a solid does not form a continuum, because much volume separating the electrons from the nucleus represents a vast empty space [2]. Reaction-diffusion CNNs have been used to model some phenomena with important practical backgrounds, which were described by PDEs.

Generally speaking, in a reaction-diffusion CNN, every cell has n state variables, but only m ($m \leq n$) state variables couple directly to their nearest neighbors via “reaction-diffusion”. Consequently, each cell has the following state equations:

$$\begin{aligned} \dot{V}_a &= f_a(V_a, V_b) + I_a, \\ \dot{V}_b &= f_b(V_a, V_b), \end{aligned} \quad (7)$$

where

$$\begin{aligned} V_a &= [V_1, V_2, \dots, V_m]^T, \quad V_b = [V_{m+1}, V_{m+2}, \dots, V_n]^T, \\ f_a &= [f_1(V_a, V_b), f_2(V_a, V_b), \dots, f_m(V_a, V_b)]^T, \\ f_b &= [f_{m+1}(V_a, V_b), f_{m+2}(V_a, V_b), \dots, f_n(V_a, V_b)]^T, \\ I_a &= [D_1 \nabla^2 V_1, D_2 \nabla^2 V_2, \dots, D_m \nabla^2 V_m]^T. \end{aligned} \quad (8)$$

The cell equilibrium point $Q_i = (V_a^i, V_b^i) (\in \mathbb{R}^n)$ of equation (7) can be determined via

$$\begin{aligned} f_a(V_a, V_b) &= 0, \\ f_b(V_a, V_b) &= 0. \end{aligned} \quad (9)$$

The Jacobian matrix at the equilibrium point Q_i has the following form:

$$J(Q_i) = [a_{kl}(Q_i)]_{n \times n} = \begin{bmatrix} A_{aa}(Q_i) & A_{ab}(Q_i) \\ A_{ba}(Q_i) & A_{bb}(Q_i) \end{bmatrix}, \quad (10)$$

where $A_{kl}(Q_i)$ are called cell parameters and

$$\begin{aligned}
 A_{aa}(Q_i) &= \begin{bmatrix} \frac{\partial f_1}{\partial V_1} & \dots & \frac{\partial f_1}{\partial V_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_m} \end{bmatrix}, \\
 A_{ab}(Q_i) &= \begin{bmatrix} \frac{\partial f_1}{\partial V_{m+1}} & \dots & \frac{\partial f_1}{\partial V_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_{m+1}} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}, \\
 A_{ba}(Q_i) &= \begin{bmatrix} \frac{\partial f_{m+1}}{\partial V_1} & \dots & \frac{\partial f_{m+1}}{\partial V_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}, \\
 A_{bb}(Q_i) &= \begin{bmatrix} \frac{\partial f_{m+1}}{\partial V_{m+1}} & \dots & \frac{\partial f_{m+1}}{\partial V_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_{m+1}} & \dots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}.
 \end{aligned} \tag{11}$$

The local state equations at the cell equilibrium point Q_i are defined via

$$\begin{aligned}
 \dot{V}_a &= A_{aa}V_a + A_{ab}V_b + I_a, \\
 \dot{V}_b &= A_{ba}V_a + A_{bb}V_b.
 \end{aligned} \tag{12}$$

Definition 1.

$$Y_Q(s) = sI - A_{aa} - A_{ab}(sI - A_{bb})^{-1}A_{ba} \tag{13}$$

is called the admittance matrix at the cell equilibrium point Q_i .

Lemma 2. A reaction-diffusion CNN cell is called locally active at the equilibrium point Q_i if and only if, its admittance matrix at Q_i satisfies at least one of the following four conditions [4].

- (1) $Y_Q(s)$ has a pole in $\text{Re}[s] > 0$.
- (2) $Y_Q^H(i\omega) = \bar{Y}_Q(i\omega) + Y_Q(i\omega) < 0$ for some $\omega = \omega_0$, where ω_0 is any real number.
- (3) $Y_Q(s)$ has a simple pole $s = i\omega_p$ on the imaginary axis, where its associated residue matrix:

$$k_1 = \begin{cases} \lim_{s \rightarrow i\omega_p} (s - i\omega_p) Y_Q(s), & \text{if } \omega_p < \infty \\ \lim_{\omega_p \rightarrow \infty} \frac{Y_Q(i\omega_p)}{i\omega_p}, & \text{if } \omega_p = \infty \end{cases} \tag{14}$$

is either a complex number or a negative real number.

- (4) $Y_Q(s)$ has a multiple pole on the imaginary axis.

Definition 3. The cell equilibrium point Q_i is called stable if and only if, all the real parts of eigenvalue λ_i of Jacobian matrix at the equilibrium point Q_i are negative [2].

Definition 4. A “reaction-diffusion” CNN with n state variables and m ports is said to be operating on the “edge of chaos” with respect to an equilibrium point Q_i if and only if, Q_i is both locally active and stable when $I_a = 0$ [4].

Using the above lemma and definitions, the bifurcation of CNN with respect to an equilibrium point can be divided into three parts: the edge of chaos domains (the locally active and stable domains), the locally active and unstable domains, and the locally passive domains. Numerical simulations indicated that many complex dynamical behaviors, such as oscillatory patterns, chaotic patterns, or divergent patterns, may emerge if the selected cell parameters are located in or nearby the edge of chaos domains.

3. Analytical Criteria for Local Activity of CNN with Five State Variables and One Port

For the reaction-diffusion CNN with five state variables and one port, its local state equations have the form

$$\begin{aligned}
 \dot{V}_a &= A_{aa}V_a + A_{ab}V_b + I_a, \\
 \dot{V}_b &= A_{ba}V_a + A_{bb}V_b,
 \end{aligned} \tag{15}$$

where

$$V_a = [V_1], \quad V_b = [V_2 \ V_3 \ V_4 \ V_5]^T, \quad I_a = [I_1],$$

$$A_{aa} = [a_{11}], \quad A_{ab} = [a_{12} \ a_{13} \ a_{14} \ a_{15}],$$

$$A_{ba} = [a_{21} \ a_{31} \ a_{41} \ a_{51}]^T,$$

$$A_{ba} = \begin{bmatrix} a_{21} \\ a_{31} \\ a_{41} \\ a_{51} \end{bmatrix}, \quad A_{bb} = \begin{bmatrix} a_{22} & a_{23} & a_{24} & a_{25} \\ a_{32} & a_{33} & a_{34} & a_{35} \\ a_{42} & a_{43} & a_{44} & a_{45} \\ a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}. \tag{16}$$

The corresponding CNN cell admittance matrix $Y_Q(s)$ is given by [1].

$$\begin{aligned}
 Y_Q(s) &= sI - A_{aa} - A_{ab}(sI - A_{bb})^{-1}A_{ba} \\
 &= s - a_{11} - \frac{T_1s^3 + K_1s^2 + L_1s + \Delta_1}{s^4 + Ts^3 + Ks^2 + Ls + \Delta},
 \end{aligned} \tag{17}$$

where $T, T_1, K, K_1, L, L_1, \Delta, \Delta_1$ are the parameters of a_{ij} 's.

Theorem 5. A necessary and sufficient condition for $Y_Q(s)$ to satisfy condition (1) in Lemma 2 is that $\exists s$, such that $g(s) = 0$ ($\text{Re}[s] > 0$), and any one of the following conditions holds.

- (1) $f(s) \neq 0$.
- (2) $f(s) = 0$, and $m > n$, where s is m and n orders zero point of $g(s)$ and $f(s)$, respectively, where $f(s) = T_1s^3 + K_1s^2 + L_1s + \Delta_1$, $g(s) = s^4 + Ts^3 + Ks^2 + Ls + \Delta$.

Proof. Obviously proved. \square

Denote

$$E = -a_{11}, \quad F = -TT_1 + K_1 - a_{11}(-2K + T^2),$$

$$P = LL_1 - K\Delta_1 - \Delta K_1,$$

$$Q = \Delta_1 + KK_1 - TL_1 - LT_1,$$

$$G = -a_{11}(2\Delta + K^2 - 2LT) - Q,$$

$$H = -a_{11}(L^2 - 2\Delta K) - P,$$

$$I = -a_{11}\Delta^2 - \Delta\Delta_1,$$

$$g(Q) = EQ^4 + FQ^3 + GQ^2 + HQ + I,$$

$$h(\lambda) = -(TT_1 - K_1)\lambda^3 - Q\lambda^2 - P\lambda - \Delta\Delta_1,$$

$$\lambda_{1,2}^* = \frac{-Q \pm \sqrt{Q^2 - 3(TT_1 - K_1)P}}{3(TT_1 - K_1)},$$

$$p = -\frac{3F^2}{16E^2} + \frac{G}{2E}, \quad q = \frac{F^3}{32E^3} - \frac{FG}{8E^2} + \frac{H}{4E},$$

$$w_{1,2} = \frac{-1 \pm i\sqrt{3}}{2},$$

$$D = \frac{q^2}{4} + \frac{p^3}{27}, \quad A_j = \left(-\frac{q}{2} \pm D^{1/2}\right)^{1/3},$$

$$x_1 = A_1 + A_2, \quad x_2 = w_1A_1 + w_2A_2,$$

$$x_3 = w_2A_1 + w_1A_2, \quad \Omega_i = x_j - \frac{F}{4E}, \quad i = 1, 2, 3. \quad (18)$$

Theorem 6. Let the following parameters be defined as in Theorem 5, then $Y_Q^H(i\omega) < 0$ for some $w = w_0 \in R$ if any one of the following conditions holds.

- (1) $a_{11} > 0$.
- (2) $a_{11} = 0$, $TT_1 - K_1 > 0$.
- (3) $a_{11} = 0$, $TT_1 - K_1 = 0$, $Q > 0$.
- (4) $a_{11} = 0$, $TT_1 - K_1 = 0$, $Q < 0$, $\Delta\Delta_1 > 0$.
- (5) $a_{11} = 0$, $TT_1 - K_1 = 0$, $Q < 0$, $P \geq 0$, $\Delta\Delta_1 - P^2/Q/4 > 0$, $\Delta\Delta_1 \leq 0$,
- (6) $a_{11} = 0$, $TT_1 - K_1 = 0$, $Q = 0$, $P > 0$.
- (7) $a_{11} = 0$, $TT_1 - K_1 = 0$, $Q = 0$, $P \leq 0$, $\Delta\Delta_1 > 0$.
- (8) $a_{11} = 0$, $TT_1 - K_1 < 0$, $\Delta\Delta_1 > 0$.
- (9) $a_{11} = 0$, $TT_1 - K_1 < 0$, $\Delta\Delta_1 \leq 0$, and $\lambda_j^* \geq 0$, $h(\lambda_j^*) < 0$, for $j = 1$ or 2 .
- (10) $a_{11} < 0$, $D > 0$, $\Omega_1 > 0$, $g(\Omega_1) < 0$.
- (11) $a_{11} < 0$, $D < 0$, and $\Omega_j \geq 0$, $g(\Omega_j) < 0$, for $j = 1, 2$ or 3 .
- (12) $a_{11} < 0$, $D = 0$, $p = q = 0$, $g(-F/4E) < 0$.
- (13) $a_{11} < 0$, $D = 0$, $q^2/4 = -p^3/27 \neq 0$, and $\Omega_j \geq 0$, $g(\Omega_j) < 0$, for $j = 1$ or 2 .

Proof. $Y_Q^H(i\omega) = \bar{Y}_Q(i\omega) + Y_Q(i\omega) = 2 \operatorname{Re}[Y_Q(i\omega)]$, so $Y_Q(i\omega)$ to satisfy condition (2) in Lemma 2 equals to $\operatorname{Re}[Y_Q(i\omega)] < 0$,

$$\begin{aligned} \operatorname{Re}[Y_Q(i\omega)] &= \operatorname{Re} \left[i\omega - a_{11} \right. \\ &\quad \left. - \frac{T_1(i\omega)^3 + K_1(i\omega)^2 + L_1(i\omega) + \Delta_1}{(i\omega)^4 + T(i\omega)^3 + K(i\omega)^2 + L(i\omega) + \Delta} \right] \\ &= \frac{E\omega^8 + F\omega^6 + G\omega^4 + H\omega^2 + I}{(\omega^4 - K\omega^2 + \Delta)^2 + (L\omega - T\omega^2)}. \end{aligned} \quad (19)$$

(1) If $a_{11} > 0$, then $\operatorname{Re}[Y_Q(i\omega)] < 0$ when ω is large enough (See (1) of Theorem 6).

(2) If $a_{11} = 0$, then

$$\begin{aligned} \operatorname{Re}[Y_Q(i\omega)] &= -\frac{(TT_1 - K_1)\omega^6 + Q\omega^4 + P\omega^2 + \Delta\Delta_1}{(\omega^4 - K\omega^2 + \Delta)^2 + (L\omega - T\omega^2)}. \end{aligned} \quad (20)$$

Let $f(\lambda) = -Q\lambda^2 - P\lambda - \Delta\Delta_1$,

(I) If $TT_1 - K_1 > 0$, then $\operatorname{Re}[Y_Q(i\omega)] < 0$ when ω is large enough (See (2) of Theorem 6).

(II) If $TT_1 - K_1 = 0$, then

(i) If $Q > 0$, then $\operatorname{Re}[Y_Q(i\omega)] < 0$ when ω is large enough (See (3) of Theorem 6).

(ii) If $Q < 0$,

(a) If $\Delta\Delta_1 > 0$, $\exists \omega_0 \in R$, such that $\operatorname{Re}[Y_Q(i\omega_0)] < 0$ (See (4) of Theorem 6).

(b) If $\Delta\Delta_1 \leq 0$, solve $f'(\lambda^*) = 0$, we can get $\lambda^* = -0.5P/Q$, $f(\lambda^*) = 0.25P^2/Q - \Delta\Delta_1$. Then when $P \geq 0$, $\Delta\Delta_1 - 0.25P^2/Q > 0$, $\exists \omega_0 > \lambda^*$, such that $\operatorname{Re}[Y_Q(i\omega_0)] < 0$ (See (5) of Theorem 6).

(iii) If $Q = 0$, then $\operatorname{Re}[Y_Q(i\omega)] = -P\omega^2 - \Delta\Delta_1$.

(a) If $P > 0$, then $\operatorname{Re}[Y_Q(i\omega)] < 0$ when ω is large enough (See (6) of Theorem 6).

(b) If $P \leq 0$, $\Delta\Delta_1 > 0$, then $\exists \omega_0$, such that $\operatorname{Re}[Y_Q(i\omega_0)] = -\Delta\Delta_1 < 0$ (See (7) of Theorem 6).

(III) If $TT_1 - K_1 < 0$, let $h(\lambda) = -(TT_1 - K_1)\lambda^3 - Q\lambda^2 - P\lambda - \Delta\Delta_1$,

(i) If $\Delta\Delta_1 > 0$, then $\exists \omega_0$, such that $\operatorname{Re}[Y_Q(i\omega_0)] < 0$ (See (8) of Theorem 6).

(ii) If $\Delta\Delta_1 \leq 0$, solve $h(\lambda^*) = 0$, we can get

$$\lambda_{1,2}^* = \frac{-Q \pm \sqrt{Q^2 - 3(TT_1 - K_1)P}}{3(TT_1 - K_1)}. \quad (21)$$

Then, for $i = 1, 2$, if $\lambda_i^* \geq 0$, $h(\lambda_i^*) < 0$, then $\exists \omega_0$, such that $\operatorname{Re}[Y_Q(i\omega_0)] < 0$ (See (9) of Theorem 6).

- (3) If $a_{11} < 0$, let $g(Q) = EQ^4 + FQ^3 + GQ^2 + HQ + I$, then $g'(Q) = 4EQ^3 + 3FQ^2 + 2GQ + H$. Let $x = \Omega + (F/4E)$, then the above becomes $g'(Q) = 4E(x^3 + px + q) = 4Ef(x)$, then x_i , $i = 1, 2, 3$ are the roots of $f(x) = 0$, Ω_i are the roots of $g'(\Omega) = 0$. If any one of the (10)–(13) of Theorem 6 holds, we can get $\text{Re}[Y_Q(i\omega_0)] < 0$.

So, if any one of conditions (1)–(13) holds, $\text{Re}[Y_Q(i\omega_0)] < 0$. $Y_Q(s)$ Satisfies condition (2) in Lemma 2. This completes the proof. \square

Theorem 7. For $j = 1$, or 2, let

$$\begin{aligned} w_j &= \frac{(\sqrt{K+2\sqrt{\Delta}} + (-1)^j \sqrt{K-2\sqrt{\Delta}})}{2}, \\ A_j &= L - 3w_j^{*2}, \\ w_j^* &= \sqrt{\frac{(K + (-1)^j \sqrt{K^2 - 4\Delta})}{2}}, \\ B_j &= 2Kw_j^* - 4w_j^{*3}, \quad A_{1j} = \Delta_1 - K_1w_j^{*2}, \\ B_{1j} &= L_1w_j^* - T_1w_j^{*3}. \end{aligned} \quad (22)$$

Then $Y_Q(s)$ satisfies condition (3) of Lemma 2, if any one of the following conditions holds.

- (I) $\Delta > 0$, $K > 2\sqrt{\Delta}$, $T = L = 0$, and any one of the following conditions holds.

- (1) $K_1w_1^2 - \Delta_1 \neq 0$.
- (2) $K_1w_1^2 - \Delta_1 = 0$, $(L_1 - T_1w_1^2)(w_2^2 - w_1^2) > 0$.
- (3) $K_1w_2^2 - \Delta_1 \neq 0$.
- (4) $K_1w_2^2 - \Delta_1 = 0$, $(L_1 - T_1w_2^2)(w_1^2 - w_2^2) > 0$.

- (II) $K > 0$, $\Delta_1 \neq 0$, $\Delta = 0$, $L = KT \neq 0$, and any one of the following conditions holds.

- (1) $T\Delta_1 < 0$.
- (2) $T(K - T_1K) - \Delta_1 + KK_1 \neq 0$.
- (3) $T(K - T_1K) - \Delta_1 + KK_1 = 0$, $T(\Delta_1 - KK_1) + K(L_1 - T_1K) < 0$.

- (III) $\Delta = 0$, $\Delta_1L > 0$.

- (IV) $\Delta < 0$, or $K > 0$, $K^2 - 4\Delta > 0$, and $2L = T(K + \sqrt{K^2 - 4\Delta})(K + \sqrt{K^2 - 4\Delta} > 0)$ or $2L = T(K - \sqrt{K^2 - 4\Delta})(K - \sqrt{K^2 - 4\Delta} > 0)$ and any one of the following conditions holds for $j = 1$, or 2.

- (1) $A_jB_{1j} - A_{1j}B_j \neq 0$.
- (2) $A_jB_{1j} - A_{1j}B_j = 0$, $A_jA_{1j} - B_jB_{1j} > 0$.

Proof. Let $f(s) = T_1s^3 + K_1s^2 + L_1s + \Delta_1$, $g(s) = s^4 + Ts^3 + Ks^2 + Ls + \Delta$, obviously, ∞ is not a single pole of $Y_Q(s)$ on the imaginary axis.

If $Y_Q(s)$ has a simple pole $s = i\omega$ on the imaginary axis, where its associated residue

$$k_1 = \lim_{s \rightarrow i\omega} (s - i\omega)Y_Q(s) = \lim_{s \rightarrow i\omega} (s - i\omega) \frac{f(s)}{g(s)} = \frac{f(s)}{g'(s)} \Big|_{s=i\omega} \quad (23)$$

is either a complex number or a negative real number, then $k_1 \neq 0$, so $f(s) \neq 0$, which implies that $i\omega$ is not a zero point of $f(s) = 0$, $i\omega$ is not a removed pole of $Y_Q(s)$.

- (I) If $Y_Q(s)$ has four poles $s = \pm i\omega_1, \pm i\omega_2$ ($\omega_1 \neq \omega_2 \neq 0$) on the imaginary axis. In this case, $g(s) = (s^2 + \omega_1^2)(s^2 + \omega_2^2) = s^4 + (\omega_1^2 + \omega_2^2)s^2 + \omega_1^2\omega_2^2$. Hence we obtain $T = L = 0$, $K = \omega_1^2 + \omega_2^2 > 0$, $\Delta = \omega_1^2\omega_2^2 > 0$. Then, we can get $K + 2\sqrt{\Delta} = (\omega_1 + \omega_2)^2$, $K - 2\sqrt{\Delta} = (\omega_1 - \omega_2)^2$, which implies that $K > 2\sqrt{\Delta}$, $\omega_{1,2} = (\sqrt{K + 2\sqrt{\Delta}} \pm \sqrt{K - 2\sqrt{\Delta}})/2$. So,

$$\lim_{s \rightarrow \pm i\omega_1} (s \mp i\omega_1)Y_Q(s) = \frac{w_1(L_1 - T_1w_1^2) \pm i(K_1w_1^2 - \Delta_1)}{2w_1(w_2^2 - w_1^2)}, \quad (24)$$

$$\lim_{s \rightarrow \pm i\omega_2} (s \mp i\omega_2)Y_Q(s) = \frac{w_2(L_1 - T_1w_2^2) \pm i(K_1w_2^2 - \Delta_1)}{2w_2(w_1^2 - w_2^2)}. \quad (25)$$

Then, when condition (I) in Theorem 7 holds, k_1 is a complex number or a negative real number. $Y_Q(s)$ satisfies condition (3) in Lemma 2.

- (II) If $Y_Q(s)$ has a simple pole $s = 0$ and two conjugate poles $\pm i\omega$ ($\omega \neq 0$) on the imaginary axis, and another pole is $a \neq 0$.

In this case, it follows that $\Delta = 0$, $\Delta_1 \neq 0$, and $g(s)$ has the form:

$$\begin{aligned} g(s) &= s(s^2 + \omega^2)(s - a) \\ &= s^4 - as^3 + \omega^2s^2 - a\omega^2, \end{aligned} \quad (26)$$

which implies that $T = -a$, $K = \omega^2 > 0$, $L = -a\omega^2 = KT$, $\Delta = 0$, $\Delta_1 = 0$, Therefore,

$$g(s) = s(s^2 + K)(s + T) \quad (27)$$

- (1) The residue of $Y_Q(s)$ at $s = 0$ is

$$\lim_{s \rightarrow 0} sY_Q(s) = \frac{\Delta_1}{(s^2 + K)(s + T)} = \frac{\Delta_1}{KT}. \quad (28)$$

Then, we conclude that if $K > 0$, $\Delta_1 \neq 0$, $\Delta = 0$, $L = KT \neq 0$, $T\Delta_1 < 0$, k_1 is a negative real number. $Y_Q(s)$ satisfies condition (3) in Lemma 2 (See (1) of (II) in Theorem 7).

- (2) The residue of $Y_Q(s)$ at $s = \pm i\sqrt{K}$ is

$$\lim_{s \rightarrow \pm i\sqrt{K}} (s \mp i\sqrt{K}) Y_Q(s) = \frac{T(\Delta_1 - KK_1) + K(L_1 - T_1K) \mp i\sqrt{K}(T(L_1 - T_1K) - \Delta_1 + KK_1)}{2K(K + T^2)}. \quad (29)$$

Consequently, we conclude that if (2) or (3) in (II) in Theorem 7 holds, k_1 is either an imaginary number or a negative real number. $Y_Q(s)$ satisfies condition (3) in Lemma 2.

(III) If $Y_Q(s)$ has a simple pole $s = 0$ on the imaginary axis, and the other poles are a_i , $\text{Re}[a_i] \neq 0$, $i = 1, 2, 3$, it follows that $\Delta_1 \neq 0$, $\Delta = 0$, and $g(s)$ has the form

$$\begin{aligned} g(s) &= s(s - a_1)(s - a_2)(s - a_3) \\ &= s(s^3 - (a_1 + a_2 + a_3)s^2 \\ &\quad + (a_1a_2 + a_1a_3 + a_2a_3)s - a_1a_2a_3). \end{aligned} \quad (30)$$

Therefore we obtain that $\Delta = 0$, $T = -(a_1 + a_2 + a_3)$, $K = a_1a_2 + a_1a_3 + a_2a_3$, $L = -a_1a_2a_3 \neq 0$, hence the residue of $Y_Q(s)$ at $s = 0$ is

$$\lim_{s \rightarrow 0} s Y_Q(s) = \frac{\Delta_1}{a_1a_2a_3} = -\frac{\Delta_1}{L}. \quad (31)$$

Then, when $\Delta = 0$, $\Delta_1 L > 0$, k_1 is a negative real number. $Y_Q(s)$ satisfies condition (3) in Lemma 2 (See (III) of Theorem 7).

(IV) If $Y_Q(s)$ has two conjugate poles $\pm i\omega$ ($\omega > 0$) on the imaginary axis, and the other poles are $\text{Re}[a] \neq 0$, $\text{Re}[b] \neq 0$. In this case, $g(s)$ has the form

$$\begin{aligned} g(s) &= (s - a)(s - b)(s + \omega^2) \\ &= s^4 - (a + b)s^3 + (ab + \omega^2)s^2 \\ &\quad - (a + b)\omega^2s + ab\omega^2. \end{aligned} \quad (32)$$

Therefore, we obtain that $T = -(a + b)$, $K = ab + \omega^2$, $L = -(a + b)\omega^2$, $\Delta = ab\omega^2 \neq 0$. Then, $ab = K - \omega^2$, $\Delta = (K - \omega^2)\omega^2 \Leftrightarrow \omega^4 - K\omega^2 + \Delta = 0$. Solving it, we have

$$\omega_j^* = \sqrt{\frac{K \pm \sqrt{K^2 - 4\Delta}}{2}}, \quad j = 1, 2. \quad (33)$$

It implies that $\Delta < 0$ or $K > 0$, $K^2 - 4\Delta \geq 0$, and $T = -(a + b) = L/\omega^2$. Then, the residue of $Y_Q(s)$ at $s = \pm i\omega_j^*$ is

$$\begin{aligned} \lim_{s \rightarrow \pm i\omega_j^*} (s \mp i\omega_j^*) Y_Q(s) \\ = \frac{-(A_{1j}A_j + B_{1j}B_j) \pm i(A_{1j}B_j - A_jB_{1j})}{A_j^2 + B_j^2}. \end{aligned} \quad (34)$$

Hence, if condition (IV) in Theorem 7 holds, $k_1 < 0$ or it is an imaginary number. $Y_Q(s)$ satisfies condition (3) in Lemma 2.

Therefore, when any one of conditions (I)–(IV) holds, $Y_Q(s)$ satisfies condition (3) in Lemma 2. This completes the proof. \square

Theorem 8. $Y_Q(s)$ has a multiple pole on the imaginary axis if any one of the following conditions holds.

- (I) $\Delta = L = 0$, $\Delta_1 \neq 0$.
- (II) $\Delta = L = K = 0$, and $\Delta_1 \neq 0$ or $L_1 \neq 0$.
- (III) $\Delta = L = K = T = 0$, and $\Delta_1 \neq 0$ or $L_1 \neq 0$ or $K_1 \neq 0$.
- (IV) $T = L = 0$, $K > 0$, $\Delta = (K/2)^2$, and any one of the following conditions holds.

- (1) $2\Delta_1 \neq KK_1$.
- (2) $2L_1 \neq KT_1$.

Proof. Let $f(s) = T_1s^3 + K_1s^2 + L_1s + \Delta_1$, $g(s) = s^4 + Ts^3 + Ks^2 + Ls + \Delta$. Obviously, when the conditions (I)–(III) hold, 0 is the multiply poles of $Y_Q(s)$.

If $Y_Q(s)$ has two multiply nonzero poles $\pm i\omega$ ($\omega > 0$), then $g(s)$ has the form:

$$g(s) = (s^2 + \omega^2)^2 = s^4 + 2\omega^2s^2 + \omega^4, \quad (35)$$

which implies that $T = L = 0$, $K = 2\omega^2 > 0$, $\Delta = \omega^4 = (K/2)^2 > 0$.

If $\pm i\omega$ are the multiply poles of $Y_Q(s)$, then $f(i\omega) \neq 0$ where

$$\begin{aligned} f(\pm i\omega) &= (\Delta_1 - K_1\omega^2) \pm i\omega(L_1 - T_1\omega^2) \\ &= \left(\Delta_1 - \frac{KK_1}{2}\right) \pm i\omega\left(L_1 - \frac{T_1K}{2}\right). \end{aligned} \quad (36)$$

Obviously, when any one of (1)–(2) of (IV) in Theorem 8 holds, $f(i\omega) \neq 0$, any one of $\pm i\omega$ is a multiply pole of $Y_Q(s)$.

So, when any one of condition (I)–(IV) holds, $Y_Q(s)$ satisfies condition (4) in Lemma 2. This completes the proof.

When any one of Theorems 5–8 holds, $Y_Q(s)$ satisfies Lemma 2, which implies that the reaction-diffusion CNN with five state variables and one port at the equilibrium point is active.

These theorems can be implemented by a computer program for calculating the bifurcation diagram of the general corresponding CNN to determine emergence of complex dynamic patterns of the corresponding CNN. \square

TABLE 1: Cell parameters and corresponding dynamic properties of the reaction-diffusion CNN of HBV mutation-selection of HBV infection.

| No. | u | k | Equilibrium point | Eigenvalues | Dynamic pattern |
|-----|-----|-----|-------------------|---|-----------------------|
| 1 | 2 | 1.0 | 20,0,0,19,98 | 69.4396, 29.8638, -0.0090, -33.3646, -71.9897 | Convergent, divergent |
| 2 | 2 | 3.0 | 20,0,0,20,98 | -0.0097, 53.0149, 69.4400, -56.5150, -71.9902 | Convergent, divergent |
| 3 | 2 | 4.9 | 20,0,0,20,98 | -0.0098, 68.2343, 69.4485, -71.6962, -72.0368 | Convergent, divergent |
| 4 | 2 | 5.1 | 20,20,50,0,0 | -0.2043 ± 0.3878 <i>i</i> , -0.0000, -2.6014, -2.5000 | Convergent |
| 5 | 2 | 10 | 10,20,99,0,0 | -2.7130, -0.3935 ± 0.4583 <i>i</i> , -0.2192, -2.2808 | Convergent |
| 6 | 2 | 24 | 4,20,239,0,0 | -3.2812, -0.8094 ± 0.2709 <i>i</i> , -0.3768, -2.1232 | Convergent |
| 7 | 2 | 39 | 3,20,389,0,0 | -4.4393, -1.2723, -0.6884, -0.4059, -2.0941 | Convergent |
| 8 | 5 | 1.0 | 50,0,0,19,38 | 67.9709, 28.4103, -0.0075, -34.9127, -73.5210 | Convergent, divergent |
| 9 | 5 | 3.0 | 50,0,0,19,38 | -0.0092, 51.5420, 67.9713, -58.0426, -73.5215 | Convergent, divergent |
| 10 | 5 | 4.9 | 50,0,0,19,38 | -0.0095, 66.7552, 67.9801, -73.2207, -73.5651 | Convergent, divergent |
| 11 | 5 | 5.1 | 49,19,19,0,0 | -0.0920 ± 0.2787 <i>i</i> , -0.0092, -5.5160, -5.4909 | Convergent |
| 12 | 5 | 10 | 25,19,39,0,0 | -0.1829 ± 0.3778 <i>i</i> , -0.2375, -5.5343, -5.2625 | Convergent |
| 13 | 5 | 24 | 10,20,95,0,0 | -5.5708, -0.4446 ± 0.4784 <i>i</i> , -0.3915, -5.1085 | Convergent |
| 14 | 5 | 39 | 6,20,155,0,0 | -5.6298, -0.7151 ± 0.4210 <i>i</i> , -0.4343, -5.0657 | Convergent |
| 15 | 9 | 1.0 | 90,0,0,18,20 | 66.0620, 26.5823, -0.0055, -37.0867, -75.6121 | Convergent, divergent |
| 16 | 9 | 3.0 | 90,0,0,18,20 | -0.0085, 49.6418, 66.0624, -60.1432, -75.6126 | Convergent, divergent |
| 17 | 9 | 4.9 | 90,0,0,18,20 | -0.0091, 64.8330, 66.0717, -75.3029, -75.6527 | Convergent, divergent |
| 18 | 9 | 5.1 | 88,18,10,0,0 | -0.0531 ± 0.2110 <i>i</i> , -0.0106, -9.5037, -9.4894 | Convergent |
| 19 | 9 | 10 | 45,19,21,0,0 | -0.1047 ± 0.2973 <i>i</i> , -0.2431, -9.5106, -9.2569 | Convergent |
| 20 | 9 | 24 | 19,20,52,0,0 | -0.2481 ± 0.4288 <i>i</i> , -9.5339, -0.3897, -9.1103 | Convergent |
| 21 | 9 | 39 | 12,20,86,0,0 | -0.4014 ± 0.4931 <i>i</i> , -9.5671, -0.4300, -9.0700 | Convergent |

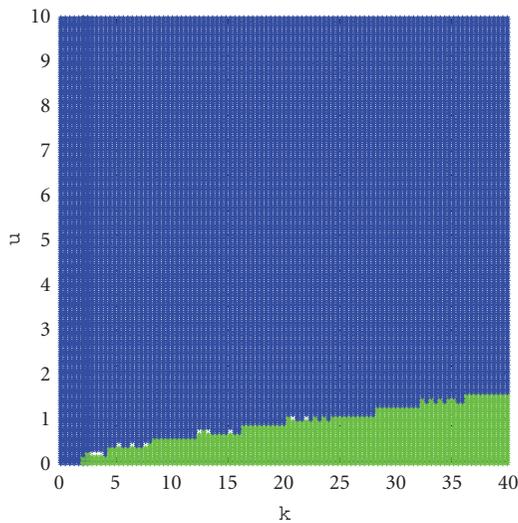


FIGURE 1: Bifurcation diagrams of equation (38) at the equilibrium points Q_1 at $k \in [0, 40]$, $u \in [0, 10]$.

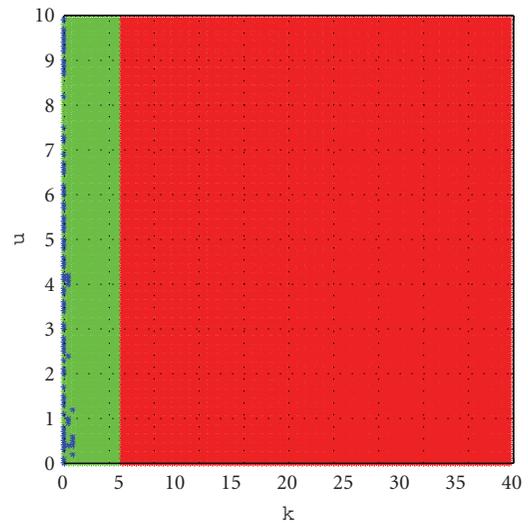


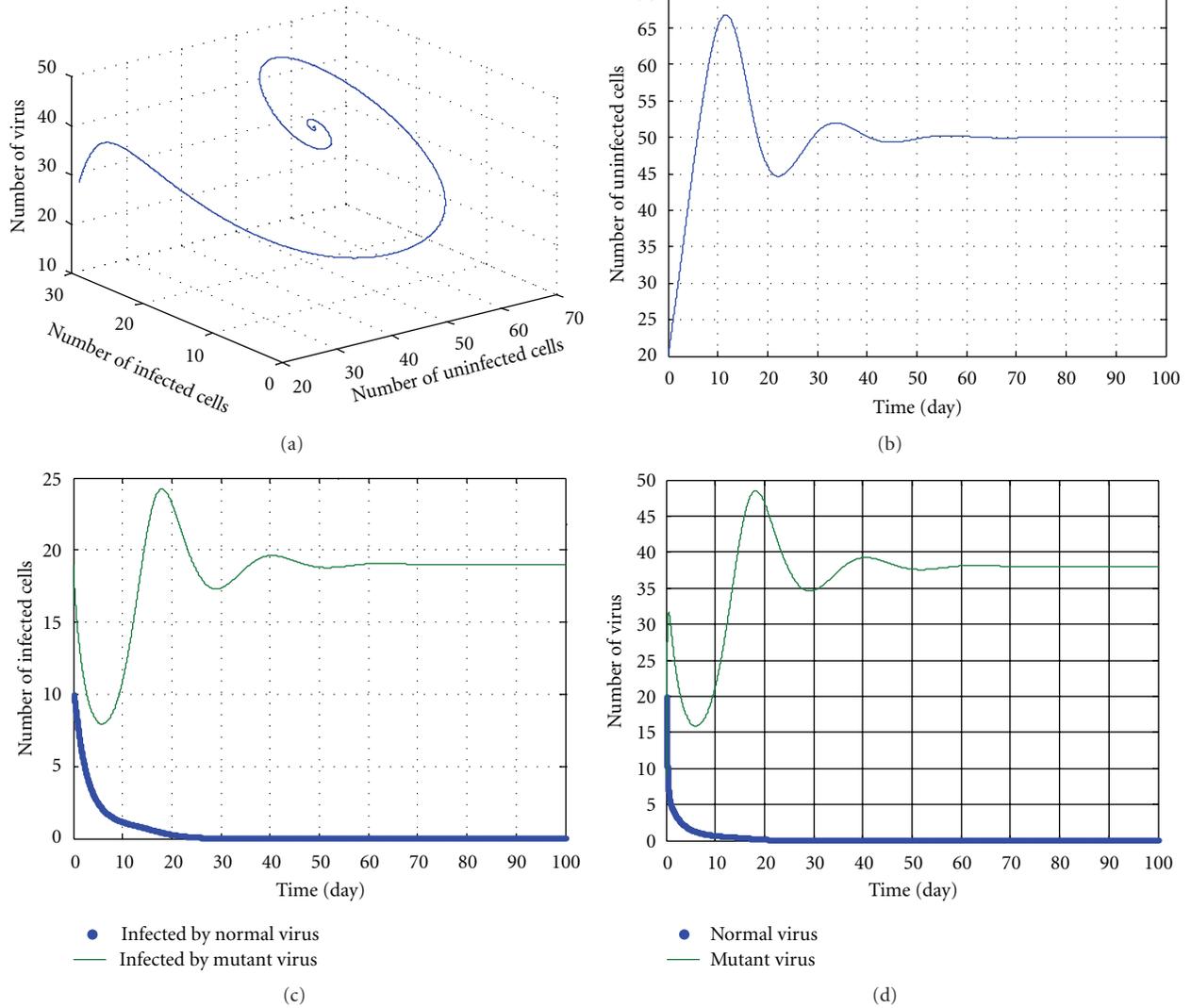
FIGURE 2: Bifurcation diagrams of equation (38) at the equilibrium points Q_2 at $k \in [0, 40]$, $u \in [0, 10]$.

4. Analysis and Simulations of Reaction-Diffusion CNN of HBV Mutation-Selection Model

Life systems consist of locally coupled homogeneous media. Mostly, dynamics of life systems are suitable to be described

via locally connected reaction-diffusion CNNs. It may be expected that reaction-diffusion CNN will become a promising candidate for modeling life phenomena.

In Chapter 11 “Timing the emergence of resistance” (Page 110) of the book “Virus dynamic: mathematical principles of immunology and virology” (Oxford university press),

FIGURE 3: The kinetic trajectories of equation (37) when $u = 5, k = 3$.

Nowak et al. proposed a mathematical model which describes the mutation selection of HBV infection during the therapy [31]:

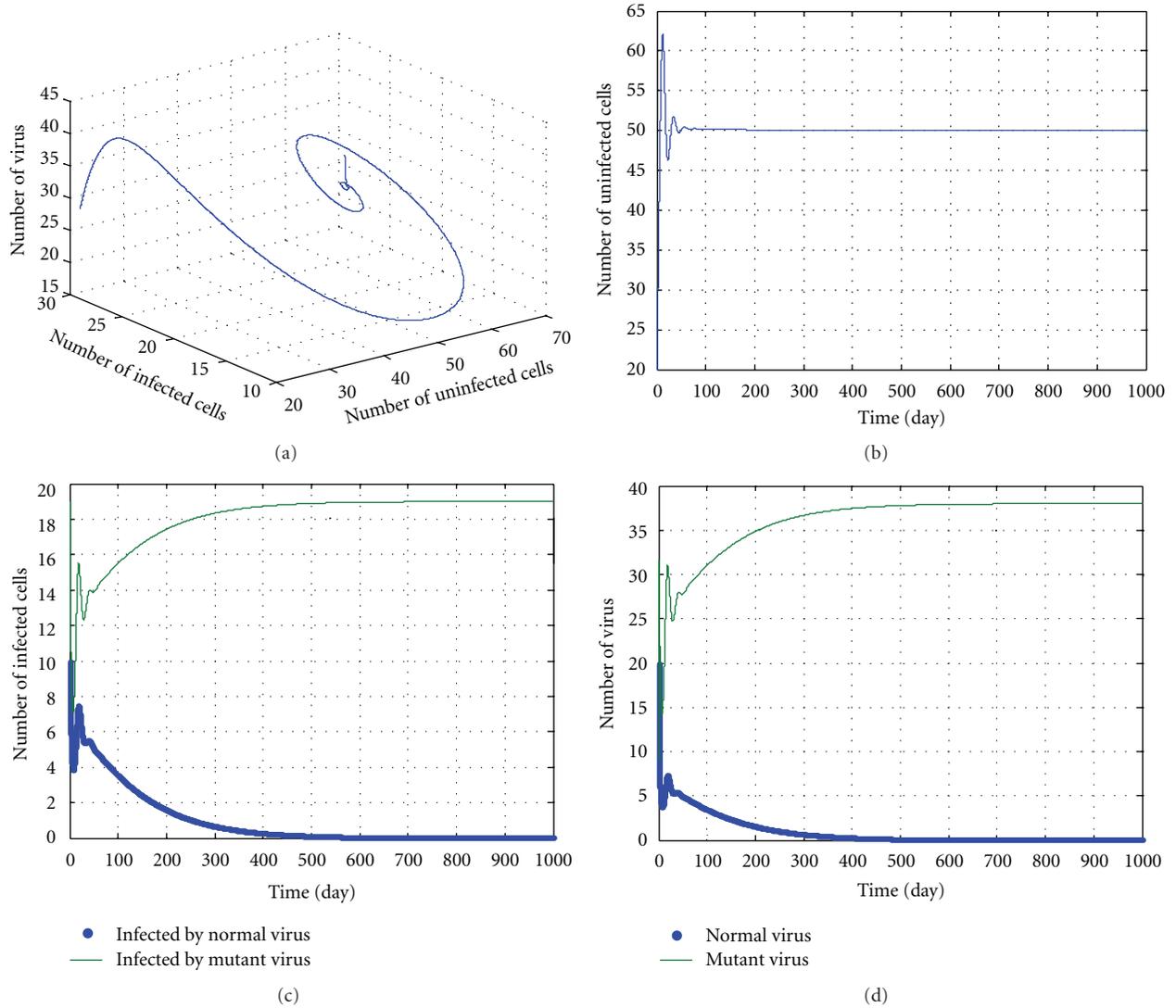
$$\begin{aligned}
 \frac{dx}{dt} &= \lambda - dx - bvx - b_n xv_n, \\
 \frac{dy}{dt} &= b(1-e)vx - ay, \\
 \frac{dv}{dt} &= ky - uv, \\
 \frac{dy_n}{dt} &= bev_x + b_n xv_n - ay_n, \\
 \frac{dv_n}{dt} &= k_n y_n - uv_n,
 \end{aligned} \tag{37}$$

where the five variables— x , y , v , y_n , v_n represent the numbers of uninfected cells, infected cells infected by normal virus, normal virus, infected cells infected by mutated virus, and mutant viruses, respectively. λ is the rate of reproduction

of uninfected cells. Uninfected cells die at rate dx and become infected at rate $b xv$ by normal virus and infected at rate $b_n xv_n$ by mutated virus. Infected cells infected by normal and mutated virus are removed at rate ay and ay_n , respectively. Normal virus is produced at rate ky and removed at rate uv , mutated virus is produced at rate ky_n and removed at rate uv_n . e is the rate constant describing the probability of mutation of virus (usual 10^{-5} – 10^{-3}), a , b , b_n , d , e , k , k_n , u , λ are positive constants. The model was briefly analyzed in Nowak's book.

The reaction-diffusion CNN of HBV mutation selection of model has the form:

$$\begin{aligned}
 \frac{dx_{ij}}{dt} &= \lambda - dx_{ij} - bx_{ij}v_{ij} - b_n x_{ij}v_{nij} + D_1 \nabla^2 x_{ij}, \\
 \frac{dy_{ij}}{dt} &= b(1-e)x_{ij}v_{ij} - ay_{ij}, \\
 \frac{dv_{ij}}{dt} &= ky_{ij} - uv_{ij},
 \end{aligned}$$


 FIGURE 4: The kinetic trajectories of equation (37) when $u = 5$, $k = 4.9$.

$$\begin{aligned} \frac{dy_{nij}}{dt} &= bex_{ij}v_{ij} + b_n x_{ij}v_{nij} - ay_{nij}, \\ \frac{dv_{nij}}{dt} &= k_n y_{nij} - uv_{nij}, \end{aligned} \quad (38)$$

where $\nabla^2 x_{ij} = x_{i-1j} + x_{i+1j} + x_{ij-1} + x_{ij+1} - 4x_{ij}$.

Let equation (38) be zeros ($D_1 = 0$) and solve it, we can get the two equilibrium points:

$$Q_1 = \left(\frac{\lambda}{d}, 0, 0, 0, 0 \right), \quad (39)$$

$$Q_2 = \left(x_0, \frac{u(au - b_n k_n u x_0)(\lambda - dx_0)}{bk_n x_0 (b_n k_n e x_0 + au - k_n x_0)}, \frac{(au - b_n k_n u x_0)(\lambda - dx_0)}{bx_0 (b_n k_n e x_0 + au - k_n x_0)}, \frac{eu(\lambda - dx_0)}{b_n k_n e x_0 + au - k_n x_0}, \frac{k_n e(\lambda - dx_0)}{b_n k_n e x_0 + au - k_n x_0} \right), \quad (40)$$

where $x_0 = au/((1-e)bk)$ and Q_1, Q_2 stand for the patient's complete recovery and HBV persistent infection, respectively.

Consequently, the Jacobian matrix at the equilibrium point Q_i ($i = 1, 2$) is

$$J(Q_i) = \begin{bmatrix} -d - bv - b_n v_n & 0 & -bx & 0 & -b_n x \\ b(1-e)v & -a & b(1-e)x & 0 & 0 \\ 0 & k & -u & 0 & 0 \\ bev + b_n v_n & 0 & bex & -a & b_n x \\ 0 & 0 & 0 & k_n & -u \end{bmatrix}. \quad (41)$$

Taking k, u as variables, and $\lambda = 10$, $a = 0.5$, $b = 0.01$, $b_n = 0.005$, $e = 0.0001$, $k_n = 10$, and $d = 0.01$, using Theorems 5–8, we can calculate the bifurcation of the reaction-diffusion CNN model equation (38) at the equilibrium

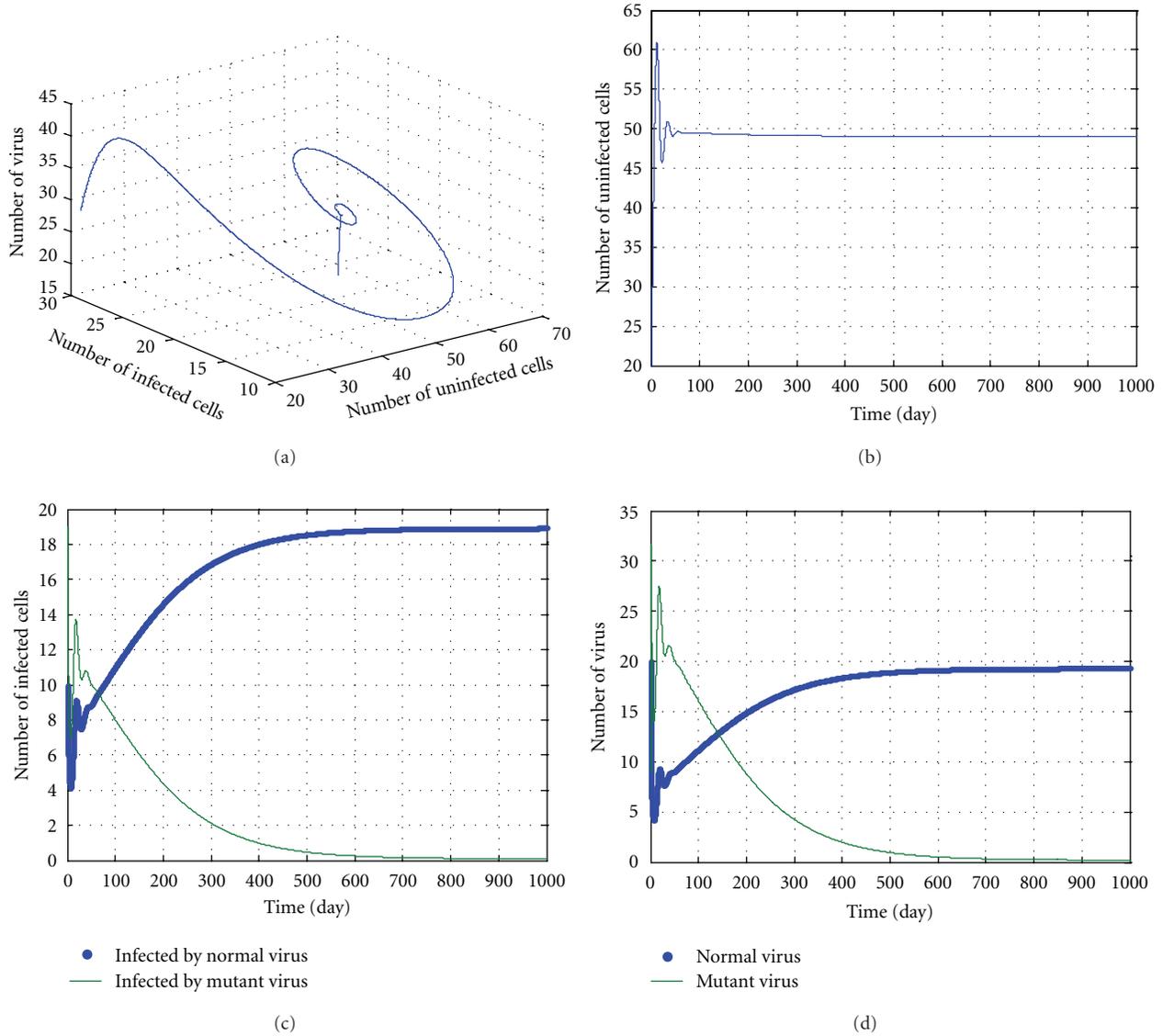


FIGURE 5: The kinetic trajectories of equation (37) when $u = 5$, $k = 5.1$.

points Q_1 and Q_2 at $k \in [0, 40]$, $u \in [0, 10]$, see Figures 1 and 2.

In Figures 1 and 2, the domains are coded as follows: edge of chaos (locally active and stable) domain (shown red), locally active and unstable domain (shown green) and locally passive domain (shown blue). From Figure 1(a), we can see that the bifurcation at equilibrium point Q_1 does not exist at the edge of chaos domain.

Take $\lambda = 10$, $k = 0.01$, $a = 0.5$, $b = 0.01$, $b_n = 0.005$, $k_n = 10$, $e = 0.0001$, and $k = 1.0, 3.0, 4.9, 5.1, 10, 24, 39$, $u = 2, 5, 9$, we model the dynamic trajectories of equation (37) using MATLAB, see Table 1.

In the following discussions, we select some parameters in No. 8-14 and $u = 5$, $k = 12.5$. The simulation results are shown in Figures 3, 4, 5, 6, and 7. During the simulation, we reached a new conclusion.

From Table 1 and Figures 3–7, we can conclude that

(I) when k is smaller (less than 5),

- (1) these parameters are located in the green domain (the local and unstable domains);
- (2) regardless of the value of u , the dynamic pattern of equation (37) is convergent or divergent depending on initial values;
- (3) the No. 1, No. 4, and No. 5 variables in equation (37) increase and the No. 2 and No. 3 variables in equation (37) decrease to 0. This means the numbers of the mutant virus and of infected cells infected by mutant virus both increase, and the numbers of normal virus and of the infected cells infected by normal virus both decrease, even to near zero. Also, the

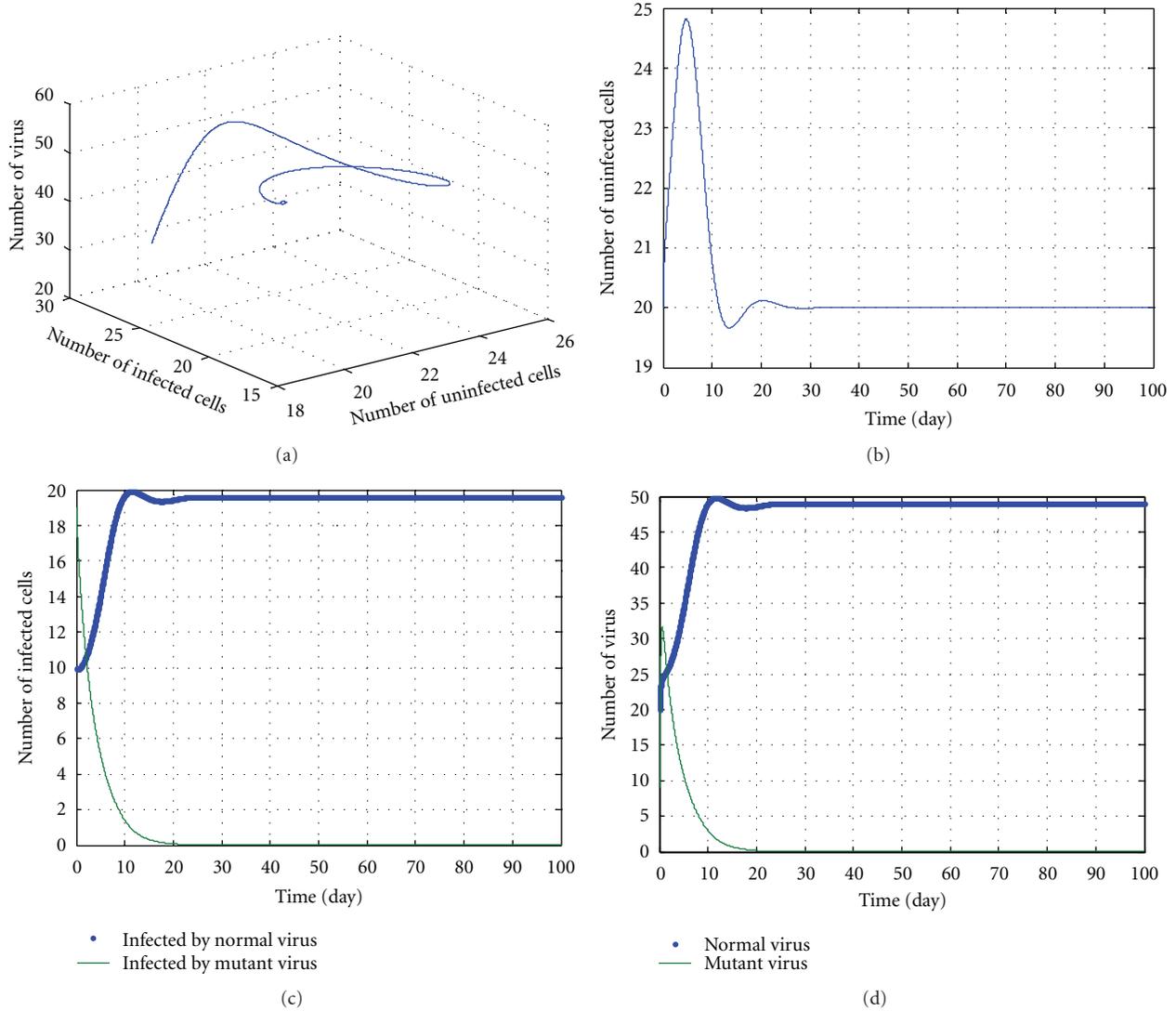


FIGURE 6: The kinetic trajectories of equation (37) when $u = 5$, $k = 12.5$.

No. 1 variable in equation (37), that is, the number of uninfected cells increases as compared to the initial number. All these indicate that the potency is perfect except for some virus mutation. The potency is ideal.

(II) When k is larger (greater than 5) but less than a threshold value (according to initial values and parameters, for example 12.5 in Figure 6), we can conclude the following.

- (1) These parameters are located in the red domain (edge of chaos).
- (2) Regardless of the value of u , the dynamic pattern of equation (37) is convergent.
- (3) The No. 1, No. 2, and No. 3 variables in equation (37) increase and No. 4 and No. 5 variables decrease to 0. This means that the number of uninfected cells, the numbers of the normal

virus, and of the cells infected by normal virus all increase. Meanwhile, the numbers of the mutant virus and of the cells infected by mutant virus both decrease, even to near zero. All these imply that the drug cannot clean the normal virus, but can destroy the mutant virus and increase the infection cells. The potency is also ideal.

(III) When $k < 40$ and greater than a threshold value (according to initial values and parameters),

- (1) these parameters are located in the red domain (edge of chaos);
- (2) regardless of the u value, the dynamic pattern of equation (37) is convergent;
- (3) The No. 2 and No. 3 variables in equation (37) increase and No. 1, No. 4, and No. 5 variables decrease, which means the numbers of the

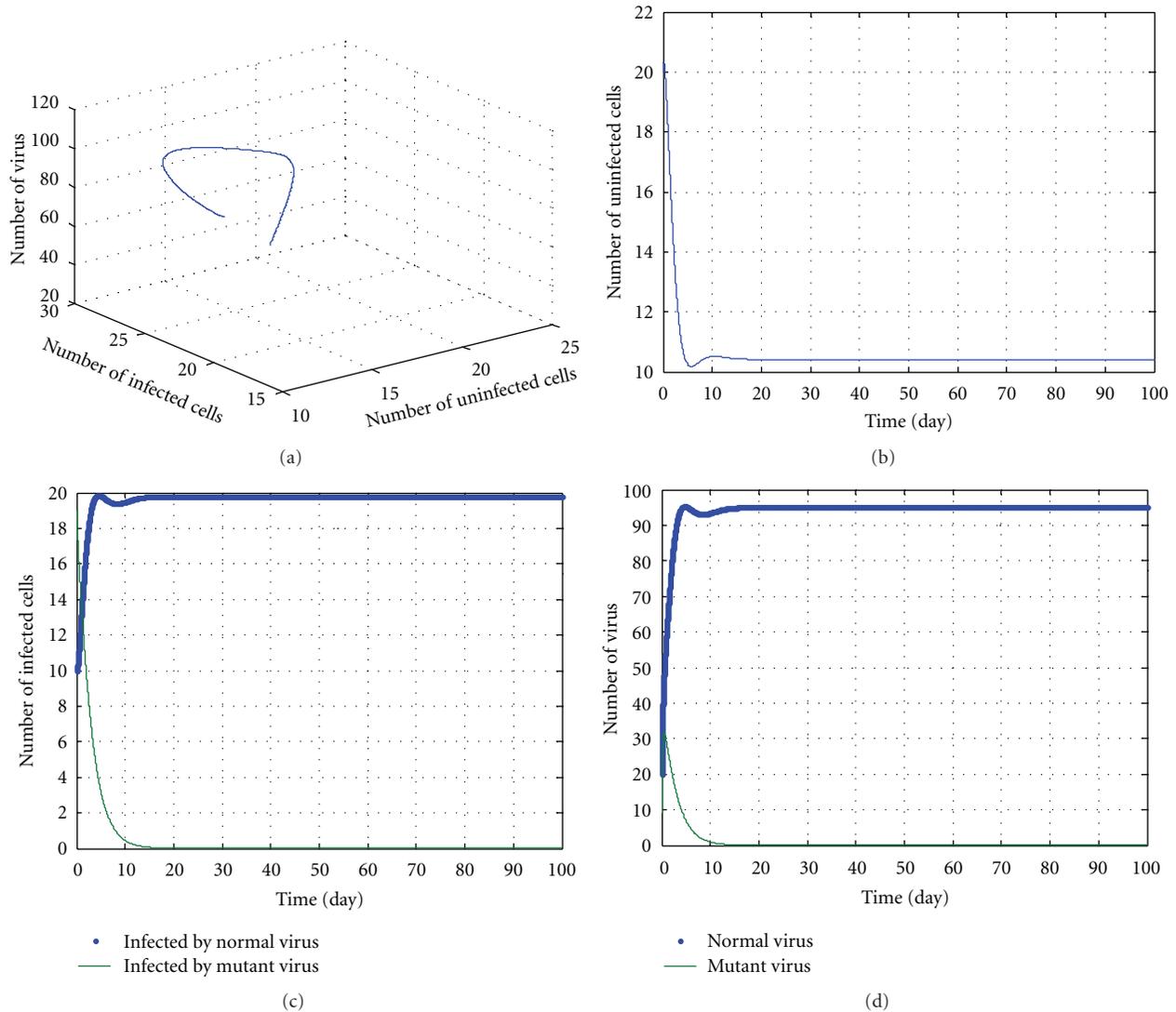


FIGURE 7: The kinetic trajectories of equation (37) when $u = 5$, $k = 24$.

mutant virus and of the uninfected cells decrease, and the number of normal virus increases. These imply that although the drug can prevent the mutation of the HBV effectively, it may destroy uninfected cells and the liver. The potency is not ideal.

5. Conclusions

The local activity of CNN has provided a powerful tool for studying the emergence of complex patterns in a homogeneous lattice formed by coupled cells. Based on the local activity principle, the analytic criteria for the local activity in reaction-diffusion CNN with five state variables and one port are set up. The analytical criteria include four theorems, which provide the inequalities involving the parameters of the CNN. The inequalities can be used for calculating the bifurcation diagram to determine emergence of complex

dynamic patterns of the reaction-diffusion CNN. As an application example, a reaction-diffusion CNN of HBV mutation-selection model is analyzed and simulated, and the bifurcation diagrams are calculated. Numerical simulations show this CNN model may explain certain complex mutant conditions during the therapy. We conclude that the local activity theory provides a practical tool for the study of the complex dynamics of certain coupled nonlinear systems.

Acknowledgment

This work is supported in part by NSFC 70890084, 60921061, 90920305.

References

- [1] L. O. Chua, "CNN: a version of complexity," *International Journal of Bifurcation and Chaos*, vol. 7, no. 10, pp. 2219–2425, 1997.

- [2] L. O. Chua, "Local activity is the origin of complexity," *International Journal of Bifurcation and Chaos*, vol. 15, no. 11, pp. 3435–3456, 2005.
- [3] L. O. Chua and L. Yang, "Cellular neural networks: theory and applications," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 10, pp. 1273–1290, 1988.
- [4] L. O. Chua, "Passivity and complexity," *IEEE Transactions on Circuits and Systems*, vol. 46, no. 1, pp. 71–82, 1999.
- [5] S. Y. Chen, H. Y. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
- [6] S. Y. Chen, W. Huang, and C. Cattani, "Traffic dynamics on complex networks: a survey," *Mathematical Problems in Engineering*, vol. 2012, Article ID 732698, 23 pages, 2012.
- [7] S. Y. Chen, M. Zhao, G. Wu, C. Yao, and J. Zhang, "Recent advances in morphological cell image analysis," *Computational and Mathematical Methods in Medicine*. In press.
- [8] S. Y. Chen, Y. Zheng, C. Cattani, and W. Wang, "Modeling of biological intelligence for SCM system optimization," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 769702, 10 pages, 2012.
- [9] R. Dogaru and L. O. Chua, "Edge of chaos and local activity domain of Fitzhugh-Nagumo," *International Journal of Bifurcation and Chaos*, vol. 8, no. 2, pp. 671–693, 1998.
- [10] R. Dogaru and L. O. Chua, "Edge of chaos and local activity domain of the Brusselator CNN," *International Journal of Bifurcation and Chaos*, vol. 8, no. 6, pp. 1107–1130, 1998.
- [11] R. Dogaru and L. O. Chua, "Edge of chaos and local activity domain of the Gierer-Meinhardt CNN," *International Journal of Bifurcation and Chaos*, vol. 8, no. 12, pp. 2321–2340, 1998.
- [12] L. Q. Min, K. R. Crouse, and L. O. Chua, "Analytical criteria for local activity and applications to the Oregonator CNN," *International Journal of Bifurcation and Chaos*, vol. 10, no. 1, pp. 25–71, 2000.
- [13] L. Q. Min, K. R. Crouse, and L. O. Chua, "Analytical criteria for local activity of reaction-diffusion CNN with four state variables and applications to the Hodgkin-Huxley equation," *International Journal of Bifurcation and Chaos*, vol. 10, no. 6, pp. 1295–1343, 2000.
- [14] L. Q. Min and G. Chen, "Local activity of the Van der Pol CNN," *International Journal of Bifurcation and Chaos*, vol. 14, no. 7, pp. 2211–2222, 2004.
- [15] L. Q. Min and N. Yu, "Analytical criteria for local activity of CNN with two-port and application to biochemical model," *Journal of University of Science and Technology Beijing*, vol. 7, no. 4, pp. 305–314, 2000.
- [16] L. Q. Min and N. Yu, "Some analytical criteria for local activity of two-port CNN with three or four state variables: analysis and applications," *International Journal of Bifurcation and Chaos*, vol. 12, no. 5, pp. 931–963, 2002.
- [17] L. Q. Min, J. Wang, X. Dong, and G. Chen, "Some analytical criteria for local activity of three-port CNN with four state variables: analysis and applications," *International Journal of Bifurcation and Chaos*, vol. 13, no. 8, pp. 2189–2239, 2003.
- [18] L. Q. Min and N. Yu, "Application of local activity theory of cellular neural network with two ports to the coupled Lorenz-cell model," *Communications in Theoretical Physics*, vol. 37, no. 6, pp. 759–767, 2002.
- [19] M. Itoh and L. O. Chua, "Advanced image processing cellular neural networks," *International Journal of Bifurcation and Chaos*, vol. 17, no. 4, pp. 1109–1150, 2007.
- [20] Y. Meng, L. Q. Min, and Y. Ji, "Application of local activity theory to rossler CNN model," in *Proceedings of the 9th International Conference for Young Computer Scientists (ICYCS '08)*, pp. 2846–2851, November 2008.
- [21] H. Li, X. Liao, C. Li, H. Huang, and C. Li, "Edge detection of noisy images based on cellular neural networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 9, pp. 3746–3759, 2011.
- [22] W. Borges Sampaio, E. Moraes Diniz, A. Corrêa Silva, A. Cardoso de Paiva, and M. Gattass, "Detection of masses in mammogram images using CNN, geostatistic functions and SVM," *Computers in Biology and Medicine*, vol. 41, no. 8, pp. 653–664, 2011.
- [23] S. U. Alver, B. Cuma, and U. N. Osman, "Application of cellular neural network (CNN) to the prediction of missing air pollutant data," *Atmospheric Research*, vol. 101, no. 1-2, pp. 314–326, 2011.
- [24] A. Pirouzmmand and K. Hadad, "Cellular neural network to the spherical harmonics approximation of neutron transport equation in x-y geometry—part I: modeling and verification for time-independent solution," *Annals of Nuclear Energy*, vol. 38, no. 6, pp. 1288–1299, 2011.
- [25] A. Grabowski, R. A. Kosiński, and M. Dźwiarek, "Vision safety system based on cellular neural networks," *Machine Vision and Applications*, vol. 22, no. 3, pp. 581–590, 2011.
- [26] J. J. Martínez, J. Garrigós, J. Toledo, E. Fernández, and J. Manuel Ferrández, "Implementation of a CNN-based retinomorphic model on a high performance reconfigurable computer," *Neurocomputing*, vol. 74, no. 8, pp. 1290–1297, 2011.
- [27] M. Di Marco, M. Grazzini, and L. Pancioni, "Global robust stability criteria for interval delayed full-range cellular neural networks," *IEEE Transactions on Neural Networks*, vol. 22, no. 4, pp. 666–671, 2011.
- [28] S. Long and D. Xu, "Stability analysis of stochastic fuzzy cellular neural networks with time-varying delays," *Neurocomputing*, vol. 74, no. 14-15, pp. 2385–2391, 2011.
- [29] L. Wan and Q. Zhou, "Attractor and ultimate boundedness for stochastic cellular neural networks with delays," *Nonlinear Analysis. Real World Applications*, vol. 12, no. 5, pp. 2561–2566, 2011.
- [30] M. Syed Ali and P. Balasubramaniam, "Global asymptotic stability of stochastic fuzzy cellular neural networks with multiple discrete and distributed time-varying delays," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 7, pp. 2907–2916, 2011.
- [31] M. A. Nowak et al., *Virus Dynamics—Mathematical Principles of Immunology and Virology*, chapter 11, Oxford university press, 2000.

Research Article

Wavelet Transform Fuzzy Algorithms for Dermoscopic Image Segmentation

Heydy Castillejos,¹ Volodymyr Ponomaryov,¹ Luis Nino-de-Rivera,¹ and Victor Golikov²

¹National Polytechnic Institute 04430, Mexico city, DF, Mexico

²UNACAR, 24180 Ciudad del Carmen, CaM, Mexico

Correspondence should be addressed to Heydy Castillejos, heydycastillejos@gmail.com

Received 1 September 2011; Revised 17 January 2012; Accepted 20 January 2012

Academic Editor: Carlo Cattani

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

This paper presents a novel approach to segmentation of dermoscopic images based on wavelet transform where the approximation coefficients have been shown to be efficient in segmentation. The three novel frameworks proposed in this paper, W-FCM, W-CPSFCM, and WK-Means, have been employed in segmentation using ROC curve analysis to demonstrate sufficiently good results. The novel W-CPSFCM algorithm permits the detection of a number of clusters in automatic mode without the intervention of a specialist.

1. Introduction

According to the World Health Organization, skin cancer is the most common form of human cancer. It is estimated that over one million new cases occur annually. Additionally, the diagnosis of skin cancer is particularly important because melanoma can be cured with a simple excision if detected early.

The term “skin cancer” refers to three different conditions that are listed below in ascending order of mortality:

- (i) basal cell carcinoma (or basal cell *carcinoma epithelioma*),
- (ii) squamous cell carcinoma (the first stage of which is called *actinic keratosis*),
- (iii) *melanoma*.

Melanoma is generally the most serious form of skin cancer because it tends to spread (metastasize) throughout the body quickly.

To diagnosis skin cancer, doctors usually remove all or a part of the growth by performing a biopsy, but this is considered an invasive technique. Alternatively, a technique called dermatoscopy reduces the need for a biopsy by utilizing

a dermatoscope. Dermatoscopy is a particularly helpful standard method of diagnosing the malignancy of skin lesions [1]. One of the major advantages of dermatoscopy is an increase in accuracy compared with naked-eye examination (up to 20% in the case of sensitivity and up to 10% in the case of specificity), thereby reducing the frequency of unnecessary surgical excisions of benign lesions [2–4].

In addition, several instruments designed for a computer-aided diagnosis (CAD) of skin lesions have been proposed. These usually work in four steps: data acquisition of skin (dermoscopic images), segmentation, feature extraction, and classification. The most relevant step is the segmentation process because it provides fundamental information to the next stages. Image segmentation is the process of adequately grouping pixels into a few regions, where pixels within a group share some similar characteristics. Automated analysis of the edges, colors, and shape of the lesion relies upon an accurate segmentation, and this is an important first step in any CAD system. However, irregular shape, non-uniform color, and ambiguous structures make the problem difficult.

Image segmentation can be classified into three categories (a) *Supervised*: these methods require the intervention of the analyst, who should specify the sections of skin or

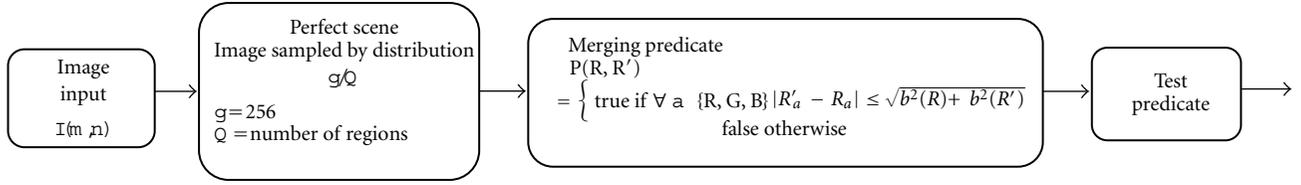


FIGURE 1: Block diagram of statistical region merging.

lesions within the image [4, 5]. (b) *Automatic*: also known as unsupervised methods, these systems attempt to find the lesion borders without any intervention from an analyst. (c) *Semiautomatic*: this describes a combination of manual and automatic segmentation.

Many segmentation methods have been developed for border detection in pigmented skin images, and they traditionally focus on dermoscopic images.

In [6], automatic adaptive thresholding (AT) has been proposed where the main idea is to segment an image comparing the color of each pixel with a threshold. In [7], the authors use a variant of a region growing and merging technique called statistical region merging (SRM) to segment an image. The SRM technique has been proven to be robust for the segmentation of color images, improving the detection rate of skin lesions.

A recent review of methods for segmentation of skin lesions in dermoscopic images [5] suggests that clustering is the most popular segmentation technique, most likely due to its robustness.

However, in some cases, it can be difficult to perform good segmentation because of hair occlusions within the pigmented skin lesion. In this case, preprocessing methods, such as those proposed in the following papers [8, 9], should be employed before segmentation.

Texture is an important characteristic of image analysis for both natural scenes and medical images. The *wavelet transform* (WT) provides an ideal representation of texture analysis presenting spatial-frequency properties via a pyramid of tree structures that is similar to subband decomposition. The hierarchical decomposition enables an analysis of the high frequencies in the image, which is important for the segmentation process.

Several techniques use the image features within a WT domain during the segmentation process. In Bello [10], image data are first decomposed into channels for a selected set of resolution levels using wavelet packets transform. Then, Markov random field (MRF) segmentation is applied to the subband coefficients for each scale, starting with the coarsest level and propagating the segmentation process from the current level to segmentation at the next level. Strickland and Hahn [11] use the image features extracted in the WT domain for detection of microcalcifications in mammograms using a matching process and some *a priori* knowledge on the target objects. Zhang and Desai [12] employ a Bayes classifier on wavelet coefficients to determine an appropriate scale and threshold that can separate segmentation targets from other features.

In this paper, we propose an efficient approach for the segmentation of dermoscopic images based on a clustering process. Our novel approach uses feature extraction in wavelet transform space before proceeding to the segmentation process. The main difference with other algorithms presented in the literature is in the usage of information from three color channels (RGB space) in wavelet transform space gathering the color channels via a *nearest neighbor interpolation* (NNI).

The paper is organized as follows. Section 2 presents a methodology; Section 3 exposes the proposed framework; Section 4 discusses the evaluation criteria applied; Section 5 contains the simulation results; Section 6 contains the contribution of this work; Section 6 concludes the paper.

2. Methodology

2.1. Statistical Region Merging. In [7], the authors use a variant of a region growing and merging technique called statistical region merging (SRM) to segment an image. The SRM technique has been proven to be robust for the segmentation of color images, improving the detection rate of skin lesions. This framework includes the following strategy.

First, a registered image I in RGB space is considered an observation of a true image I^* , in which pixels are perfectly represented by a family of distributions from each of the observed color channels. The color channel values for every pixel are replaced by Q independent random variables with values from $(0, g/Q)$, where the value Q represents the number of regions that should be generated.

The predicate of regions is defined as

$$P(R, R') = \begin{cases} \text{true} & \text{if } \forall a \in \{R, G, B\} |R'_a - R_a| \\ & \leq \sqrt{b^2(R) + b^2(R')} \\ \text{false} & \text{otherwise,} \end{cases} \quad (1)$$

where R and R' represent the two regions being tested and R_a denotes the observed average. R and $R_{|p|}$ are the set of regions with p pixels. The SRM framework is shown in Figure 1.

2.2. K-Means Clustering Algorithm. This algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other [13]. It works in an iterative manner according to the following steps.

- (1) Choose initial centroids m_1, \dots, m_k of the clusters C_1, \dots, C_k .
- (2) Calculate new cluster membership. A feature vector x_j is assigned to the cluster C_i if and only if

$$i = \operatorname{argmin}_{k=1, \dots, K} \|x_j - m_k\|^2. \quad (2)$$

- (3) Recalculate the centroids for the clusters according to

$$m_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j, \quad (3)$$

where x_j belong to dataset $X = \{x_1, \dots, x_i, \dots, x_N\}$.

- (4) If none of the cluster centroids were changed, finish the algorithm. Otherwise go to step 2.

2.3. Fuzzy C-Means Algorithm. The algorithm finds the center of “ n ” number of clusters iteratively by adjusting their position and evaluating an objective function. Additionally, it permits more flexibility by introducing partial membership to other clusters. The classical algorithm uses the following objective function:

$$E = \sum_{j=1}^N \sum_{i=1}^C \mu_{ij}^k \|x_i - c_j\|^2, \quad (4)$$

where μ_{ij}^k is the fuzzy membership of the pixel x_i , the cluster identified by its center c_j , and k is a constant that defines the fuzziness of the resulting partitions.

The membership value is proportional to the probability that a pixel belongs to some specific cluster where the probability is only dependent on the distance between the pixel and each independent cluster center. Consequently, the criterion E has a minimal value when the pixels are nearby the corresponding cluster center. Higher membership values are assigned to these nearby pixels while lower membership values are assigned to the pixels that are far from a center. This algorithm runs with the clusters number and initial center positions as previously determined. The algorithm then determines how many pixels belong to each cluster. The membership function and centers are determined as follows:

$$\mu_{ij} = \frac{1}{\sum_{m=1}^C \|x_i - c_j\| / \|x_i - c_m\|^{2/(k-1)}}, \quad (5)$$

$$c_i = \frac{\sum_{j=1}^N \mu_{ij}^k x_j}{\sum_{j=1}^N \mu_{ij}^k}. \quad (6)$$

The FCM algorithm runs four simple steps.

- (1) The center is initialized with the first value “ t ” of the data to be equal to zero, and this value is used as a counter for the number of iterations.
- (2) The fuzzy partition membership functions μ_{ij} are initialized according to (5).

TABLE 1: Member functions of “Distance.”

| Fuzzy set | Function | Center | Variance |
|----------------|----------|--------|----------|
| <i>Minimum</i> | Gauss | 15 | 16 |
| Shorter | Gauss | 53 | 24 |
| Short | Gauss | 105 | 30 |
| Regular | Gauss | 150 | 30 |
| Large | Gauss | 222 | 45 |
| Maximum | Gauss | 255 | 15 |

TABLE 2: Member functions of “Size.”

| Fuzzy set | Function | Center | Variance |
|------------|----------|--------------|--------------|
| <i>Min</i> | Gauss | 9000 | 1.789e + 005 |
| Small | Gauss | 3.015e + 005 | 1.626e + 005 |
| Medium | Gauss | 6.53e + 005 | 1.968e + 005 |
| Big | Gauss | 9.728e + 005 | 2.236e + 005 |
| Max | Gauss | 1.44e + 006 | 2.862e + 005 |

- (3) The value “ $t = t + 1$ ” is changed and novel centers are computed using (6).

- (4) Steps 2 and 3 run until criterion E converges.

Criterion E approaches a minimum value when its variations are decreased according to the restriction that a user selects. The algorithm can also be interrupted if a user determines that only a certain number of iterations are required [13].

2.4. Cluster Preselection Fuzzy C-Means. The FCM algorithm is one of the most common procedures for image segmentation but has the following drawback: the number of clusters needs to be predetermined by a user. Therefore, the user may not select the correct number of clusters for a given specific application. Therefore, a method that uses fuzzy logic to find the number of clusters can reproducibly select the correct number of clusters. To achieve this, we take into consideration the difference between the max (V_{\max}) and the min (V_{\min}) values of intensity in an image $D = V_{\max} - V_{\min}$. Using these proportions, the algorithm determines the optimal number of clusters. Specifically, image data are analyzed to determine the centers, thus reducing the operational time of the FCM algorithm. The first data classification for our fuzzy system is called “Distance” and has a total of six fuzzy sets, “minimum,” “shorter,” “short,” “regular,” “large,” and “maximum” (see Table 1). The classification for our fuzzy system called “Size” has a total of five fuzzy sets, “Min,” “Small,” “Medium,” “Big,” and “Max” (see Table 2). Finally, the classification for our fuzzy system called “Cluster” has five fuzzy sets, “Very few,” “Few,” “Some,” “Many,” and “Too Many” (see Table 3).

Finally, the fuzzy system “cluster” contains five fuzzy sets that are applied in the determination of the centers using 30 fuzzy rules, reducing the operational time of the FCM

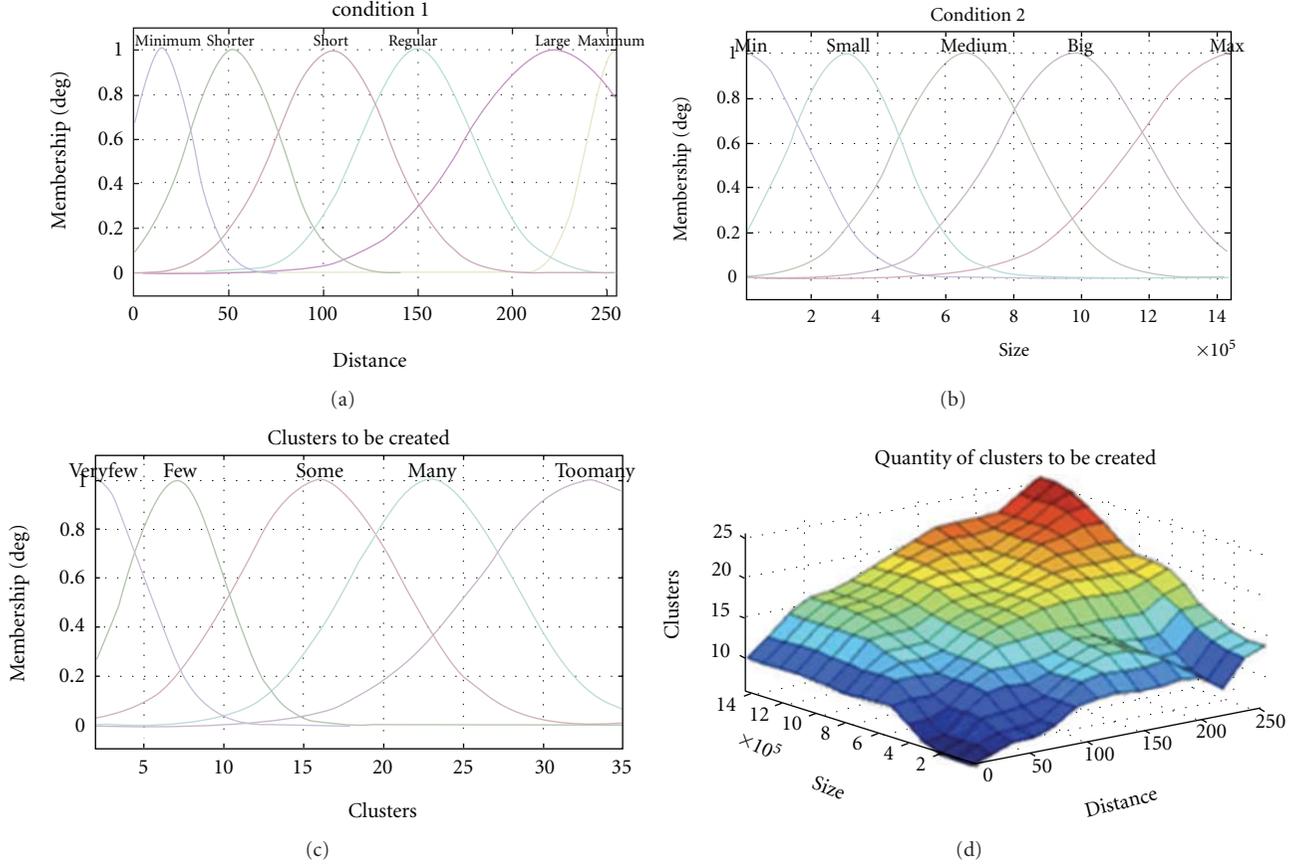


FIGURE 2: Preselection of the number of clusters.

TABLE 3: Member functions of “Clusters.”

| Fuzzy set | Function | Center | Variance |
|-----------|----------|--------|----------|
| Very Few | Gauss | 2 | 3 |
| Few | Gauss | 7 | 3 |
| Some | Gauss | 16 | 5 |
| Many | Gauss | 23 | 5 |
| Too Many | Gauss | 33 | 7 |

algorithm. The overall response of the fuzzy system can be represented as follows:

$$Q(c) = \max_i \left\{ \min \left\{ \min \left\{ \mu_{\text{distance}}(d_0), \mu_{\text{size}}(s_0), \mu_i(d_0, s_0, c) \right\} \right\} \right\}, \quad (7)$$

where $i = 1, 2, \dots, 30$ fuzzy rules, *distance* is {minimum, shorter, short, regular, large, maximum} and *size* is {min, small, medium, big, max}.

In Figure 2 we can see the membership functions with a gaussian distribution of three conditions and the total number of clusters.

In the second phase, the number of clusters and their centers are already known, simply requiring dividing the difference D into the “ N ” clusters to determine the centers:

$$c_j = j \frac{D}{N} \quad j = 1, 2, 3, \dots, N, \quad (8)$$

where “ N ” represents the number of clusters to be created and “ j ” is a counter to define all the centers.

This looks like a hard type of algorithm, but the centers are still rather far from the final ones. Therefore, there are still a certain number of iterations that should be applied to find them, but the number of iterations is far fewer than for the original system, reducing the required computation time.

The RGB image is decomposed into its three-color channels, and the Euclidean distance is employed [13] to determine the difference between the three distances for each color:

$$\begin{aligned} d_1(x_{\text{red}}, x_{\text{blue}}) &= \sqrt{\sum_{k=1}^P (x_{\text{red}}^k - x_{\text{blue}}^k)^2}, \\ d_2(x_{\text{red}}, x_{\text{green}}) &= \sqrt{\sum_{k=1}^P (x_{\text{red}}^k - x_{\text{green}}^k)^2}, \\ d_3(x_{\text{green}}, x_{\text{blue}}) &= \sqrt{\sum_{k=1}^P (x_{\text{green}}^k - x_{\text{blue}}^k)^2}. \end{aligned} \quad (9)$$

Two distances that are more alike are combined into one grayscale image and then processed as a corrected image. The proposed method is then used to determine the number of clusters to be created.

The CPSFCM consists of the following steps.

- (1) Divide the RGB image into three different images, use (9) to find two images that are more similar to each other, and use them to create a new grayscale image.
- (2) Calculate the distance between intensity levels in the image D , and obtain the size of an image.
- (3) Use this data with the fuzzy preselective system and obtain the number of centers to be created.
- (4) Use (8) to obtain the approximate centers. The initial value “ t ” is equal to zero, and it is used as a counter for the number of the iterations.
- (5) The fuzzy partition membership functions μ_{ij} are initialized according to (5).
- (6) Let the value be “ $t = t + 1$ ” and compute the new centers using (6).
- (7) The steps 5 and 6 should be performed until criterion E converges.

3. Proposed Framework

In the proposed approach, the procedure consists of the following stages: a digital color image $I[n, m]$ is separated into R , G , and B channels in color space, where each channel image is decomposed calculating their wavelet coefficients using Mallat’s pyramid algorithm [14]. Using the chosen wavelet family, the original image is decomposed into four subbands. These subbands, labeled as LH, HL, and HH, represent the finest scale wavelet coefficient (detail images), while the subband LL corresponds to coarse level coefficients (approximation image), noted below as D_h^j , D_v^j , D_d^j , and A^{2^j} , respectively at given scale 2^j , for $j = 1, 2, \dots, J$, where J is the number of scales used in the discrete wavelet transform (DWT). Finally, the DWT can be represented as follows:

$$W_i = |W_i| \exp(j \ i),$$

$$|W_i| = \left(\sqrt{|D_{h,i}|^2 + |D_{v,i}|^2 + |D_{d,i}|^2} \right),$$

$$i = \begin{cases} \alpha_i & \text{if } D_{h,i} > 0 \\ \pi - \alpha_i & \text{if } D_{h,i} < 0 \end{cases} \quad \alpha_i = \tan^{-1} \left(\frac{D_{v,i}}{D_{h,i}} \right).$$
(10)

Therefore, W_i is considered a new image for each color channel. The following process, conducted in wavelet transform space, consists of several stages: the classic segmentation method is applied to each channel image; the segmented image corresponding to the red channel is interpolated with the segmented image corresponding to the green channel, and after applying the NNI process, the resulting image is interpolated with the segmented image corresponding to the blue channel using NNI again. Finally, this image is considered the output of the segmentation procedure.

The block diagram in Figure 3 explains in detail the operations for the following: (a) image segmentation using the K-Means algorithm where WT is applied, named WK-Means, (b) image segmentation using the FCM algorithm where WT is applied, named W-FCM, and, finally, (c) image segmentation using the CPSFCM algorithm where WT is applied, named W-CPSFCM.

4. Evaluation Criteria

Different objective measures are used in the literature for the purpose of evaluation of the performance of border detection in dermoscopic images.

Objective measures require a ground truth (GT) image, which is determined by a dermatologist manually drawing the border around the lesion. Using a GT image, Garnavi et al. [15] calculated the operation exclusive disjunction (XOR) measure. Other metrics used in segmentation performance are presented in [16, 17] and include the *sensitivity* and *specificity*, precision and recall, *true positive rate*, *false positive rate*, *pixel misclassification probability*, and the *weighted performance index*.

Below, let us consider the sensitivity and specificity measure. Sensitivity and specificity are statistical measures of the performance of a binary classification test, commonly used in medical studies. In the context of segmentation of skin lesions, sensitivity measures the proportion of actual lesion pixels that are correctly identified as such. Specificity measures the proportion of background skin pixels that are correctly identified. We give the following definitions.

TP: true positive, lesion pixels correctly classified as lesion.

FP: false positive, skin pixels incorrectly identified as lesion.

TN: true negative, skin pixels correctly identified as skin.

FN: false negative, lesion pixels incorrectly identified as skin,

where, in each of the above categories, the sensitivity and specificity are given by

$$\text{sensitivity} = \frac{TP}{TP + FN},$$

$$\text{specificity} = \frac{TN}{FP + TN}.$$
(11)

We also apply the receiver operating characteristic (ROC) analysis that permits us to evaluate the image segmentation quality in terms of the ability of human observers or a computer algorithm using image data to classify patients as “positive” or “negative” with respect to any particular disease. This characteristic represents the second level of diagnostic efficacy in the hierarchical model described by Fryback and Thornbury [17]. The points of the ROC curve are obtained by sweeping the classification threshold from the most positive classification value to the most negative and can be used to produce quantitative summary measures of the ROC curve for this measure called the *area under the ROC curve* (AUC).

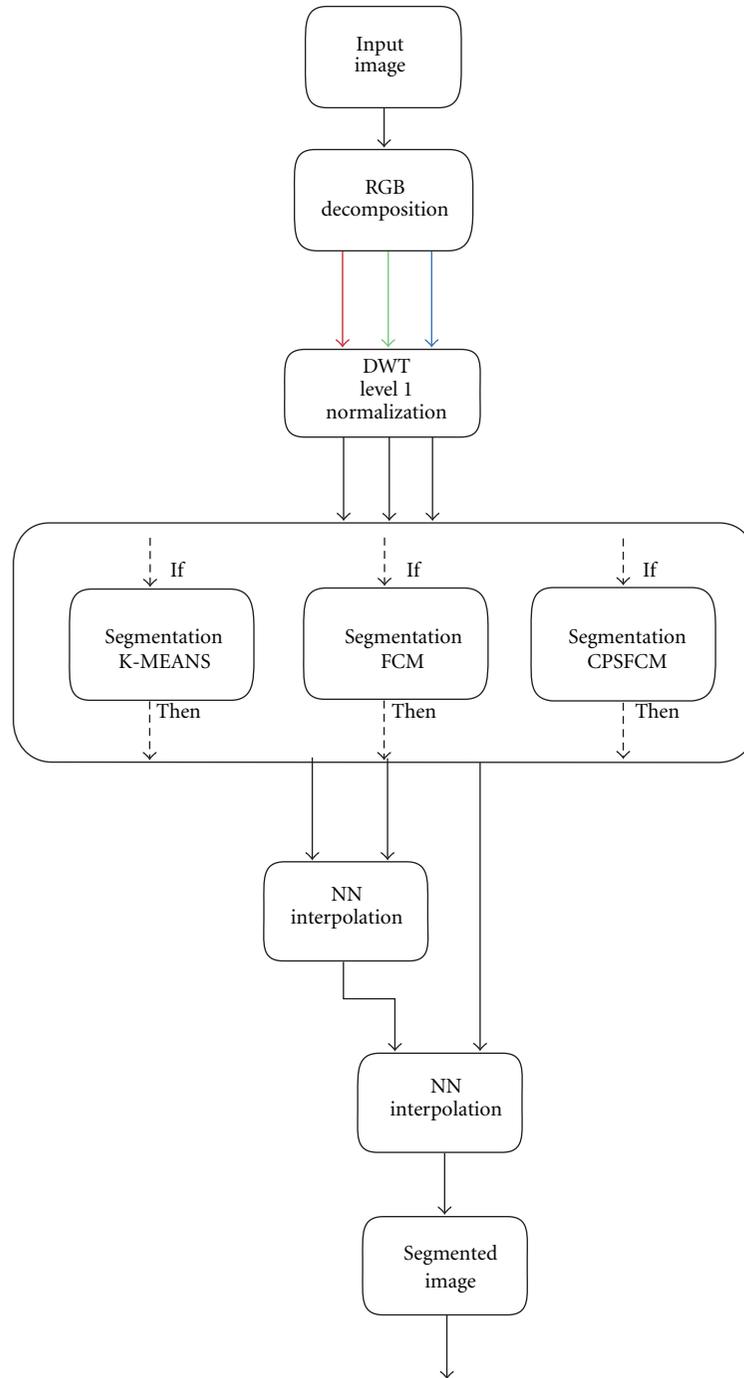


FIGURE 3: Block diagram of the proposed algorithms: WK-MEANS, W-FCM, and W-CPSFCM.

5. Simulation Results

The proposed segmentation algorithms were evaluated on a set of 50 images of dermoscopic images obtained from <http://www.dermoscopyatlas.com> [18]. These images do not contain occlusions because the preprocessing procedure has already been applied. The *GT* images were obtained via human-based segmentation (see Figure 5). The dataset

presents 24-bit color images in JPEG format with 600×600 pixel size. Below, we expose only five different images with different texture characteristics where the *sensitivity* and *specificity* are used as the evaluation criteria for segmentation accuracy. We also plotted the ROC curves to examine the classifier performance. Additionally, the diagnostic performance was quantified by the AUC measure. Figure 4 shows the dermoscopic images used in this study.

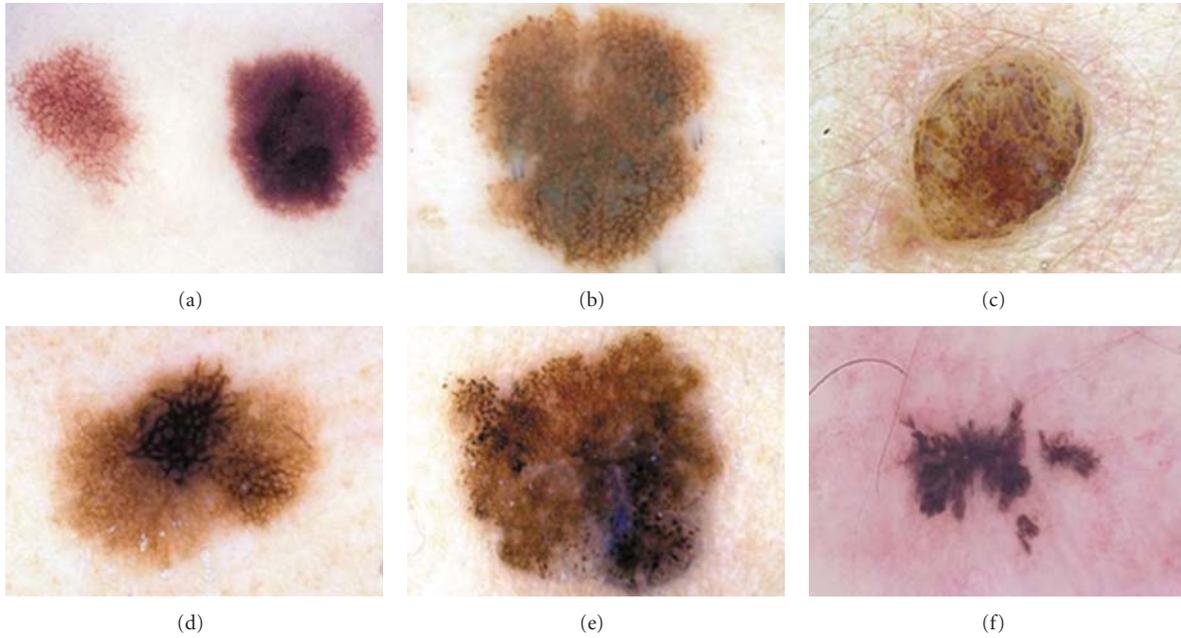


FIGURE 4: Dermoscopic images used in this study: (a) *Clark's nevus* (lesion 1), (b) *Clark nevus's* (lesion 2), (c) *dermal nevus* (lesion 3), (d) *melanoma* (lesion 4), (e) *melanoma* (lesion 5), (f) *recurrent nevus* (lesion 6).

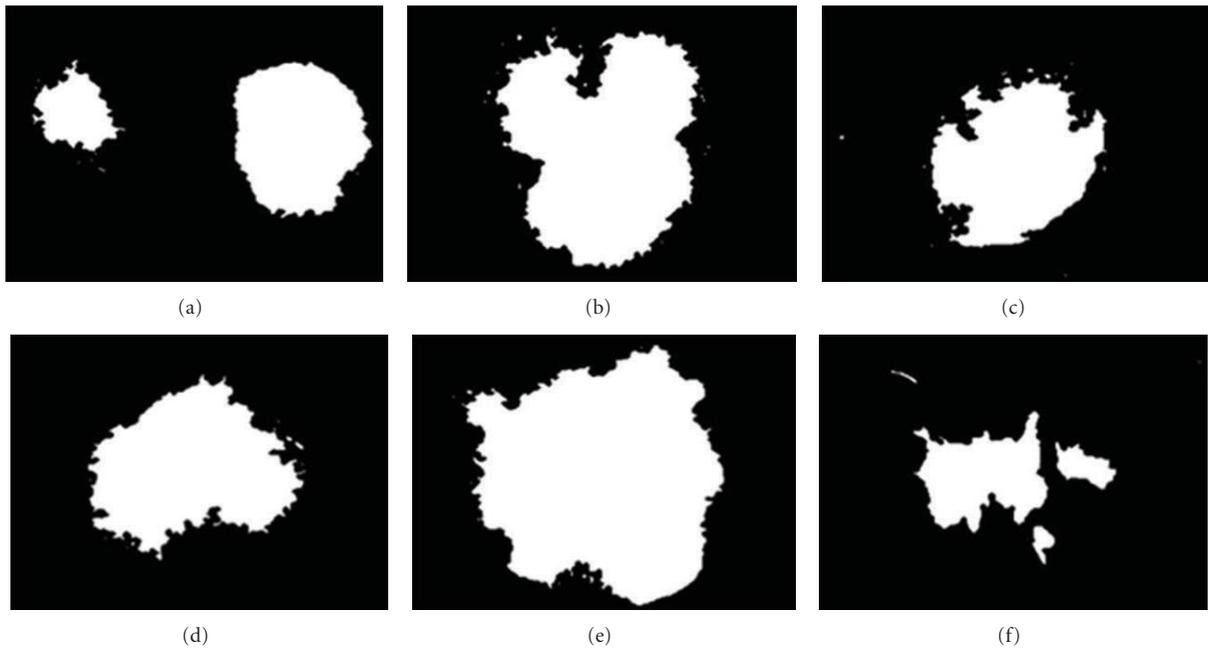


FIGURE 5: GT of dermoscopic images used in this study: (a) *Clark nevus's* (lesion 1), (b) *Clark's nevus* (lesion 2), (c) *dermal nevus* (lesion 3), (d) *melanoma* (lesion 4), (e) *melanoma* (lesion 5), (f) *recurrent nevus* (lesion 6).

The simulation results in Table 4 present the AUC values for the proposed framework based on different wavelet families and confirm their improved performance compared to classical techniques. The maximum AUC value is obtained when WF Daubechies 4 is used followed by the WAF π_6 .

According to [16], AUC values should be greater than 0.8 to be considered a good test, but our study is focused on the

best approximation of a segmented image to GT, meaning that the goal is to achieve an AUC value of approximately one.

Figures 6(a) and 6(b) present the skin lesions and their corresponding GT. In Figures 6(c) and 6(f), it is easy to note that the segmentation procedure has selected the area only around the lesion. On the other hand, in Figures 6(g)

TABLE 4: AUC simulation results using different segmentation algorithms.

| | Lesion 1 | Lesion 2 | Lesion 3 | Lesion 4 | Lesion 5 | | Lesion 1 | Lesion 2 | Lesion 3 | Lesion 4 | Lesion 5 |
|----------|--------------|--------------|--------------|----------|--------------|---------------------|----------|----------|----------|--------------|----------|
| | | | | | | Without wavelet | | | | | |
| CPSFCM | 0.954 | 0.915 | 0.530 | 0.914 | 0.946 | W-CSPFCM | 0.798 | 0.787 | 0.886 | 0.906 | 0.921 |
| FCM | 0.967 | 0.936 | 0.955 | 0.954 | 0.960 | W-FCM | 0.826 | 0.929 | 0.901 | 0.935 | 0.913 |
| K-Means | 0.969 | 0.935 | 0.955 | 0.952 | 0.959 | WK-Means | 0.858 | 0.957 | 0.922 | 0.950 | 0.925 |
| SRM | 0.856 | 0.930 | 0.877 | 0.929 | 0.801 | | | | | | |
| | | | | | | WF Coiflets 3 | | | | | |
| W-CSPFCM | 0.851 | 0.841 | 0.923 | 0.948 | 0.932 | W-CSPFCM | 0.832 | 0.956 | 0.887 | 0.929 | 0.943 |
| W-FCM | 0.966 | 0.948 | 0.956 | 0.961 | 0.963 | W-FCM | 0.874 | 0.953 | 0.926 | 0.953 | 0.931 |
| WK-Means | 0.871 | 0.959 | 0.928 | 0.953 | 0.928 | WK-Means | 0.898 | 0.961 | 0.941 | 0.965 | 0.934 |
| | | | | | | WF Daubechies 4 | | | | | |
| W-CSPFCM | 0.886 | 0.956 | 0.961 | 0.958 | 0.961 | W-CSPFCM | 0.811 | 0.758 | 0.868 | 0.914 | 0.936 |
| W-FCM | 0.969 | 0.945 | 0.957 | 0.959 | 0.970 | W-FCM | 0.846 | 0.940 | 0.911 | 0.943 | 0.920 |
| WK-Means | 0.874 | 0.964 | 0.937 | 0.960 | 0.939 | WK-Means | 0.878 | 0.960 | 0.931 | 0.957 | 0.931 |
| | | | | | | WF Biorthogonal 6.8 | | | | | |
| W-CSPFCM | 0.878 | 0.939 | 0.913 | 0.955 | 0.947 | W-CSPFCM | 0.811 | 0.763 | 0.870 | 0.911 | 0.935 |
| W-FCM | 0.966 | 0.949 | 0.956 | 0.962 | 0.964 | W-FCM | 0.844 | 0.939 | 0.910 | 0.942 | 0.919 |
| WK-Means | 0.869 | 0.958 | 0.927 | 0.953 | 0.928 | WK-Means | 0.875 | 0.960 | 0.929 | 0.959 | 0.932 |

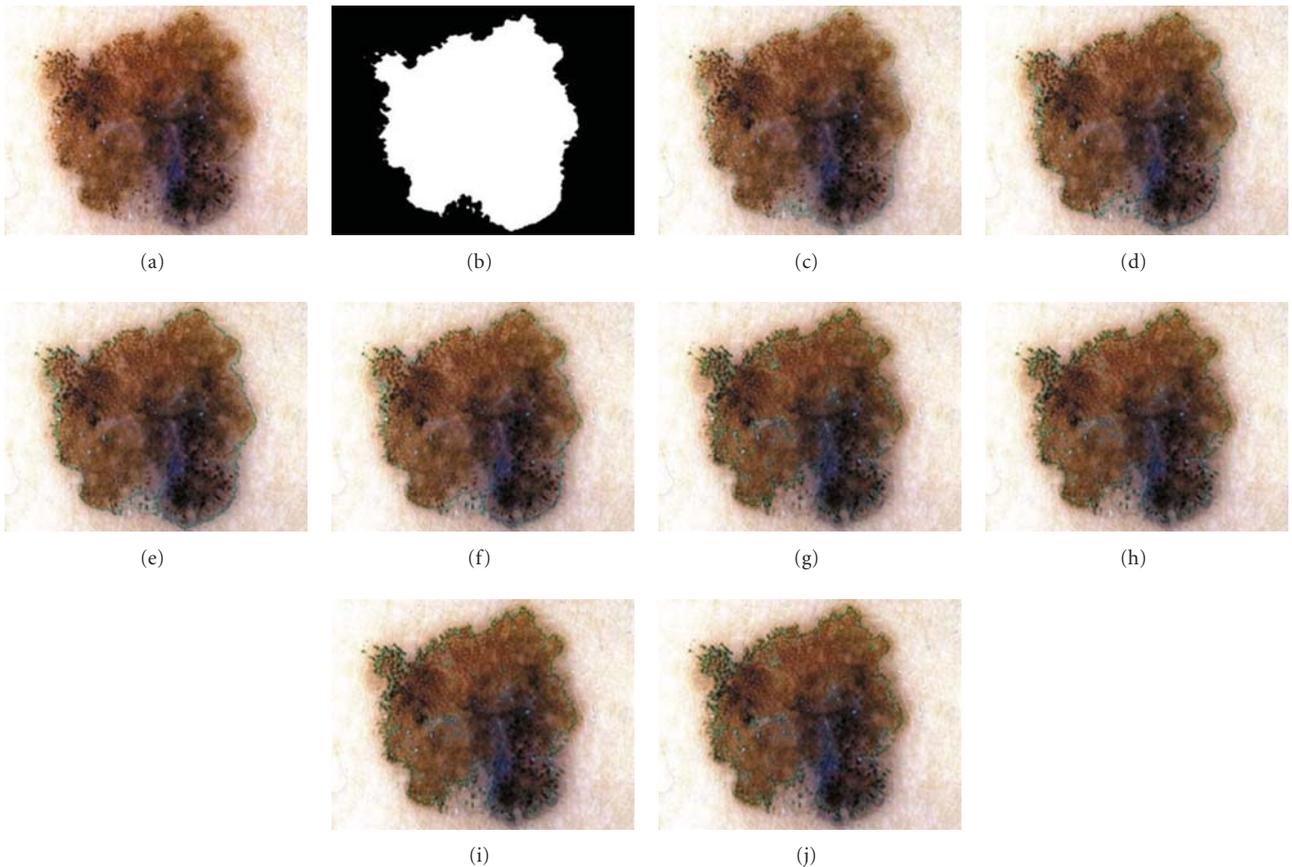


FIGURE 6: Image segmentation results under different algorithms using (a) *melanoma*, (b) *ground truth*, (c) FCM, (d) W-FCM with WF Coiflets 3, (e) W-FCM with Daubechies 4, (f) W-FCM with WF biorthogonal 6.8, (g) W-FCM with WAF up_2 , (h) W-FCM with WAF π_6 , and (i) W-FCM with WAF fup_2 , (j) W-FCM with WAF e_2 .

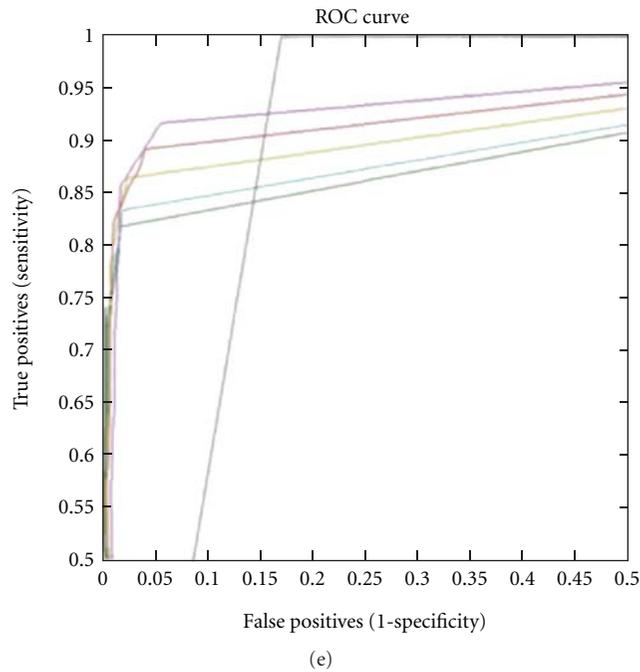
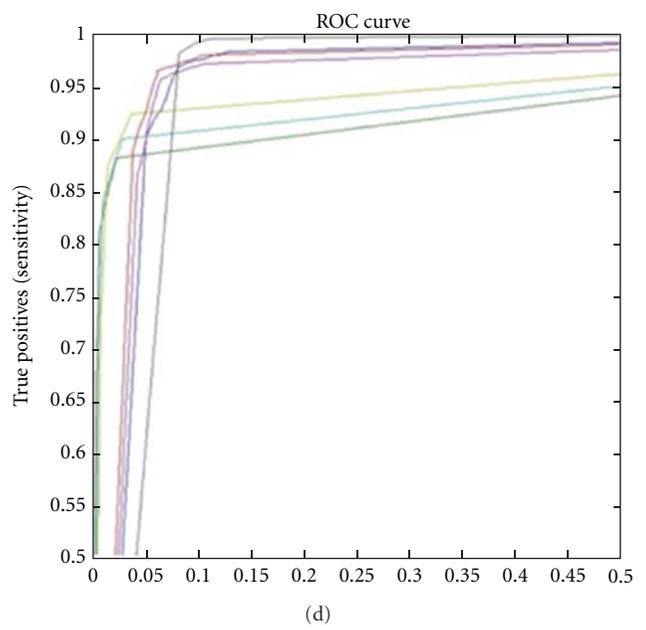
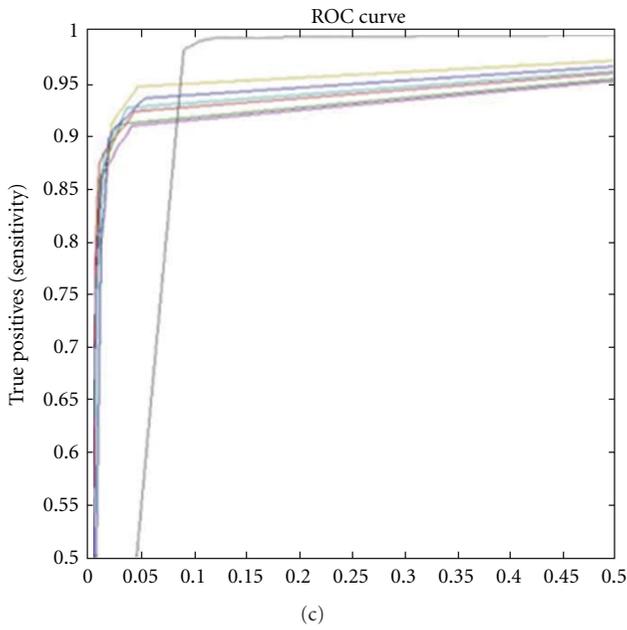
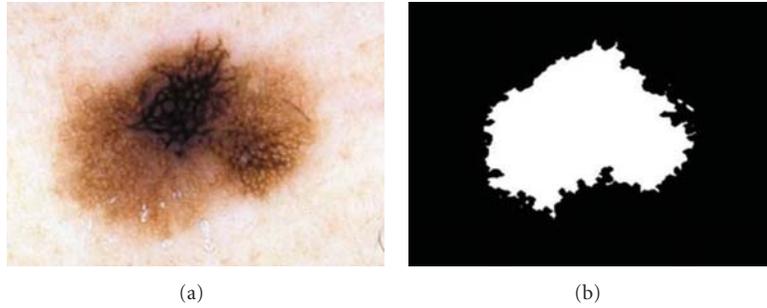


FIGURE 7: (a) Lesion 4 *melanoma*, (b) *ground truth* image; ROC curves for (c) WK-Means algorithm, (d) FCM algorithm, and (e) W-CPSFCM: for WF Daubechies 4 (dark blue), for WF biorthogonal 6.8 (red), for WF Coiflets 3 (purple), for WAF up_2 (dark green), for WAF fup_2 (aqua), and for WAF π_6 (light green); FCM (black).

and 6(j), where WAF results are presented, one can see that together with segmentation of the lesion border, there are some skin areas included in the lesion segment.

Figure 7 presents ROC curves for lesion 4 comparing the classic and proposed algorithms. In particular, Figure 7(c) presents the ROC curves for the WK-means and K-Means algorithms where one can see the superiority of the proposed WK-Means algorithm that uses WAF π_6 (see ROC curve in light green color). Figure 7(d) presents the ROC curves for the W-FCM and FCM algorithms where it is easy to observe the better performance of WK-Means that employs the WF biorthogonal 6.8 (see ROC curve in red color). Finally, in Figure 7(e), the ROC curves for the W-CPSFCM and CPS-FCM algorithms have confirmed the better performance of the first one for WF biorthogonal 6.8 usage (see ROC curve in red color). The method marks one boundary around the principal lesion and sometimes other discontinues regions that can be the regions of speared lesions. All marked clusters are important for evaluation and classification processes.

6. Contributions of This Work

Many authors have studied the segmentation problem in dermoscopic images. The principal contribution of the current proposal resides in the use of information from all color channels together during image segmentation. We first propose an approach that involves the wavelet transform space via decomposition process in the segmentation process, employing different wavelet families. Then, the interpolation procedure between every two channels is used, finally gathering detail information of three color channels of the output segmented image. Another achievement of the proposed framework, in our opinion, is the designed pre-selective clusters system, which determines the number of clusters automatically, to analyze with the color channel images. This preselective system optimizes the FCM framework. A disadvantage of the proposed preselective system consists of an additional program intervention that may be needed for clusters with zero pixel values during channel interpolation stage.

7. Conclusion

In this paper, we present three novel algorithms *W-FCM*, *W-CPSFCM*, and *WK-Means* that are applied in segmentation of dermoscopic images. All of these frameworks are compared with analogue ones that do not apply wavelet transform. The segmentation objective measures have been performed using sensibility, specificity, and AUC metric. The ROC curve analysis is also applied confirming that the usage of wavelet transform features is very promising in segmentation of dermoscopic images producing sufficiently good results. The proposed W-CPSFCM algorithm employs an additional procedure permitting to find a number of clusters in automatic mode without the intervention of a specialist. In the future, we suppose to develop the lesion classification framework using the current segmentation method.

Acknowledgments

The authors thank the National Polytechnic Institute, Mexico, and CONACYT (Project 81599) for their support to realize this work.

References

- [1] G. Argenziano and H. P. Soyer, "Dermoscopy of pigmented skin lesions—a valuable tool for early diagnosis of melanoma," *Lancet Oncology*, vol. 2, no. 7, pp. 443–449, 2001.
- [2] M. E. Vestergaard, P. Macaskill, P. E. Holt, and S. W. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting," *British Journal of Dermatology*, vol. 159, no. 3, pp. 669–676, 2008.
- [3] G. Argenziano, G. Fabbrocini, P. Carli, V. de Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Archives of Dermatology*, vol. 134, no. 12, pp. 1563–1570, 1998.
- [4] P. A. Ascierto, G. Palmieri, E. Celentano et al., "Sensitivity and specificity of epiluminescence microscopy: evaluation on a sample of 2731 excised cutaneous pigmented lesions," *British Journal of Dermatology*, vol. 142, no. 5, pp. 893–898, 2000.
- [5] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, pp. 148–153, 2009.
- [6] M. Silveira, J. C. Nascimento, J. S. Marques et al., "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 35–45, 2009.
- [7] M. E. Celebi, H. A. Kingravi, H. Iyatomi et al., "Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, vol. 14, no. 3, pp. 347–353, 2008.
- [8] Q. Abbas, M. E. Celebi, and I. F. García, "Hair removal methods: a comparative study for dermoscopy images," *Biomedical Signal Processing and Control*, vol. 6, no. 4, pp. 395–404, 2011.
- [9] Q. Abbas, M. E. Celebi, I. Fondón García, and M. Rashid, "Lesion border detection in dermoscopy images using dynamic programming," *Skin Research and Technology*, vol. 17, no. 1, pp. 91–100, 2011.
- [10] M. G. Bello, "A combined Markov random field and wavepacket transform-based approach for image segmentation," *IEEE Transactions on Image Processing*, vol. 3, no. 6, pp. 834–846, 1994.
- [11] R. N. Strickland and H. I. Hahn, "Wavelet transform matched filters for the detection and classification of microcalcifications in mammography," in *Proceedings of the International Conference on Image Processing*, vol. 1, pp. 422–425, Washington, DC, USA, 1995.
- [12] X. P. Zhang and M. D. Desai, "Segmentation of bright targets using wavelets and adaptive thresholding," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1020–1030, 2001.
- [13] L. A. Zadeh, "Fuzzy approach to color region extraction," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [14] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, Calif, USA, 1998.

- [15] R. Garnavi, M. Aldeen, and M. E. Celebi, “Weighted performance index for objective evaluation of border detection methods in dermoscopy images,” *Skin Research and Technology*, vol. 17, no. 1, pp. 35–44, 2011.
- [16] M. E. Celebi, G. Schaefer, and H. Iyatomi, “Objective evaluation of methods for border detection in dermoscopy images,” in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS’08)*, pp. 3056–3059, Vancouver, Canada, August 2008.
- [17] D. G. Fryback and J. R. Thornbury, “The efficacy of diagnostic imaging,” *Medical Decision Making*, vol. 11, no. 2, pp. 88–94, 1991.
- [18] The international atlas of dermoscopy and dermatoscopy, <http://www.dermoscopyatlas.com>.

Research Article

Software Tool for the Prosthetic Foot Modeling and Stiffness Optimization

Matija Štrbac¹ and Dejan B. Popović^{1,2}

¹Faculty of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia

²Aalborg University, Center for Sensory Motor Interaction, 9200 Aalborg, Denmark

Correspondence should be addressed to Matija Štrbac, sir.matija@gmail.com

Received 28 September 2011; Revised 31 December 2011; Accepted 5 January 2012

Academic Editor: Carlo Cattani

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

We present the procedure for the optimization of the stiffness of the prosthetic foot. The procedure allows the selection of the elements of the foot and the materials used for the design. The procedure is based on the optimization where the cost function is the minimization of the difference between the knee joint torques of healthy walking and the walking with the transfemoral prosthesis. We present a simulation environment that allows the user to interactively vary the foot geometry and track the changes in the knee torque that arise from these adjustments. The software allows the estimation of the optimal prosthetic foot elasticity and geometry. We show that altering model attributes such as the length of the elastic foot segment or its elasticity leads to significant changes in the estimated knee torque required for a given trajectory.

1. Introduction

Artificial feet (Figure 1) currently often used within transfemoral and transtibial prostheses comprise elements for heel impact absorption and energy accumulation during the loading of the leg. These features are instrumental for comfortable and energy-efficient gait [1]. The mechanics of the C-foot and various versions of the flexfoot-based feet follow the principles described in several patents [2–6].

The sketch of the foot (Figure 2) shows that the sole of the foot is elastic at the rear and front ends, while the area under the ankle joint is rigid.

In this paper, we use a biomechanical model that considers the elastic sole. The model that we adopted takes into account the longitudinal bending of the foot and carries over the kinematical and kinetic changes to other joints as has been done in previous studies [7]. We present here the procedure that allows the design of the leg based on the individual characteristics of the patient and the gait modality of interest.

The sketch shows the geometrical differences in the knee joint position during the gait (Figure 3) when the model with the flexible foot is used. The altered geometry results

with the difference in the distance between the knee joint and the point (COP) where the ground reaction force acts. This change results with the alteration of the knee joint torque that is required [8, 9] for the tracking of the desired trajectory. We suggest that the flexible bending model of the foot yields to reduced knee joint torques, ultimately reduces stress on the muscle activities of the knee extensors during the stance phase of the gait. In short, the goal of this study was to assist the individual design procedure of the prosthetic foot for the given user and the selected gait activity.

Human foot motion during stance can be described as a smooth progression between four phases: (1) heel strike, (2) flat foot, (3) heel lift, and (4) toe push-off. The model and proposed optimization methods are focused on the characteristics of the forefoot since the push-off phase and heel lift require the most energy during the gait cycle.

In order to justify the reasons for this research, we review shortly the state of the art that dominated recent literature. New designs and materials in prosthetic foot-ankle assemblies offer the individual with lower limb amputation wide choice among the prostheses with rather similar characteristic, which complicates the task of the prosthetic

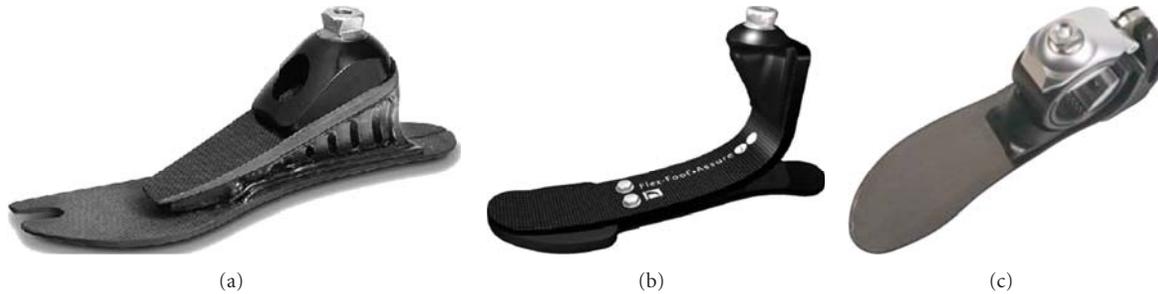


FIGURE 1: Flex-Foot Axia, Ossur, Island (a), Flex-Foot Assure, Ossur, Island (b), and 1C40 C-Walk, Otto Bock Germany (c). Adapted from [17, 18].

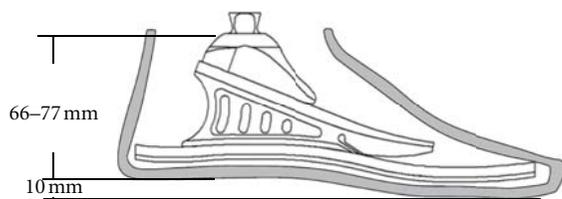


FIGURE 2: The sketch of the model of the Flex-Foot Axia, Össur, Island. Adapted from [17].

clinic team in formulating the most appropriate prescription for each patient. The most important clinical distinction is related to the internal design characteristics, which enable the component to simulate some actions of the human foot [10]. Over the last 20 years, many studies concentrated on comparing various properties of these assemblies [11–14] and came to conclusion that the optimal design choice depends on the individual walking style. However, the most popular prosthetic devices today (Carbon Copy II, Otto Bock SASH, Greissinger foot, SAFE foot, Seattle Natural, Flex-Foot) all consist of the (1) elastic keel which provides the energy transfer from the heel strike through to the toe off and the dorsiflexion required for natural ambulation and (2) the considerably stiff heel, which provides the impact absorption at heel strike and the kinetic energy required for a smooth transition between the heel strike and the toe off [1]. Our model incorporates these two elements and allows us to mathematically optimize their stiffness based on a given walking pattern for any geometry of the prosthetic foot.

This presentation concentrates on the optimization procedure. Thus, for the geometry in question, it is particularly important to find the upper bound of elasticity above which the foot becomes rigid and the lower bound below which the elasticity of the foot will result in a change in knee torque. In other words, we estimate the limits on the elasticity of a foot to match the patterns of healthy feet. The simulation uses gait and anthropometric data from healthy individuals. We present a software tool that can minimize knee torque based on biometric subject data.

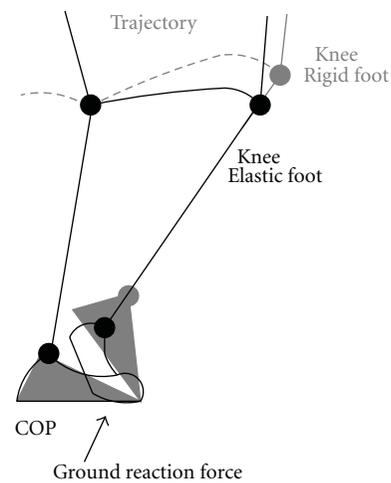


FIGURE 3: Trajectory of the knee joint for the rigid and elastic foot.

The results of this paper are relevant to the design of optimal geometry and stiffness identification for prosthetic feet. This study does not consider the powering of the ankle joint that has recently been introduced (e.g., PowerFoot One iWalk, Cambridge, MA or Proprio Foot Ossur, Reykjavik, Iceland [15]).

2. Methods

The mathematical model for inverse dynamics analysis of human gait is a set of equations based on the laws of momentum and kinetic momentum. The model presented is planar with only one active degree of freedom (DOF) at the knee joint and a bending foot that alters its geometry according to the load experienced along its length (Figure 4). The ankle joint motion was introduced since the obtained data are from the walking of the healthy individual; yet, we are not considering the power at this joint. This simplification was recommended in order to form a model that could produce the healthy human gait pattern. The list of notations is presented in the form of a table (Table 1).

TABLE 1: List of notations.

| Symbol | Definition |
|--------------|---|
| \ddot{x}_C | Horizontal acceleration of the leg |
| \ddot{y}_C | Vertical acceleration of the leg |
| COP | Position of the center of pressure |
| D_A | Length of the ankle segment |
| D_f | Length of the unyielding foot segment |
| D_T | Length of the toes segment |
| d_S | Distance from the knee to the center of mass |
| D_S | Length of the shank |
| x | Horizontal distance from the CM to COP |
| y | Vertical distance from the CM to COP |
| F_x | Horizontal ground reaction force |
| F_y | Vertical ground reaction force |
| δ_T | Amount of deflection in the elastic beam |
| φ_T | Angle of deflection in the elastic beam |
| φ_A | Externally controlled ankle joint angle |
| EI | Bending stiffness of an elastic segment |
| J_{Cz} | Moment of inertia for the leg and foot |
| m | Mass of the leg and foot |
| θ | Angle of the shank with respect to the horizontal |
| M_K | Joint torque at the knee joint |

If we assume the leg below the knee joint as one segment (no active joints), the kinetics is defined by three second-order equations, however, with the flexible geometry that introduces the ankle rotation:

$$m\ddot{x}_C = F_x - R_x, \quad (1)$$

$$m\ddot{y}_C = F_y - R_y - mg, \quad (2)$$

$$J_{Cz}\ddot{\theta} = F_{y-x} - F_{y-x} + R_x D_S \cos \theta - R_y D_S \sin \theta - M_K. \quad (3)$$

From (1) and (2), we can eliminate the knee driving forces in (3) and express the knee torque as a function of the kinematical data and the external, ground reaction forces:

$$M_K = (F_x - m\ddot{x}_C)D_S \cos \theta - F_y - m\ddot{y}_C - mg D_S \sin \theta - F_{y-x} + F_{x-y} - J_{Cz}\ddot{\theta}. \quad (4)$$

The segment kinematics need to be estimated. In our case, this was done by a stereometric system of cameras. The ground reaction forces have been estimated using the force platform. The parameters required for the optimization are also the horizontal and vertical distances from the center of mass (CM) and the position of the ground reaction force (COP) x and y .

The foot model was designed to include one elastic beam component that replaces the flexion of the metatarsal joint (D_T) and two rigid components that imitate the unyielding central portion of the human foot (D_f) and externally controlled vertical part of the foot (D_A). This model allows that the optimization uses the parameters of these three segments (geometry and elasticity). This allowed us to study

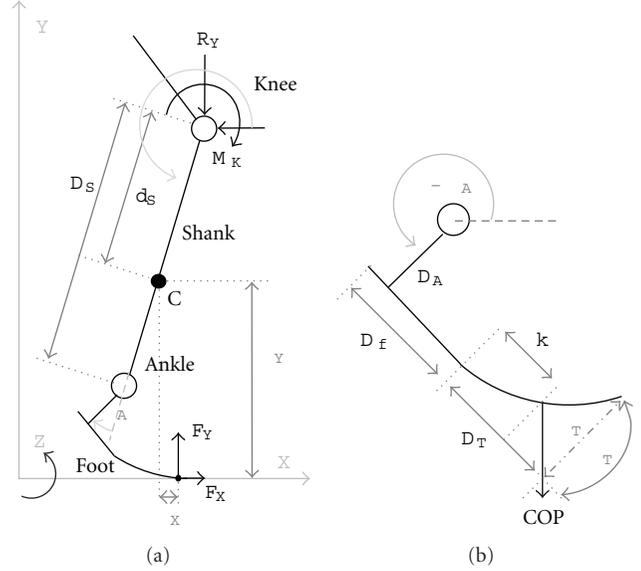


FIGURE 4: The model of the leg (a) and the foot (b) with the annotations used in the mathematical model.

the influence of the term $F_{y-x} + F_{x-y}$, which describes the moment acting at the center of mass.

The moment arm can be determined based on trigonometric formulas if the toe bending parameters represent the amount and angle of beam deflection:

$$\begin{aligned} x &= -(D_S - d_S) \sin \theta - D_A \sin \theta - \varphi_A \\ &\quad + D_f + D_T \cos \theta - \varphi_A + \delta_T \sin \theta - \varphi_A, \\ y &= -(D_S - d_S) \cos \theta - D_A \cos \theta - \varphi_A \\ &\quad - D_f + D_T \sin \theta - \varphi_A + \delta_T \cos \theta - \varphi_A. \end{aligned} \quad (5)$$

These formulas follow the leg and foot models shown in Figure 4.

According to Euler-Bernoulli beam theory, beam deflection and the angle of deflection in a simple cantilever beam can be described using the two following equations:

$$\delta_{\text{deflect}} = \frac{Pk^2}{6EI}(3L - k), \quad (6)$$

$$\varphi_{\text{deflect}} = \frac{Pk^2}{2EI}, \quad (7)$$

where P is the force perpendicular to the beam and exerted at the distance k along the length of the beam L . This is equivalent to a GRF being exerted on the foot at some distance k along the length of elastic segment. Hence, this force can be easily estimated:

$$P = F_x \sin \theta - \varphi_A + F_y \cos \theta - \varphi_A. \quad (8)$$

Since the foot is decomposed into simple beam structures, these equations are appropriate for the representation of geometrical variation required to calculate x and y during foot bending.

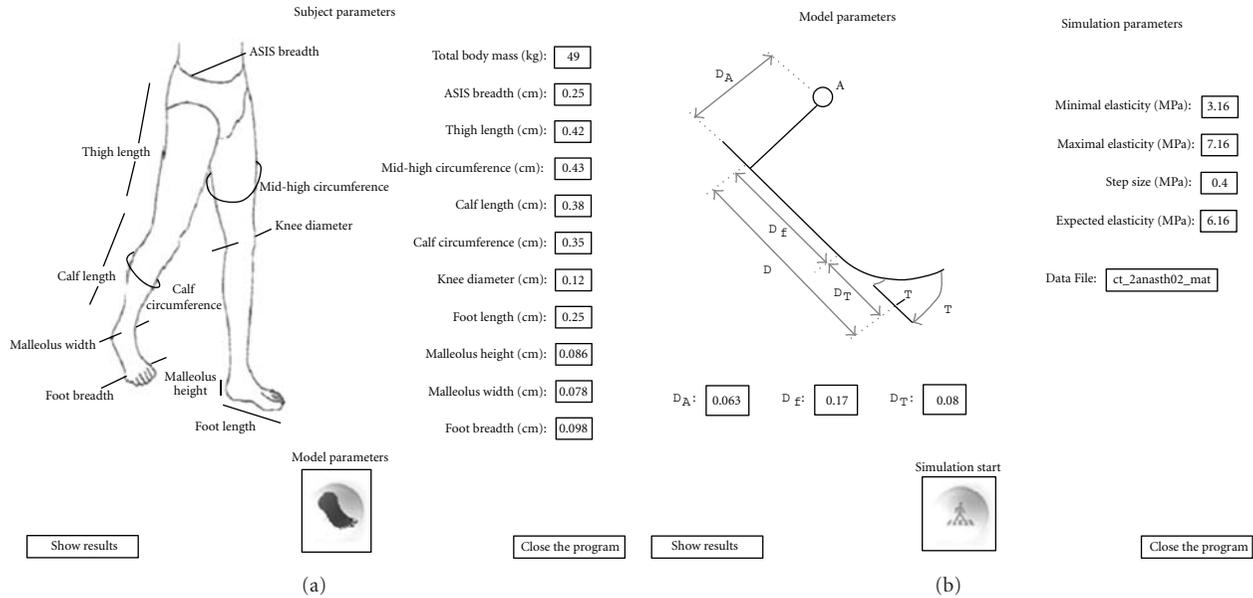


FIGURE 5: The graphical interface that defines inertial properties using the anthropometric data of the subject (a) and the graphical interface that defines the model's geometrical characteristics and simulation properties (b).

In the presented foot model, there are three geometrical parameters that can be altered to optimize a prosthesis: the length of the bending segment (D_T) and the lengths of the rigid structures (D_A, D_f). If these parameters are defined, based on (6) and (7), the only parameter left to alter is EI , which represents the stiffness of the beam. The cross-sectional moment of area I can be estimated based on the shape and geometry of the prosthetic foot of interest. In general, this moment varies throughout the foot because its geometry is not uniform. Thus, this value should be calculated as a length-dependent variable. In this presentation, we selected the cross-sectional moment of inertia $I = 1.8 \cdot 10^{-7} \text{ m}^4$. This value was selected as the typical value for some of the commercially available devices.

Here we concentrate on estimating the elasticity of the foot E for the selected value of the cross-sectional moment of inertia. However, this can be easily changed toward the analysis of the product IE ; hence the modified cross-sectional moment of inertia from the one presented here.

The software presented has been tested for various gait modalities. In this paper, we present only one dataset. These data come from the gait analysis of a female subject walking in a gait laboratory (Movement Analysis Laboratory, Rizzoli Orthopedic Institute, Bologna, Italy) and include measurements of body segment kinematics based on the CAST protocol [16]. The recordings include the ground reaction force estimates made using the Kistler force platform.

2.1. Optimization Software. A software tool for prosthetic foot analyses was developed using an interactive MATLAB-based application with a graphical user interface. The program initially shows the subject's anthropometric parameters, which will later define the segmental inertial properties at play (i.e., segment mass, the center of gravity, and the

moment of inertia). For these calculations, we used simple geometric modeling combined with the measured data in the same way as introduced in [19], bearing in mind that for the physically disabled, this process will be slightly different [20].

The anthropometric parameters shown in Figure 5 are the actual values for the real subject used in the simulation presented. The parameters need to be set before simulation to correspond to the subject whose gait is being analyzed. The button in the center of the screen will direct the user to the second screen, where he can determine the geometric characteristics of the prosthetic foot model and some of the simulation properties. The user can leave the program and the simulation at any time simply by clicking the button at the far right of the screen.

The second screen is similar to the first one, with a button in the center that starts the simulation. This screen also includes the fields for entering parameters that define the foot model and the fields determining the simulation process. The foot model parameters that need to be entered are dimensions of the one bending component D_T and two rigid segments D_f and D_A , as previously described. The lengths can be set, but the aggregate foot length (total length) must remain the same. The simulation parameters include the initial prosthetic foot elasticity and the optimization parameters. These parameters consist of the lower and upper bounds of elasticity and step size for the iterative search for the optimal modulus of elasticity. Initial elasticity is used only for the comparison of the final results where it will be presented in respect to the rigid foot model and the model with optimal elasticity, while the other parameters will have influence on the simulation as well as the final result. The last field on this screen contains the name of the file where the kinematical and dynamic data are stored. This data need

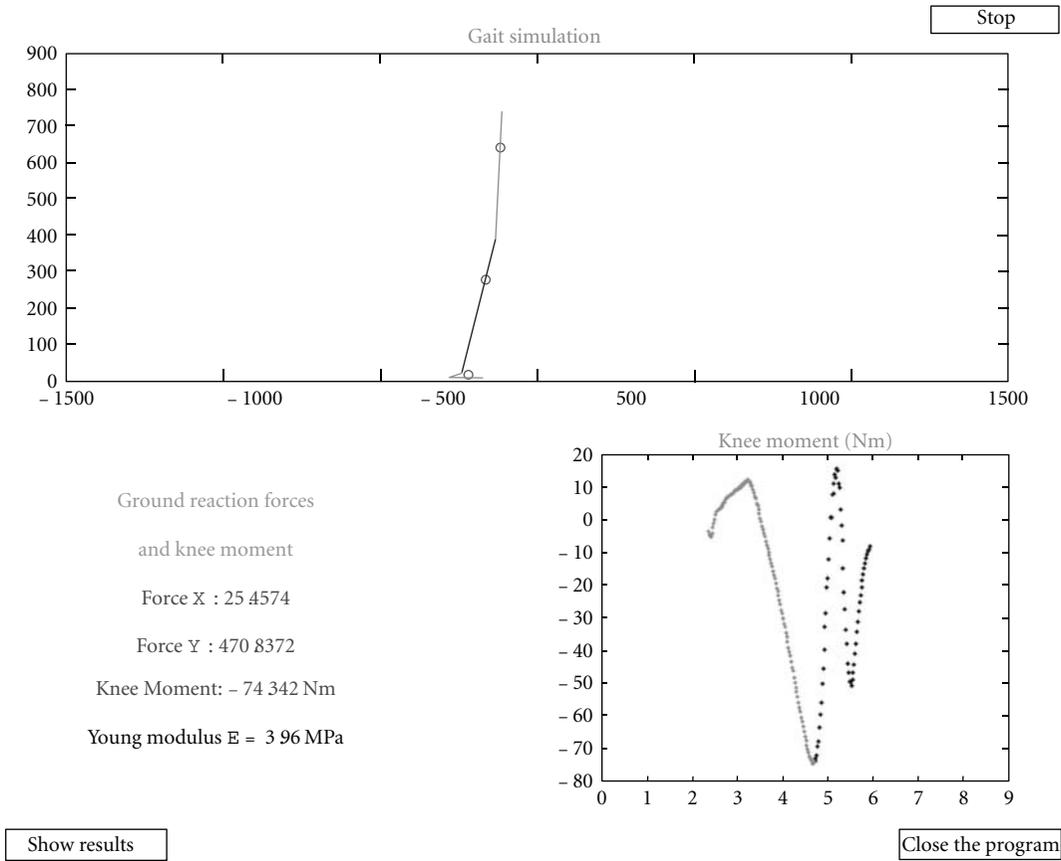


FIGURE 6: Simulation window and the values of interest in the captured instance of time.

to include hip, knee, and ankle joint center kinematics in sagittal plane, ground reaction forces, and the positions of those forces (i.e., COP).

The simulation screen includes two graphs and four text fields that track the changes in the parameters of interest in every sample of time (see Figure 6). Also, the center of mass is indicated for every segment with a circle. The figure in the upper half of the window shows the model position in the sagittal plane. The lower right part of the screen shows a graph that indicates the knee torque in the presented instance. Text fields to the left capture the vertical and the horizontal components of the ground reaction force, the knee torque in the presented time sample, and the modulus of elasticity in the current iteration. As previously mentioned, this simulation routine will be conducted at every step for the elasticity module within the range of the elasticity bounds defined in the simulation properties. This simulation protocol is intended to give the user the ability to monitor the effect of the model changes in the time domain and to identify the fragments of the gait cycle that the possible instabilities or peaks in the knee torque create. This insight into gait phase and model feature dependency thus helps to generate better model parameter allocation and better model formation.

Finally, for ease of understanding, the input variables and the optimized knee torque are presented as a function

of time. The input variables include joint center kinematics in the vertical and horizontal directions and the vertical and horizontal components of the ground reaction force. They are presented in line with the simulation results for better anticipation of the correlation. The output from the simulation includes three differently calculated knee torque signals. Two of them present different elasticity levels in the bending foot model, the optimal elasticity and the user-defined expected elasticity. The expected elasticity model is included in this software for three reasons: (1) as a reminder that the optimal knee moment does not always represent most human foot prostheses; (2) the optimization bounds should be set based on (6) and (7), and (3) to provide explicit predictions regarding different levels of model elasticity associated with the knee moment. The knee torque for the stiff foot is presented as the upper boundary of the knee torque in the elastic model. If the model knee torque exceeds this bound, then the lengths of the elastic segments must be changed.

3. Results

The input data required for the simulation are presented in Figure 7. The horizontal and vertical positions were used to estimate the joint angles required for the simulation.

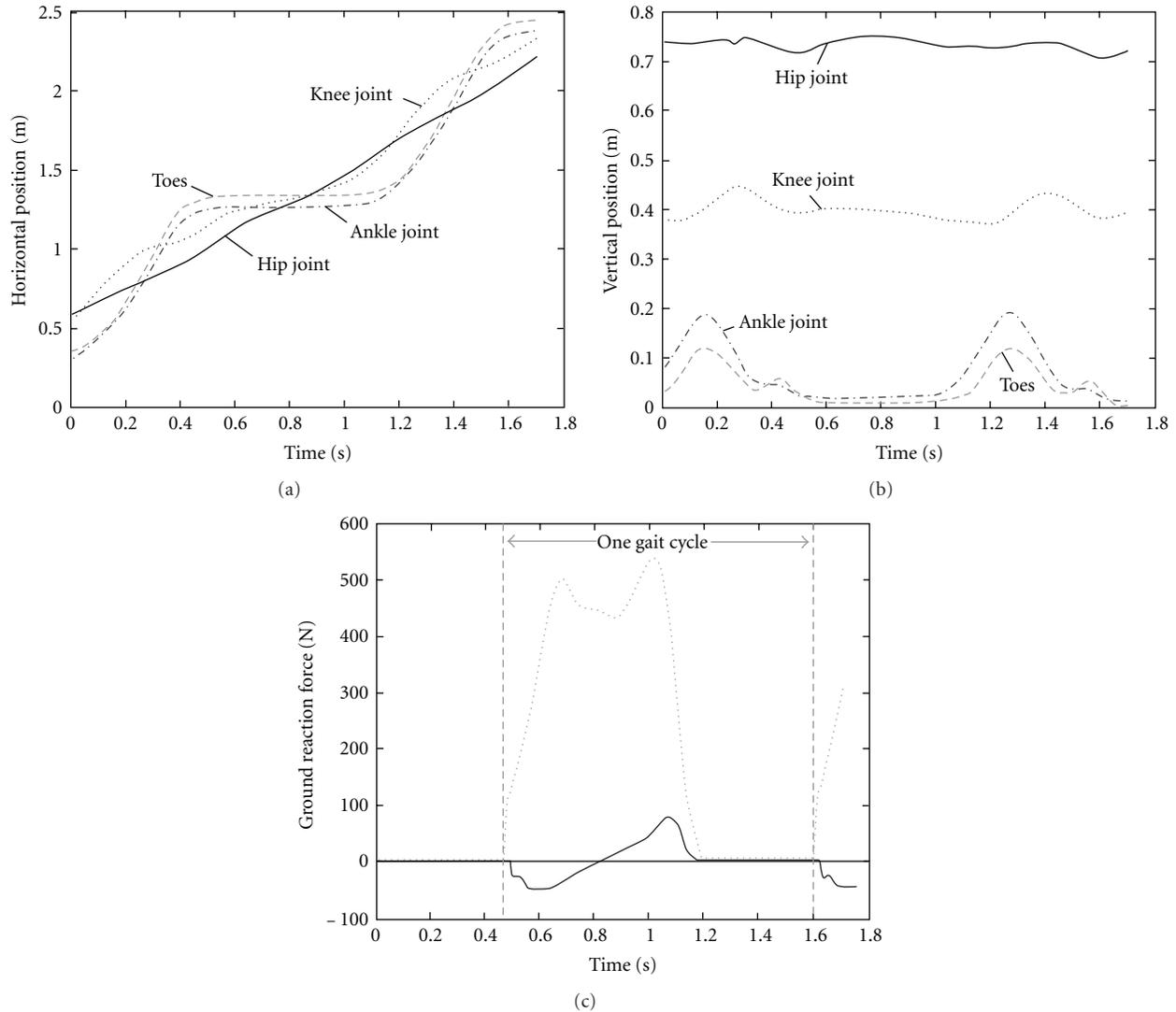


FIGURE 7: The kinematics and ground reaction forces used for simulation. The subject was a young female with body mass of 52 kg, height 1.64 m. The gait cycle lasted $T = 1.14$ s, and the stride length was 1.4 m ($v = 1.22$ m/s).

The center of pressure was estimated from the data recorded by the force platform.

We assumed that the maximal angle between the foot and the walking surface is approximately 30 degrees. Using the information about the maximum ground reaction force, we calculated (6) that the elasticity of a healthy foot is approximately 6.16 MPa. Based on this value, we chose an elasticity range between 4.16 and 8.16 MPa. We also set the upper and lower limits for maximum foot bending.

The software makes it possible to select any ratio between the lengths of the bending and stiff segments of the foot. The software estimates the optimal stiffness based on the minimal knee joint torque. From the results shown in Figure 8, it is easy to conclude that the elastic foot model produces lower knee joint torque than the stiff foot model ($\approx 50\%$) and that the increase in the ratio between the bending and stiff prosthetic segments improved this figure even more (≈ 10 Nm). The optimal elasticity in the first case was

$E = 4.36$ MPa (bending/stiff = 1 : 4); in the second case, it was 4.76 MPa (bending/stiff = 9 : 16).

4. Discussion

This paper presents the software developed for use in designing the optimal foot for prosthetic applications. The software also allows the selection of the appropriate motor unit if the prosthesis is to be externally driven. This is possible because the software determines the joint torque range required to track the desired trajectory and the dynamics that it needs to guaranty.

The software makes it possible to test the performance for various gait modalities and set the parameters to appropriate values for the amputee who would eventually use the transfemoral prosthesis.

We demonstrated that it is important to select the appropriate stiffness and geometry of the prosthetic foot to

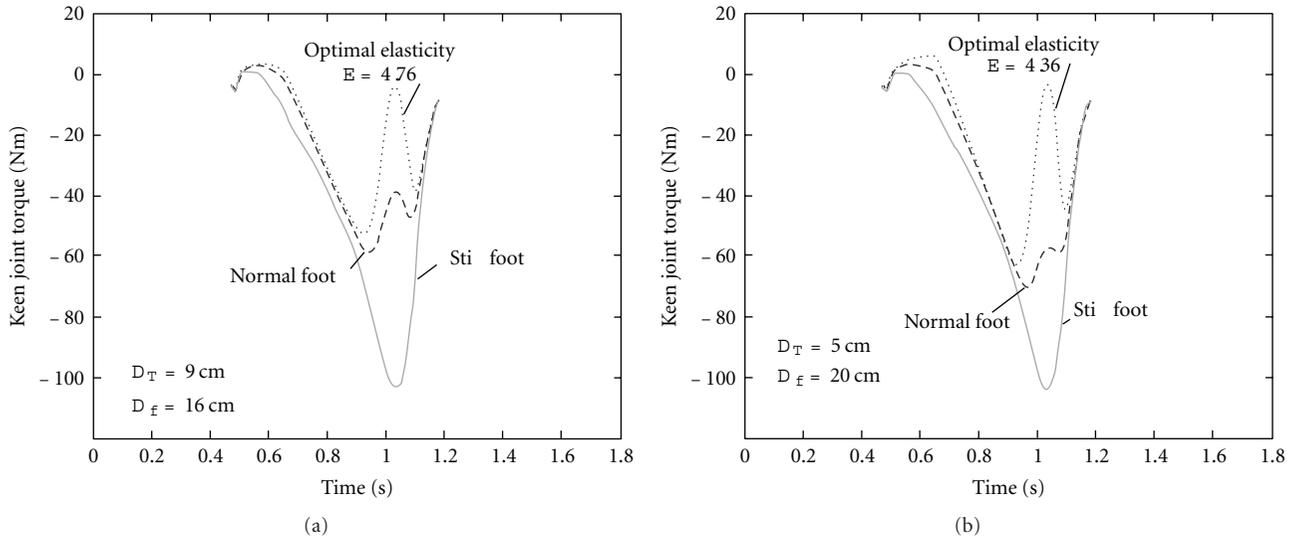


FIGURE 8: The knee joint torque calculated for a normal foot, a foot with optimal elasticity determined using the software presented in this study, and a stiff foot ($D_T + D_f = 25$ cm). (a) illustrates the case featuring a shorter bending segment ($D_T = 5$ cm), and (b) illustrates the case featuring a longer bending segment ($D_T = 9$ cm).

minimize the power needs at the knee joint. The shape of the knee joint torque will resemble the pattern of normal walking; thus, the joint torque at the hip (volitional control by the amputee) is likely to be almost normal.

The software also simulates walking on various terrains (on sloped ground, up the stairs, etc.), which is of great interest for multimodal control [21–23].

This model does not take into account the elastic deformation of the heel during initial ground contact or the impact on the heel during heel contact. These factors could be incorporated by including another bending beam on the back of the rigid segment. The elasticity and geometry of this beam could be analyzed using the methods presented here.

Acknowledgments

The work on this project was partly supported by the Ministry of Education and Science, Government of the Republic of Serbia (Project no. 175016). The authors acknowledge excellent technical support and assistance in gait measurements to Milica Djurić-Jovičić, Ph.D. student from the University of Belgrade.

References

- [1] R. Das, M. D. Burman, and S. Mohapatra, “Prosthetic foot design for transtibial prosthesis,” *Indian Journal of Biomechanics*, vol. 2009, pp. 142–146, 2009.
- [2] M. Naeder et al., “Prosthetic foot having Z shaped insert,” US Patent 5,062,859, 1991, <http://www.patents.com/us-5062859.html>.
- [3] D. L. Poggi, “Prosthetic foot having a cantilever spring keel,” US Patent 4,645,509, 1987, <http://www.patents.com/us-4645509.html>.
- [4] L. van Philips et al., “Composite prosthetic foot and leg,” US Patent 4,547,913, 1985, <http://www.patents.com/us-4547913.html>.
- [5] J. A. Sabolich et al., “Prosthetic foot,” US Patent 5,219,365, 1993, <http://www.patents.com/us-5219365.html>.
- [6] J. C. H. Goh, “Energy storing composite prosthetic foot,” US patent 5,258,039, 1993, <http://www.patents.com/us-5258039.html>.
- [7] F. Stefanovic and D. B. Popovic, “Control of the lower leg during walking: a versatile model of the foot,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 1, pp. 63–69, 2009.
- [8] D. Popovic, R. B. Stein, M. N. Oguztoreli, M. Lebedowska, and S. Jonic, “Optimal control of walking with functional electrical stimulation: a computer simulation study,” *IEEE Transactions on Rehabilitation Engineering*, vol. 7, no. 1, pp. 69–79, 1999.
- [9] S. Miyazaki, S. Yamamoto, and T. Kubota, “Effect of ankle-foot orthosis on active ankle moment in patients with hemiparesis,” *Medical and Biological Engineering and Computing*, vol. 35, no. 4, pp. 381–385, 1997.
- [10] J. E. Edelman, “Prosthetic feet: state of the art,” *Physical Therapy*, vol. 68, no. 12, pp. 1874–1881, 1988.
- [11] H. W. L. Van Jaarsveld, H. J. Grootenboer, J. De Vries, and H. F. J. M. Koopman, “Stiffness and hysteresis properties of some prosthetic feet,” *Prosthetics and Orthotics International*, vol. 14, no. 3, pp. 117–124, 1990.
- [12] A. P. Arya, A. Lees, H. C. Nirula, and L. Klenerman, “A biomechanical comparison of the SACH, Seattle and Jaipur feet using ground reaction forces,” *Prosthetics and Orthotics International*, vol. 19, no. 1, pp. 37–45, 1995.
- [13] A. E. Barr, K. L. Siegel, J. V. Danoff et al., “Biomechanical comparison of the energy-storing capabilities of SACH and Carbon Copy II prosthetic feet during the stance phase of gait in a person with below-knee amputation,” *Physical Therapy*, vol. 72, no. 5, pp. 344–354, 1992.

- [14] N. Mizuno, T. Aoyama, A. Nakajima, T. Kasahara, and K. Takami, "Functional evaluation by gait analysis of various ankle-foot assemblies used by below-knee amputees," *Prosthetics and Orthotics International*, vol. 16, no. 3, pp. 174–182, 1992.
- [15] J. A. Blaya and H. Herr, "Adaptive control of a variable-impedance ankle-foot orthosis to assist drop-foot gait," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 1, pp. 24–31, 2004.
- [16] M. G. Benedetti, F. Catani, A. Leardini, E. Pignotti, and S. Giannini, "Data management in gait analysis for clinical applications," *Clinical Biomechanics*, vol. 13, no. 3, pp. 204–215, 1998.
- [17] <http://www.ossur.com/?PageID=12639>.
- [18] http://www.ottobock.com/cps/rde/xchg/ob_com_en/hs.xsl/1792.html.
- [19] C. L. Vaughan, B. L. Davis, and J. C. Connor, *Dynamics of Human Gait*, Kiboho, Cape Town, South Africa, 1999.
- [20] R. B. Stein, E. P. Zehr, M. K. Lebidowska, D. B. Popovic, A. Scheiner, and H. J. Chizeck, "Estimating mechanical parameters of leg segments in individuals with and without physical disabilities," *IEEE Transactions on Rehabilitation Engineering*, vol. 4, no. 3, pp. 201–211, 1996.
- [21] A. H. Hansen, D. S. Childress, and S. C. Miff, "Roll-over characteristics of human walking on inclined surfaces," *Human Movement Science*, vol. 23, no. 6, pp. 807–821, 2004.
- [22] M. P. Dillon, S. Fatone, and A. H. Hansen, "Effect of prosthetic design on center of pressure excursion in partial foot prostheses," *Journal of Rehabilitation Research and Development*, vol. 48, no. 2, pp. 161–178, 2011.
- [23] M. P. Dillon and T. M. Barker, "Comparison of gait of persons with partial foot amputation wearing prosthesis to matched control group: observational study," *Journal of Rehabilitation Research and Development*, vol. 45, no. 9, pp. 1317–1334, 2008.

Research Article

On the Existence of Wavelet Symmetries in Archaea DNA

Carlo Cattani

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

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

Received 13 September 2011; Revised 27 October 2011; Accepted 29 October 2011

Academic Editor: Sheng-yong Chen

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

This paper deals with the complex unit roots representation of archaea DNA sequences and the analysis of symmetries in the wavelet coefficients of the digitalized sequence. It is shown that even for extremophile archaea, the distribution of nucleotides has to fulfill some (mathematical) constraints in such a way that the wavelet coefficients are symmetrically distributed, with respect to the nucleotides distribution.

1. Introduction

In some recent papers the existence of symmetries in nucleotide distribution has been studied for several living organisms [1–6] including mammals, fungi [1–4], and viruses [5, 6]. Thus showing that any (investigated) DNA sequence, when converted into a digital sequence, features some fractal shape of its DNA walk and an apparently random-like distribution. However, when the short wavelet transform maps the digital sequence into the space of wavelet coefficients, and these coefficients are clustered then they are located along some symmetrical shapes.

One of the main tasks of this paper is to show that although the distribution of nucleotide, in any DNA sequence, can be considered as randomly given, when we compare a random sequence (and the corresponding random walk) with a DNA sequence (and walk) it can be seen that there exists some distinctions. So that the nucleotides distribution seems to side with a random distribution with some constraints. These constraints (rules) are singled out in the following, by showing the existence of hidden geometry which underlies the structure of a DNA sequence.

In other words, nucleotides are distributed along any DNA sequence at first apparently randomly but at second analysis according to some (statistical) mathematical constraints which does not allow a given nucleotide to be arbitrarily followed by any other remaining nucleotides.

It is interesting to notice that even in the primitives organisms which billions of years ago have been colonizing

the earth under extreme conditions of life, their DNA has to fulfill the same constraints of the more evolved DNAs.

In order to achieve this goal some fundamental steps have to be taken into consideration and discussed.

- (1) Since DNA is a sequence of symbols, a map of these symbols into numbers has to be defined. In the following we will consider the complex unit roots map, which has the advantage of being unitary and distributed along the unit circle.
- (2) The indicator matrix is defined on the the indicator map. This matrix is important in order to draw the dot plot of the DNA sequence and from this plot we can see that apparently nucleotides seem to be randomly distributed. However, we will show by wavelet analysis that they look randomly distributed, while they are not.
- (3) The Ulam spiral adapted to DNA sequences is defined in order to single out some geometrical patterns.
- (4) Random walks on DNA, or short DNA walks, show that the random walks look like fractals.
- (5) The analysis of clusters of wavelet coefficients show that DNA walks have to fulfill some geometrical constraints.

In all DNA sequences, analyzed so far, for different kinds of living organisms, this geometrical symmetry [1–6] has been detected. In the following this analysis is extended also to archaea, since they might be considered at the early

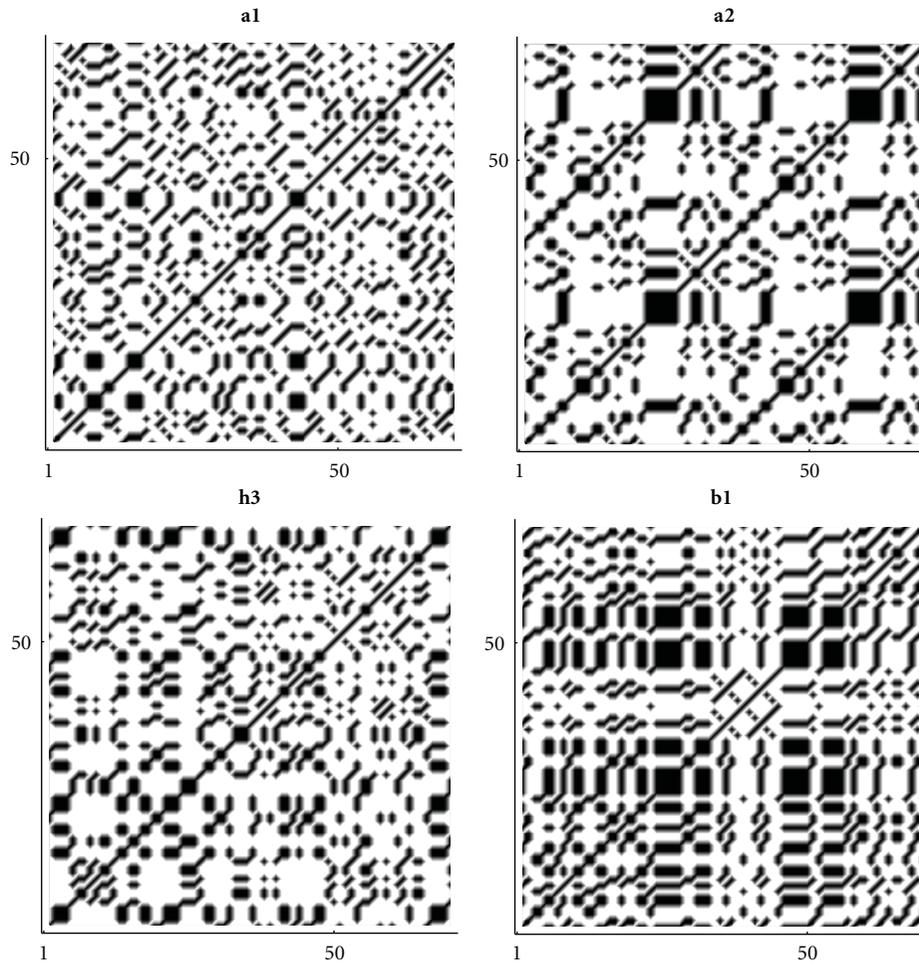


FIGURE 1: Indicator matrix for: (a1) pseudorandom 70-length sequence; (a2) pseudo-periodic 70-length sequence with period $\pi = 35$; (b1) 70-length DNA sequence of *Mycoplasma* KS1 bacter; (h3) 70-length DNA sequence of *Acidilobus* Archaea.

stage of life and their DNA is compared with more evolved microorganisms as bacteria.

It will be shown that, inspite of the many similarities with random sequences, only the wavelet analysis makes it possible to single out some distinctions. In particular, the wavelet coefficients of all (analyzed) organisms tend to fulfill a minimum principle for the energy of the signal. Also the archaea which often live in extreme environments have to fulfill the same geometrical rule of any other living organism.

The analysis of DNA by wavelets [7–9], as seen in [8–12], helps to single out local behavior and singularities [7, 13] or to express the scale invariance of coefficients [14]. Also multifractal nature of the time series [15–17] can be easily detected by wavelet analysis.

Some previous paper have studied various sequences of DNA such as leukemia tet variants, influenza viruses such as the A (H1N1) variant, mammalian, and a fungus (see [1–3, 14]) provided by the National Center for Biotechnology Information [18–21]. In all these papers it was observed

that DNA has to fulfill not only some chemical steady state given by the chemical ligands but also some symmetrical distribution of nucleotide along the sequence. In other words, base pairs have to be placed exactly in some positions.

According to previous results, it will be shown that as any other living organisms also these elementary organisms have DNA walks with fractal shape and wavelet coefficients bounded on a short-range wavelet transform. In other words, also anaerobic organism which should be understood as the most elementary at the first step of life have the same symmetries on wavelet coefficients as for more evolved organism, so that life has to fulfill some constrained distribution of nucleotides in order to give rise to some organism even at the most elementary step.

In particular, in Section 2, some remarks about the analysed data are given. Section 3 deals with some elementary plots which can easily visualize the distribution of nucleotides. The Ulam spiral plot is also proposed for the first time and it is observed a different distribution of weak/strong

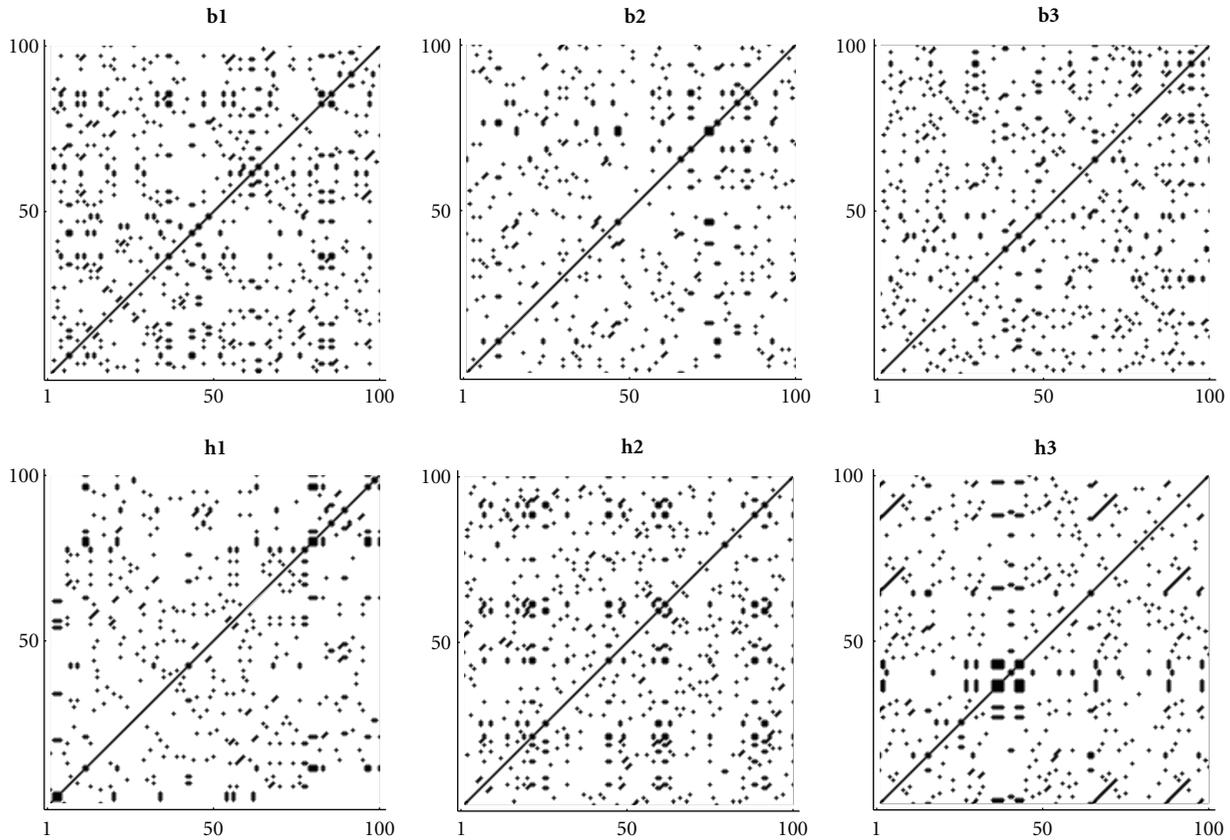


FIGURE 2: Indicator matrix for the first 100 amino acids of (h1) *Aeropyrum pernix* K1, (h2) *Acidianus hospitalis* W1, (h3) *Acidilobus saccharovorans* 345-15 (b1) *Mycoplasma putrefaciens* KS1, (b2) *Mortierella verticillata*, and (b3) *Blattabacterium* sp.

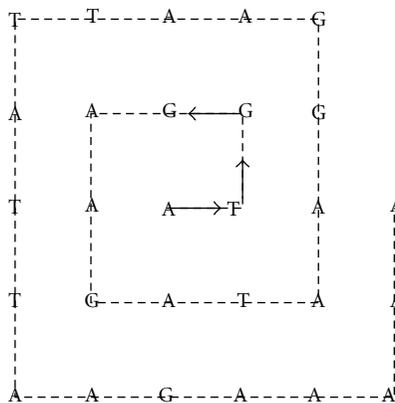


FIGURE 3: Distribution of nucleotides on a rectangular spiral.

hydrogen bonds. Section 4 provides some definitions about parameters of complexity. We will notice that all these parameters give rise to the same classification of organism. Section 4 proposes a complex numerical representation of DNA chains and random walks, while in final Section 6 the

short wavelet transform is given in order to single out some symmetries at the lower order of transform.

2. Materials and Methods

In the following we will take into consideration some genome, complete sequences of DNA, concerning the following archaea:

h1: *Aeropyrum pernix* K1, complete genome. DNA, circular, 1669696 bp, [18–21], accession BA000002.3. Lineage: *Archaea*; Crenarchaeota; Thermoprotei; Desulfurococcales; Desulfurococcaceae; *Aeropyrum*; *Aeropyrum pernix*; *Aeropyrum pernix* K1.

This organism, which was the first strictly aerobic hyperthermophilic archaeon sequenced, was isolated from sulfuric gases in Kodakara-Jima Island, Japan in 1993.

h2: *Acidianus hospitalis* W1, complete genome. DNA, circular, 2137654 bp, [18–21], accession CP002535. Lineage: *Archaea*; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae; *Acidianus*; *Acidianus hospitalis*; *Acidianus hospitalis* W1.

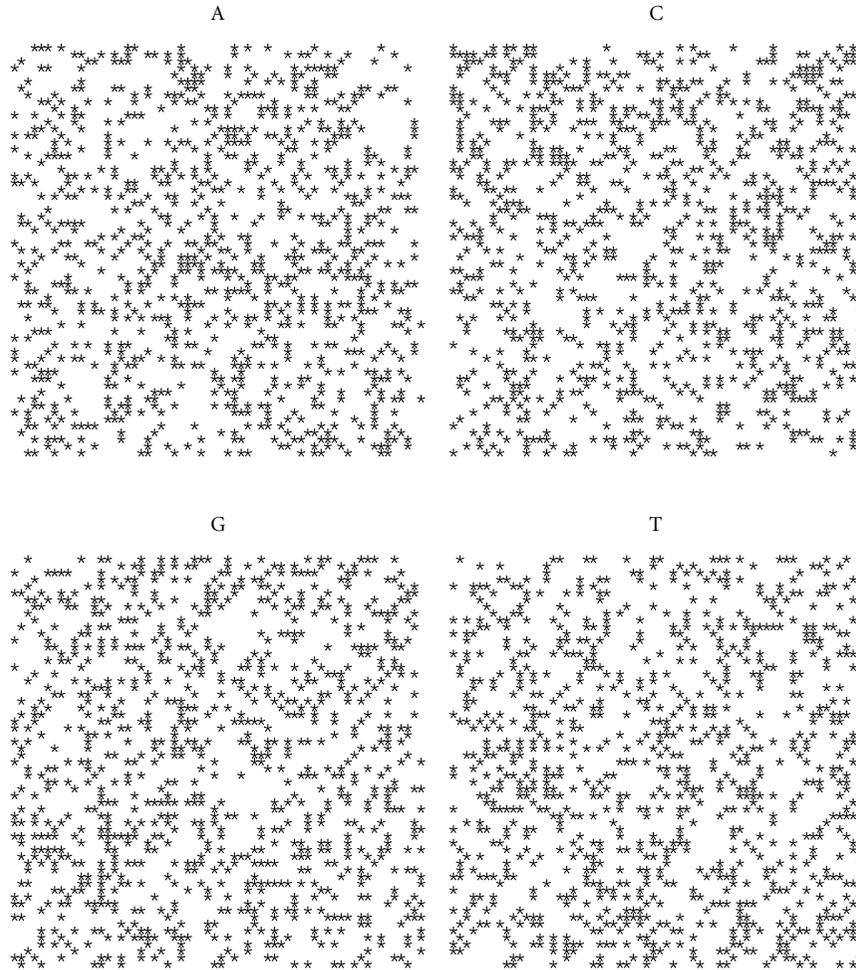


FIGURE 4: Spiral distribution of the first 3752 nucleotides for the random sequence.

h3: *Acidilobus saccharovorans* 345-15. complete genome. DNA, circular, 2137654 bp, [18–21], accession CP001742.1. Lineage: *Archaea*; Crenarchaeota; Thermoprotei; Acidilobales; Acidilobaceae; Acidilobus; *Acidilobus saccharovorans*; *Acidilobus saccharovorans* 345-15. Anaerobic bacteria found in hot springs.

to be compared with the following (aerobic/anaerobic) bacteria/fungi:

b1: *Mycoplasma putrefaciens* KS1 chromosome, complete genome. DNA, circular, length 832603 bp, [18–21], accession NC 015946. Lineage: *Bacteria*; Tenericutes; Mollicutes; Mycoplasmatales; Mycoplasmataceae; Mycoplasma; *Mycoplasma putrefaciens*; *Mycoplasma putrefaciens* KS1.

b2: *Mortierella verticillata* mitochondrion, complete genome. dsDNA, circular, length 58745 bp, [18–21], accession NC 006838. Lineage: Eukaryota; Opisthokonta; *Fungi*; *Fungi* incertae sedis; Basal fungal

lineages; Mucoromycotina; Mortierellales; Mortierellaceae; *Mortierella*; *Mortierella verticillata*.

b3: *Blattabacterium* sp. (*Periplaneta Americana*) str. BPLAN, complete genome. DNA, circular, length 636994 nt, [18–21], accession NC 013418. Lineage: *Bacteria*; Bacteroidetes/Chlorobi group; Bacteroidetes; Flavobacteria; Flavobacteriales; Blattabacteriaceae; Blattabacterium; Blattabacterium sp. (*Periplaneta Americana*); *Blattabacterium* sp. (*Periplaneta Americana*) str. BPLAN.

Moreover we will compare DNA sequences with artificial sequences of nucleotides randomly taken (see Section 4).

2.1. Archaea. Archaea are a group of elementary single-cell microorganisms, having no cell nucleus or any other membrane-bound organelles within their cells. They are similar to bacteria, since they have the same size and shape (apart few exceptions) and the generally similar cell structure. However, the evolutionary history of archaea and their biochemistry has significant differences with regard to other forms of life.

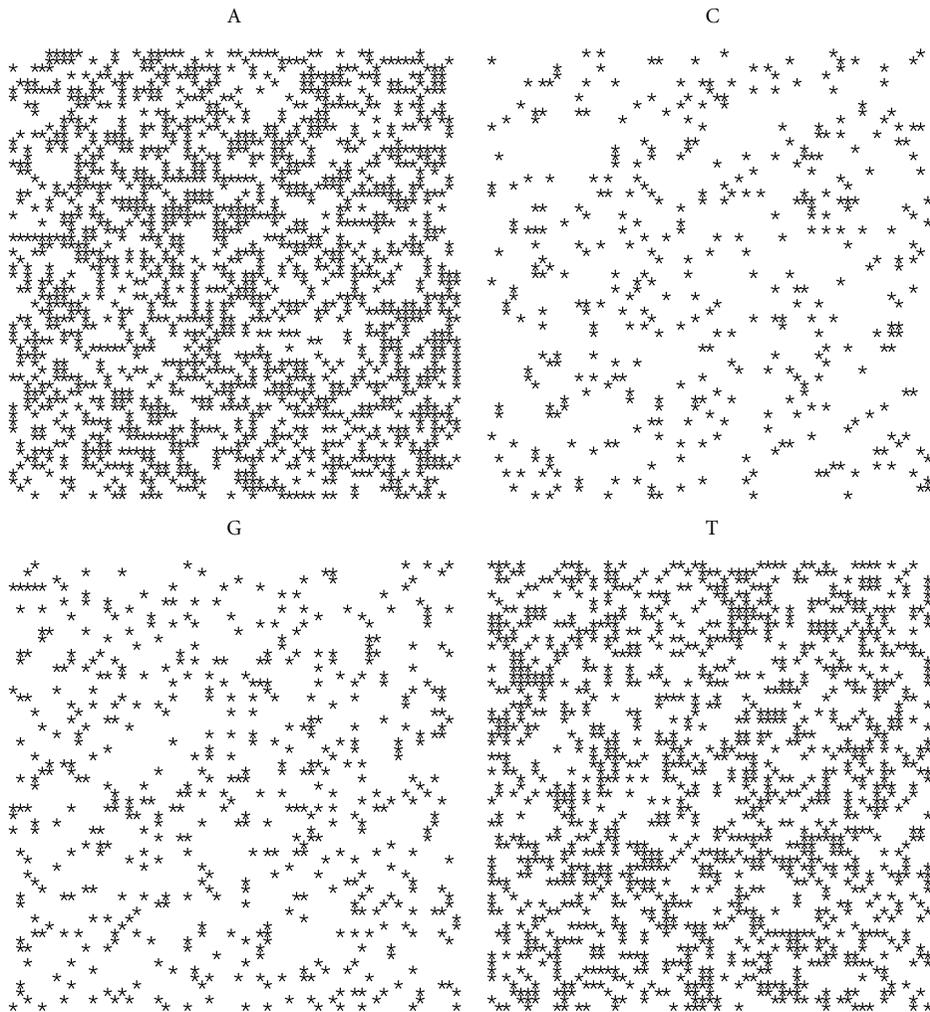


FIGURE 5: Spiral distribution of the first 3752 nucleotides for *Mycoplasma putrefaciens* KS1.

Therefore they are considered as members of a phylogenetic group distinct from bacteria and eukaryota.

Archaea during their evolution have been spreading all over the Earth in almost all habitats [22, 23] existing in a broad range of habitats, being one of the major contribution (20%) to earth's biomass. The most peculiar feature of archaea is that they can live in some environments with extreme life conditions (thus being considered as extremophiles [22, 24]). Indeed, some archaea survive to high temperatures, over 100°C, while others can live in very cold habitats or highly saline, acidic, or alkaline water. Nevertheless some archaea are living in mild conditions.

It has been also recognized that the archaea may be the most ancient organisms on the Earth, so that archaea, and eukaryotes are probably diverged early from an ancestral colony of organisms.

We will see, in the following, that archaea DNA it looks very close to random sequences so that we can assume that

the ancestral organism were evolving by random permutations from a primitive assembly of nucleotides. So that the evolution can be seen as a tendency to a steady state far from the randomness. Therefore, the bacteria's DNA (and other eukaryotes' [1–6]), as a result of the evolution, shows the existence of some hidden stability.

3. Correlation Plots

In this section we will consider some elementary plots from where it is possible to visualize autocorrelation, distribution law of nucleotides and to measure some fundamental parameters by using frequency count.

Let

$$\mathcal{A} \stackrel{\text{def}}{=} \{A, C, G, T\} \quad (1)$$

be the finite set (alphabet) of nucleotides (nucleic acids): adenine (A), cytosine (C), guanine (G), thymine (T), and

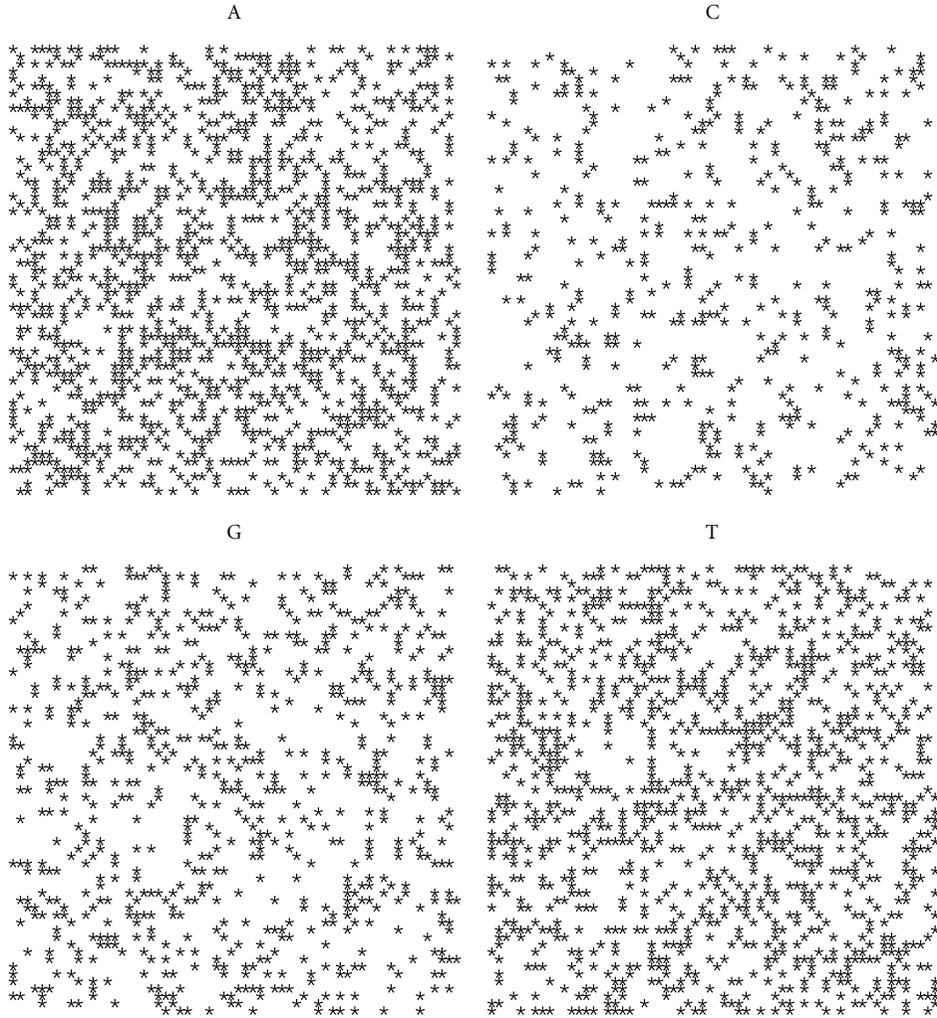


FIGURE 6: Spiral distribution of the first 3752 nucleotides for *Mortierella verticillata*.

$x \in \mathcal{A}$ any member of the alphabet. Nucleic acids are further grouped according to their ligand properties as

- (a) purine {A, G}, pyrimidine {C, T},
- (b) amino {A, C}, keto {G, T},
- (c) weak hydrogen bonds {A, T}, strong hydrogen bond {G, C}.

A DNA sequence is the finite symbolic sequence

$$\mathcal{S} = \mathbb{N} \times \mathcal{A} \quad (2)$$

so that

$$\mathcal{S} \stackrel{\text{def}}{=} \{x_h\}_{h=1, \dots, N}, \quad N < \infty \quad (3)$$

with

$$x_h \stackrel{\text{def}}{=} (h, x) = x(h), \quad (h = 1, 2, \dots, N; x \in \mathcal{A}) \quad (4)$$

being the nucleotide x at the position h .

In general we can define an ℓ -length alphabet as follows: let the ℓ -length DNA word be defined by the ℓ -combination of the 4 nucleotides (1). For each fixed length ℓ there are 4^ℓ words, however not all of them can be considered, from biological point of view, as independent instances (see, e.g., Table 1), for this we define the ℓ -length alphabet as the set of ℓ -length independent words:

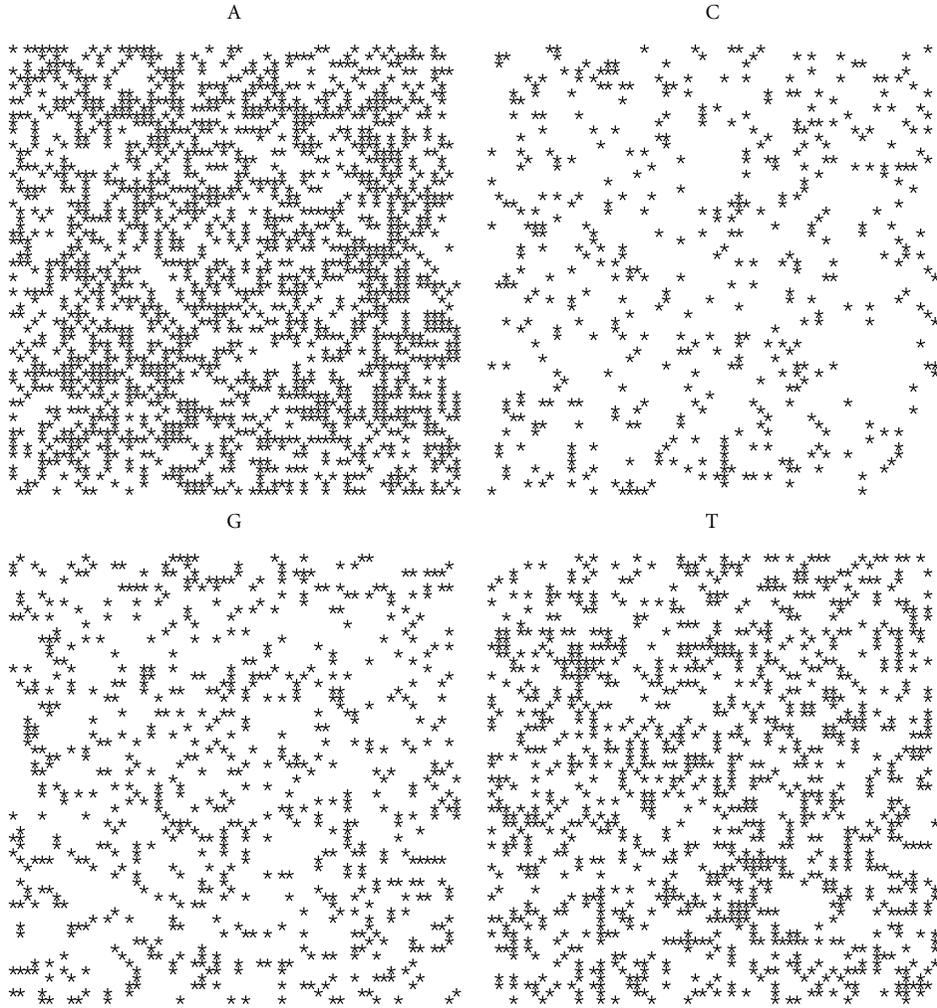
$$\mathcal{A}_\ell \stackrel{\text{def}}{=} \{a_1, a_2, \dots, a_{M_\ell}\}, \quad M_\ell \stackrel{\text{def}}{=} |\mathcal{A}_\ell| \leq 4^\ell \quad (5)$$

with $|\dots|$ cardinality of the set and

$$\ell \stackrel{\text{def}}{=} \text{length}(a_j), \quad (j = 1, \dots, M_\ell). \quad (6)$$

For instance with $\ell = 1$, the alphabet is $\mathcal{A}_1 = \mathcal{A} = \{A, C, G, T\}$, with $\ell = 3$ the alphabet is given by the 20 amino acids

$$\mathcal{A}_3 = \{M, E, Q, D, R, T, N, H, V, G, L, S, P, F, I, C, A, K, Y, W\} \quad (7)$$


 FIGURE 7: Spiral distribution of the first 3752 nucleotides for *Blattabacterium* sp.

each amino acid being represented by a 3-length word of Table 1.

Let \mathcal{S}_N be an N -length ordered sequence of nucleotides $\{A, C, G, T\}$ and \mathcal{A}_ℓ the chosen alphabet, a DNA sequence of words is the finite symbolic sequence

$$\mathcal{D}_\ell(S_N) = \mathbb{N} \times \mathcal{A}_\ell \quad (8)$$

so that

$$\mathcal{D}_\ell(S_N) \stackrel{\text{def}}{=} \{x_h\}_{h=1, \dots, N}, \quad (x \in \mathcal{A}_\ell; N < \infty) \quad (9)$$

with

$$x_h \stackrel{\text{def}}{=} (h, x), \quad (h = 1, 2, \dots, N; x \in \mathcal{A}_\ell) \quad (10)$$

being the word x at the position h .

3.1. Indicator Matrix. The 2D indicator function, based on the 1D definition given in [25], is the map

$$u : \mathcal{S} \times \mathcal{S} \longrightarrow \{0, 1\} \quad (11)$$

such that

$$u(x_h, x_k) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x_h = x_k, \\ 0 & \text{if } x_h \neq x_k, \end{cases} \quad (x_h \in \mathcal{S}, x_k \in \mathcal{S}), \quad (12)$$

with

$$u(x_h, x_k) = u(x_k, x_h), \quad u(x_h, x_h) = 1 \quad (13)$$

and, where for short, we have assumed

$$\mathcal{S} \stackrel{\text{def}}{=} \mathcal{D}_1(S_N). \quad (14)$$

According to (12), the indicator of an N -length sequence can be easily represented by the $N \times N$ sparse symmetric matrix of binary values $\{0, 1\}$ which results from the indicator matrix (see also [3–5])

$$u_{hk} \stackrel{\text{def}}{=} u(x_h, x_k), \quad (x_h \in \mathcal{S}, x_k \in \mathcal{S}; h, k = 1, \dots, N), \quad (15)$$

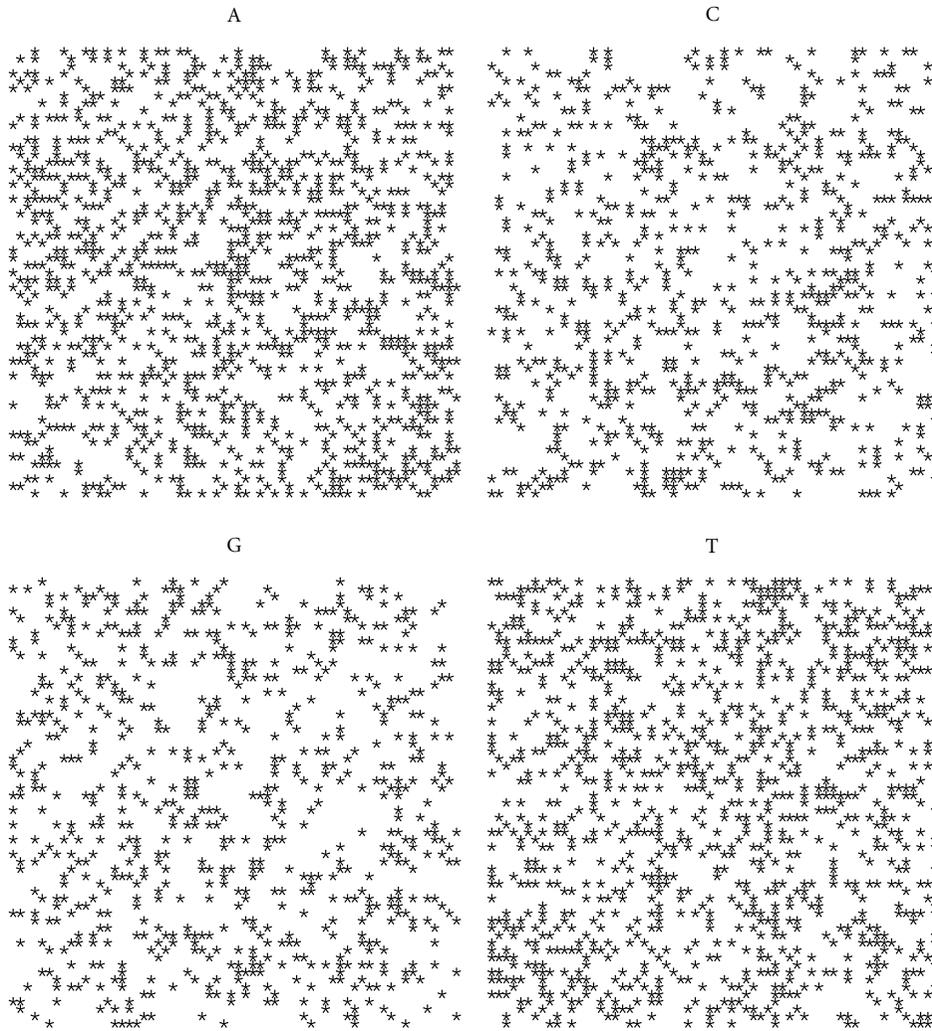


FIGURE 8: Spiral distribution of the first 3752 nucleotides for *Aeropyrum pernix* K1.

being, explicitly

$$\begin{array}{c|cccccccccccc}
 & \vdots & \dots \\
 \text{G} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots & \\
 \text{C} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\
 \text{A} & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & \dots & \\
 \text{A} & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & \dots & \\
 \text{T} & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots & \\
 \text{A} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & \dots & \\
 \text{C} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \dots & \\
 \text{T} & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots & \\
 \text{G} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots & \\
 \text{A} & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & \dots & \\
 \hline
 u_{hk} & \text{A} & \text{G} & \text{T} & \text{C} & \text{A} & \text{T} & \text{A} & \text{A} & \text{C} & \text{G} & \dots &
 \end{array} \tag{16}$$

This squared matrix can be plotted in 2 dimensions by putting a black dot where $u_{hk} = 1$ and white spot when $u_{hk} = 0$ (Figure 1) thus giving rise to the two-dimensional dot plot, which is a special case of the *recurrence plot* [26].

A simple generalization of this matrix can be considered for the alphabets \mathcal{A}_ℓ , as follows. By choosing the 3 alphabet of amino acids, the 2D indicator function is the map

$$u : \mathcal{D}_3(S_N) \times \mathcal{D}_3(S_N) \longrightarrow \{0, 1\} \tag{17}$$

such that

$$u(x_h, x_k) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x_h = x_k, \\ 0 & \text{if } x_h \neq x_k, \end{cases} \quad (x_h \in \mathcal{D}_3(S_N), x_k \in \mathcal{D}_3(S_N)), \tag{18}$$

with

$$u(x_h, x_k) = u(x_k, x_h), \quad u(x_h, x_h) = 1. \tag{19}$$

TABLE 1: Correspondence codons to amino acids.

| | | Amino acid | Codon |
|----|---|---------------|------------------------------|
| 1 | M | Methionine | ATG |
| 2 | E | Glutamic acid | GAA, GAG |
| 3 | Q | Glutamine | CAA, CAG |
| 4 | D | Aspartic acid | GAT, GAC |
| 5 | R | Arginine | CGT, CGC, CGA, CGG, AGA, AGG |
| 6 | T | Threonine | ACT, ACC, ACA, ACG |
| 7 | N | Asparagine | AAT, AAC |
| 8 | H | Histidine | CAT, CAC |
| 9 | V | Valine | GTT, GTC, GTA, GTG |
| 10 | G | Glycine | GGT, GGC, GGA, GGG |
| 11 | L | Leucine | TTA, TTG, CTT, CTC, CTA, CTG |
| 12 | S | Serine | TCT, TCC, TCA, TCG, AGT, AGC |
| 13 | P | Proline | CCT, CCC, CCA, CCG |
| 14 | F | Phenylalanine | TTT, TTC |
| 15 | I | Isoleucine | ATT, ATC, ATA |
| 16 | C | Cysteine | TGT, TGC |
| 17 | A | Alanine | GCT, GCC, GCA, GCG |
| 18 | K | Lysine | AAA, AAG |
| 19 | Y | Thyroxine | TAT, TAC |
| 20 | W | Tryptophan | TGG |
| | | Stop | TAA, TAG, TGA |

According to (12), the indicator, on the 3-alphabet of amino acids of an N -length sequence can be easily represented by the $N \times N$ sparse symmetric matrix of binary values $\{0, 1\}$:

$$u_{hk} \stackrel{\text{def}}{=} u(x_h, x_k), \quad (20)$$

$$(x_h \in \mathcal{D}_3(S_N), x_k \in \mathcal{D}_3(S_N); h, k = 1, \dots, N),$$

being, explicitly

$$\begin{array}{c|cccccccccccc}
 & \vdots & \dots \\
 \text{M} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \\
 \text{Q} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\
 \text{R} & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\
 \text{T} & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & \dots \\
 \text{T} & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & \dots \\
 \text{E} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\
 \text{R} & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\
 \text{R} & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\
 \text{K} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\
 \text{M} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \\
 \hline
 u_{hk} & \text{M} & \text{K} & \text{R} & \text{R} & \text{E} & \text{T} & \text{T} & \text{R} & \text{Q} & \text{M} & \dots & \dots
 \end{array} \quad (21)$$

With the graphical representation of this matrix we can also show the correlation of amino acids.

3.2. *Test Sequences.* In the following, in order to single out the main features of biological sequences, we will compare the DNA sequence with some test sequences.

- (1) Pseudorandom N -length sequence of nucleotides is the sequence $\{\mathcal{R}_i\}_{i=1, \dots, N}^\ell$ where r_i is a symbol randomly chosen in the alphabet \mathcal{A}_ℓ , like for example, ($\ell = 1$):

$$\{A, C, A, G, T, A, T, G, G, A, T, T, A, C, C, G, \dots\}. \quad (22)$$

- (2) Pseudoperiodic N -sequence of nucleotides with period π is the direct sum of a given π -length pseudorandom sequence, such that $N = k\pi$, ($k \in \mathbb{N}$) and $\mathcal{R}_i = \mathcal{R}_{i+\pi}$, for example,

$$\{A, C, A, G, A, C, A, G, A, C, A, G, A, C, A, G, \dots\}, \quad (\pi = 4). \quad (23)$$

When $\pi = 1$ we have a pseudorandom sequence.

If we plot the indicator matrix of some bacteria and compare it with a pseudorandom and periodic sequence, we can see that (Figure 1)

- (1) the main diagonal is a symmetry axis for the plot;
- (2) there are some motifs which are repeated at different scales like in a fractal;
- (3) periodicity is detected by parallel lines to the main diagonal (Figure 1(a2));

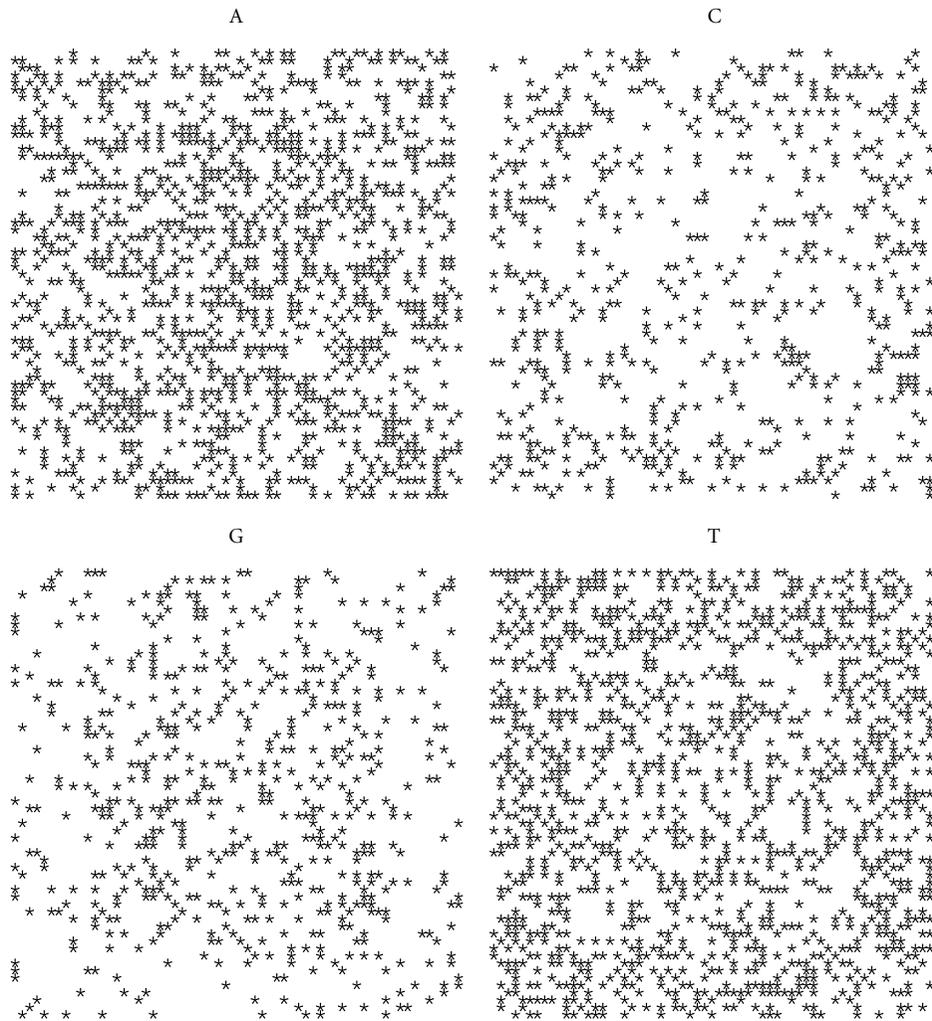


FIGURE 9: Spiral distribution of the first 3752 nucleotides for *Acidianus hospitalis* W1.

- (4) empty spaces are more distributed than filled spaces, in the sense that the matrix u_{hk} is a sparse matrix (having more 0's than 1's);
- (5) it seems that there are some square-like islands where black spots are more concentrated; these islands show the persistence of a nucleotide (Figures 1(a2) and 1(b1));
- (6) the dot plot of archaea is very similar to the dot plot of a random sequence (Figures 1(a1) and 1(h3)).

It can be noticed that DNA sequences of a living organism resemble (Figure 1) random sequences, with some short range influence, built on the same alphabet. This has been taken as an axiom of nucleotides distribution, so that DNA sequences are often considered as Markov chain [27]. However, there are some hidden rules in combining the nucleotides and these rules lead, during the evolution, to a steady distribution. In fact, the more primitive the sequence is, the more randomly distributed the nucleotides are. It seems that

as a consequence of the evolution, nucleotides move from a disordered aggregation toward a more organized structure, shown by the growing islands in the dot plot. The biological evolution is such that the challenge for the self-organization might follow from random permutations of a primitive disordered sequence so that the organization, that is, the complexity, is only the result of many arbitrary permutations of randomness. During the challenge for complexity, DNA sequence becomes “less random” and it loses some kind of energy.

From the graphical representation of the indicator matrix for bacteria and amino acids we can see a more sparse matrix, but with some typical plots (Figure 2).

3.3. Spiral Plot. In this section we consider a 2D distribution of nucleotides, following the idea given by Ulam for the distribution of primes, along an Ulam-like spiral [28]. In order to find some patterns in their distribution, nucleotides are arranged along a rectangular spiral. This is equivalent to

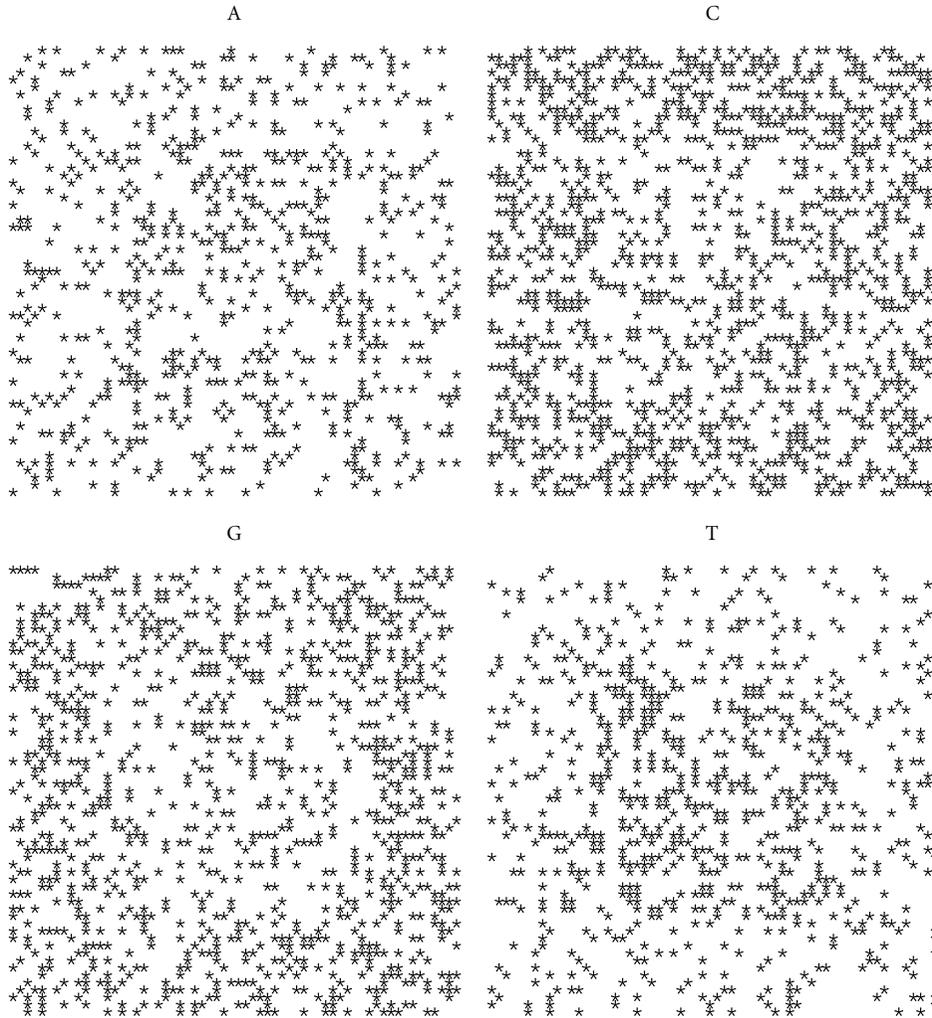


FIGURE 10: Spiral distribution of the first 3752 nucleotides for *Acidilobus saccharovorans* 345-15.

mapping the 1D sequence of integers into a 2D sequence as follows:

$$\begin{array}{lll}
 X_1 & 1 & \{0, 0\} \\
 X_2 & 2 & \{1, 0\} \\
 X_3 & 3 & \{1, 1\} \\
 X_4 & 4 & \{0, 1\} \\
 X_5 & 5 & \{-1, 1\} \\
 X_6 & 6 & \{-1, 0\} \\
 X_7 & 7 & \{-1, -1\} \\
 X_8 & 8 & \{0, -1\} \\
 X_9 & 9 & \{1, -1\} \\
 X_{10} & 10 & \{2, -1\} \\
 X_{11} & 11 & \{2, 0\} \\
 \vdots & \vdots & \vdots
 \end{array} \tag{24}$$

For instance the sequence

$$\{A, T, G, G, A, A, G, A, T, A, A, G, \dots\} \tag{25}$$

distributed along the spiral looks like Figure 3.

For each nucleotide we can draw a spiral containing the distribution of only one acid nucleic. To each organism there correspond four plots, for A, C, G, T, respectively.

Let us first note that on a random sequence (Figure 4) the four distribution are equivalent.

By comparing the spirals of bacteria, random and archaea (Figures 4, 5, 6, 7, 8, 9, 10) we can see that there is a different distribution of each nucleotide. However the more evolved organism tends to have a higher percentage of weak hydrogen bonds (Figures 5, 6 and 7), so that we can assume the following.

Conjecture 1. *During the evolution, the distribution of nucleotides changes in a such way that strong hydrogen bonds tend to become weak.*

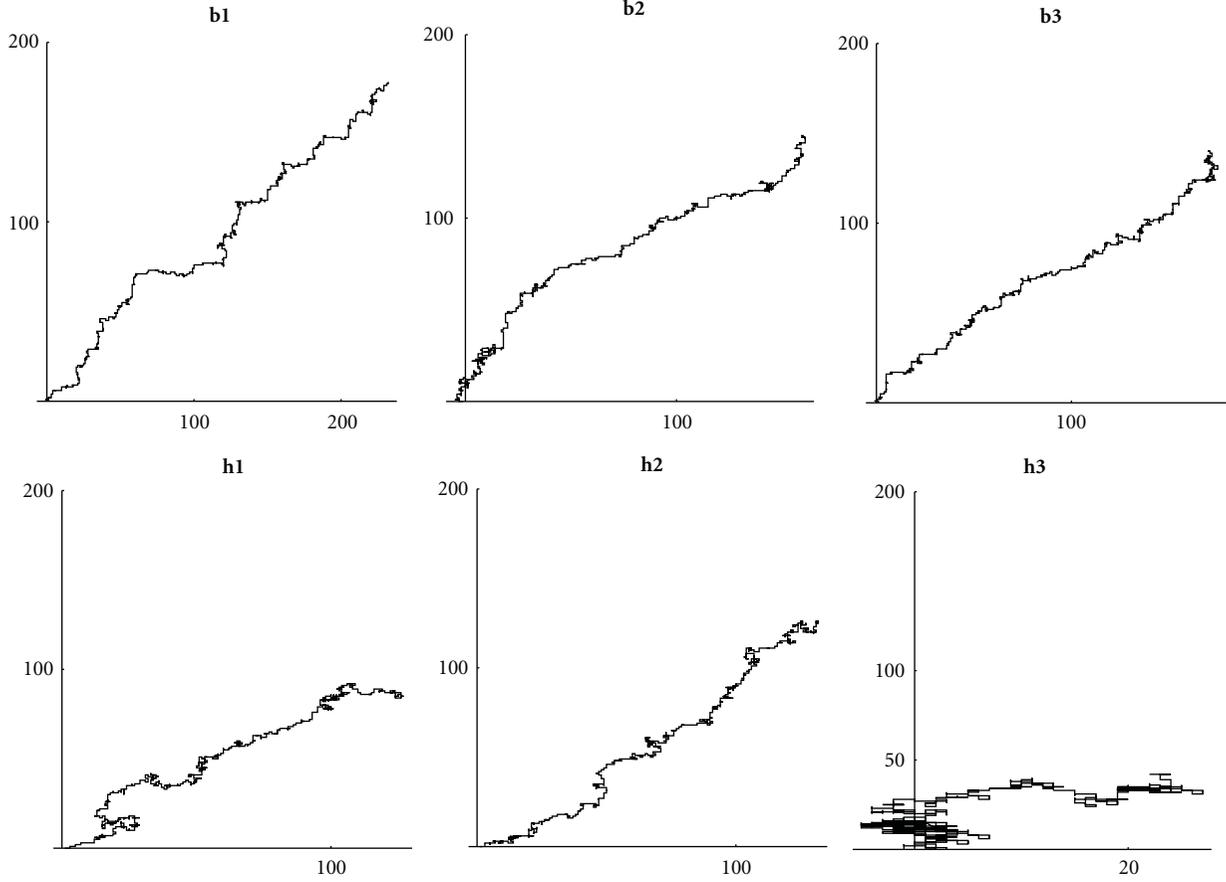


FIGURE 11: Walks on the first 200 nucleotides: (b1) *Mycoplasma putrefaciens*, (b2) *Mortierella verticillata*, (b3) *Blattabacterium*, (h1) *Aeropyrum pernix*, (h2) *Acidianus hospitalis*, and (h3) *Acidilobus saccharovorans*.

It should be noticed that along these spirals, there is a one-to-one map λ between \mathbb{N} and the points of the spiral (with integer coordinates) in \mathbb{R}^2

$$\lambda : \mathbb{N} \mapsto \gamma \subset \mathbb{R} \times \mathbb{R} \quad (26)$$

so that

$$\lambda(n) = (a, b), \quad (n \in \mathbb{N}; (a, b) \in \gamma \subset \mathbb{R} \times \mathbb{R}; a \in \mathbb{Z}, b \in \mathbb{Z}),$$

$$\lambda^{-1}(a, b) = n. \quad (27)$$

This bijective map can be considered also between \mathbb{N} and the complex space \mathbb{C} so that each natural number corresponds to a complex number (with integer coefficients)

$$\lambda(n) = z \stackrel{\text{def}}{=} a + ib, \quad (n \in \mathbb{N}; a, b \in \mathbb{Z}; z \in \mathbb{C}). \quad (28)$$

Since these spirals seem to fill in a finite region of the plane we can evaluate the complexity of each curve by typical fractal measures.

4. Parameters of Complexity

In this section we define some parameters, based on frequency distribution, which can measure the complexity of a DNA by computing the complexity of its representation in the complex plane (for a more detailed analysis see [29] and references therein).

Let \mathcal{S}_N be an N -length-ordered sequence of nucleotides, and

$$p_x(h), \quad x \in \mathcal{A}_1 = \{A, C, G, T\} \quad (29)$$

be the probability to find the nucleotide x at the position h , $1 \leq h \leq N$. According to (12) we define

$$\begin{aligned} a_h &\stackrel{\text{def}}{=} \sum_{j=1}^h u_{Aj}, & c_h &\stackrel{\text{def}}{=} \sum_{j=1}^h u_{Cj}, \\ g_h &\stackrel{\text{def}}{=} \sum_{j=1}^h u_{Gj}, & t_h &\stackrel{\text{def}}{=} \sum_{j=1}^h u_{Tj}, \end{aligned} \quad (1 \leq h \leq N) \quad (30)$$

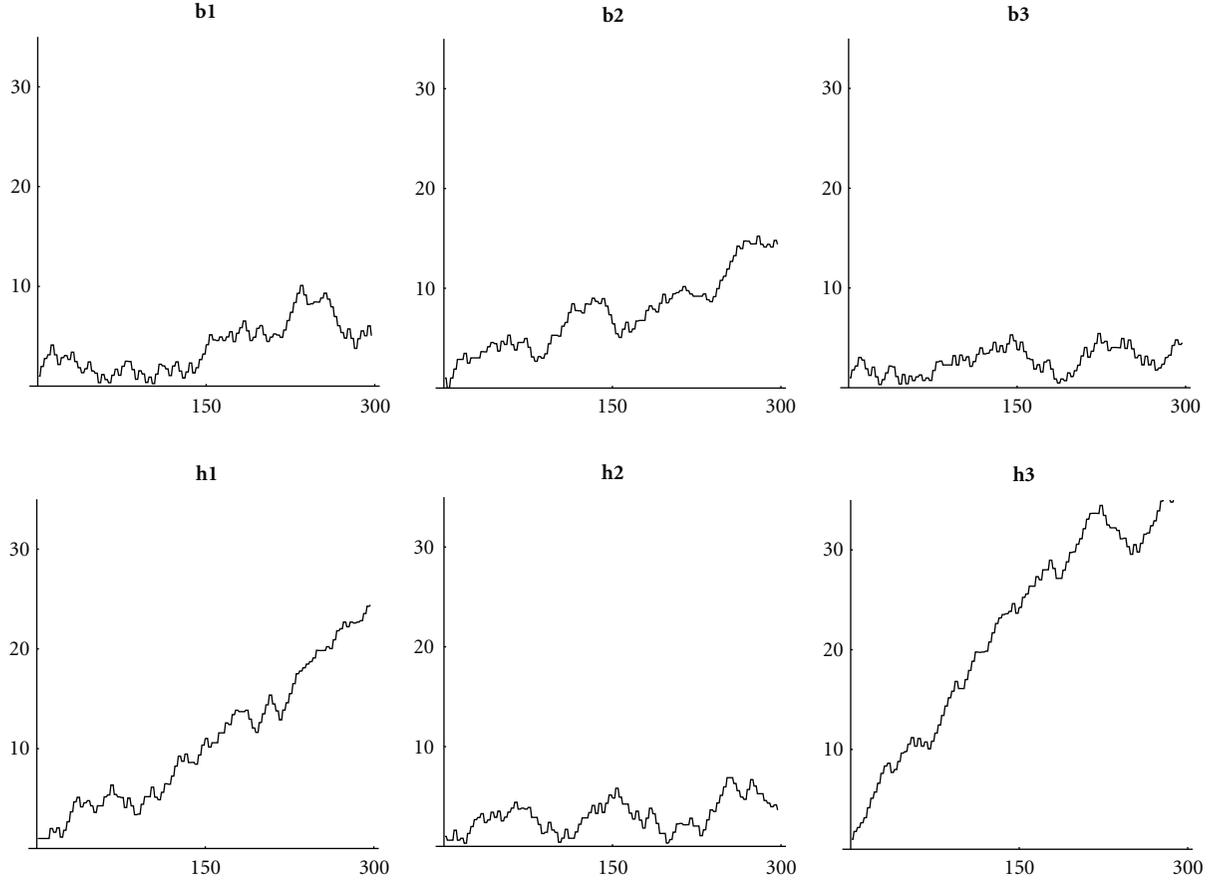


FIGURE 12: Absolute value of walks on the first 100 amino acids: **(b1)** *Mycoplasma putrefaciens*, **(b2)** *Mortierella verticillata*, **(b3)** *Blattabacterium*, **(h1)** *Aeropyrum pernix*, **(h2)** *Acidianus hospitalis*, **(h3)** *Acidilobus saccharovorans*.

as the number of nucleotides in the h -length segment of \mathcal{S}_N , so that

$$a_h + c_h + g_h + t_h = h. \quad (31)$$

The corresponding frequencies are

$$v_x(h) \stackrel{\text{def}}{=} \frac{1}{h} \sum_{j=1}^h u_{xj}, \quad x \in A_1, \quad (1 \leq h \leq N), \quad (32)$$

so that

$$\begin{aligned} v_A(h) &= \frac{a_h}{h}, & v_C(h) &= \frac{c_h}{h}, \\ v_G(h) &= \frac{g_h}{h}, & v_T(h) &= \frac{t_h}{h}. \end{aligned} \quad (33)$$

We can assume that for large sequences

$$p_x(h) \cong v_x(h). \quad (34)$$

4.1. Randomness. Since for a random sequence the frequencies of nucleotides coincide for large n ,

$$v_A(n) \cong v_C(n) \cong v_G(n) \cong v_T(n) \quad (35)$$

TABLE 2: Randomness.

| | |
|----------------------------------|-------|
| <i>Mycoplasma putrefaciens</i> | 0.696 |
| <i>Mortierella verticillata</i> | 0.779 |
| <i>Blattabacterium</i> | 0.743 |
| <i>Aeropyrum pernix</i> | 0.982 |
| <i>Acidianus hospitalis</i> | 0.828 |
| <i>Acidilobus saccharovorans</i> | 0.934 |
| pseudorandom | 0.999 |

we can define as randomness index the following:

$$\mathcal{R} \stackrel{\text{def}}{=} 1 - \sigma(v_A(n), v_C(n), v_G(n), v_T(n)) \quad (36)$$

with σ being the variance, so that $\mathcal{R} = 1$ for random sequence and $\mathcal{R} = 0$ for a nonrandom sequence. Over the first 10000 nucleotides we have the randomness value of Table 2.

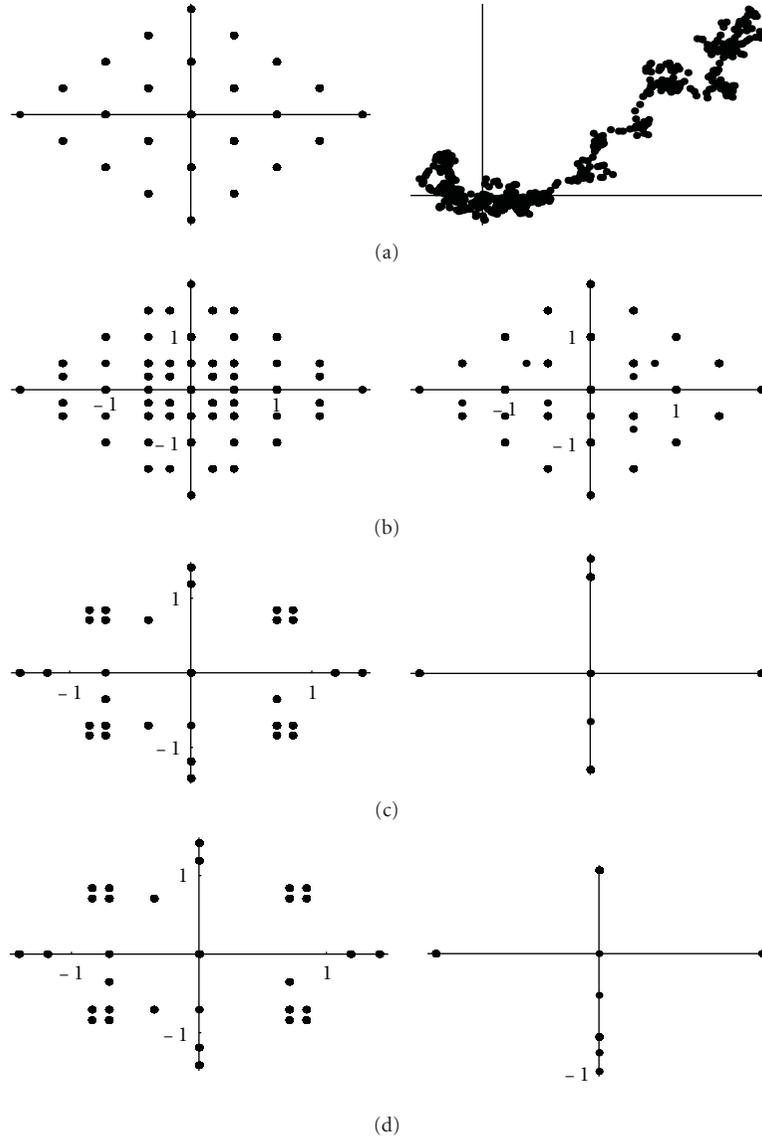


FIGURE 13: Cluster analysis of the 4th short Haar wavelet transform of a 4000-length random sequence (left) and its 2000-length random walk (right): (a) (α, α^*) ; (b) $(\beta_0^0, \beta_0^{*0})$; (c) $(\beta_0^1, \beta_0^{*1})$; (d) $(\beta_1^1, \beta_1^{*1})$.

However, if we compute the randomness index over the frequencies of amino acids in the \mathcal{A}_3 alphabet then we can observe a different distribution of values. Over the first 30000 nucleotides corresponding to 10000 amino acids, we have the randomness value of Table 3.

So that we can comment that the arising complexity of the words and alphabets shows a different randomness in each alphabet.

4.2. Complexity. As a simple measure of complexity [30–32], for an n -length sequence, the following has been proposed [33]:

$$K = \frac{1}{n} \log \frac{n!}{a_n! c_n! g_n! t_n!}. \quad (37)$$

In Table 4 the complexity of the first 100-length segment of the DNA sequences is computed. It is interesting to notice

TABLE 3: Randomness of amino acids distribution.

| | |
|----------------------------------|-------|
| <i>Mycoplasma putrefaciens</i> | 0.946 |
| <i>Mortierella verticillata</i> | 0.938 |
| <i>Blattabacterium</i> | 0.953 |
| <i>Aeropyrum pernix</i> | 0.962 |
| <i>Acidianus hospitalis</i> | 0.916 |
| <i>Acidilobus saccharouorans</i> | 0.950 |
| pseudorandom | 0.963 |

the more similarities between the archaea *Acidilobus* with the pseudorandom sequence than with the pseudoperiodic.

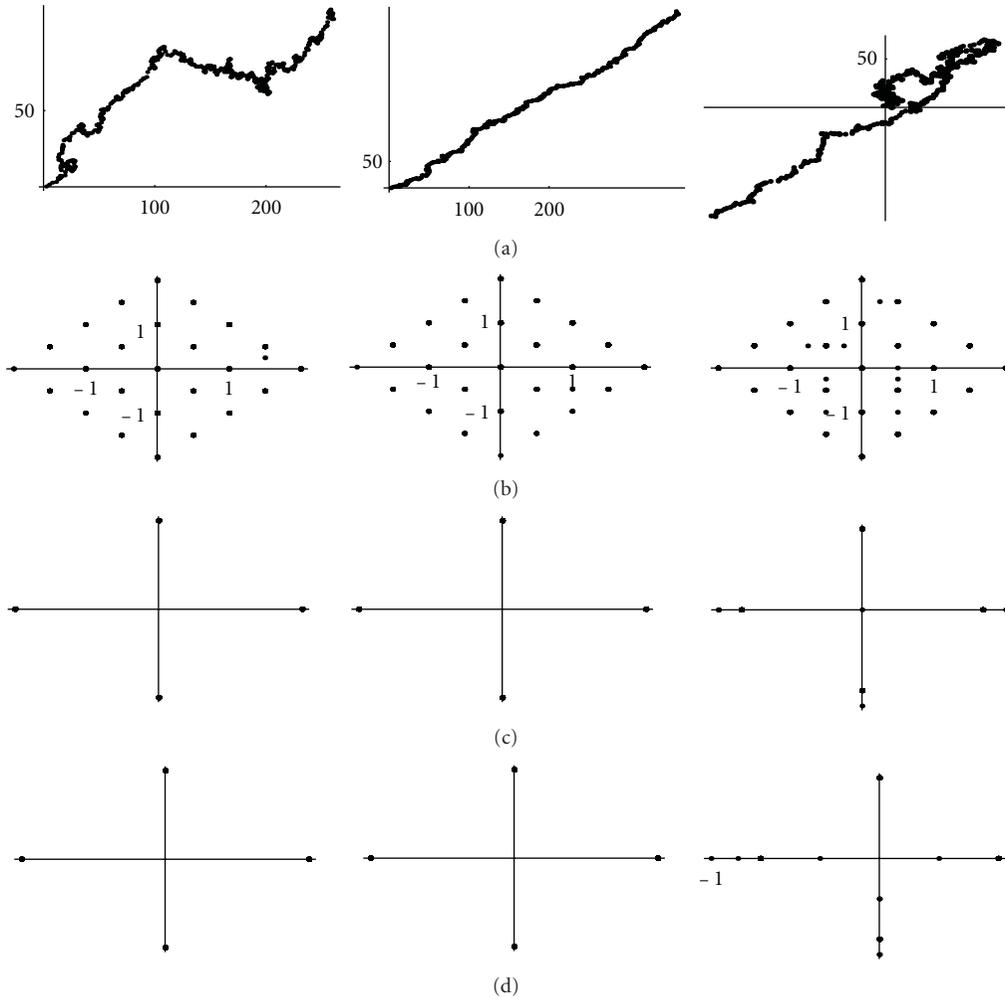


FIGURE 14: Cluster analysis of the 4th short Haar wavelet transform of the complex representation for a DNA walk on the first 2000 nucleotides of (**h1**) *Aeropyrum*, (**h2**) *Acidianus*, (**h3**) *Acidilobus saccharovorans* in the planes: (a) (α, α^*) ; (b) $(\beta_0^0, \beta_0^{*0})$; (c) $(\beta_0^1, \beta_0^{*1})$; (d) $(\beta_1^1, \beta_1^{*1})$.

TABLE 4: Complexity.

| | |
|----------------------------------|-------|
| <i>Mycoplasma putrefaciens</i> | 1.151 |
| <i>Mortierella verticillata</i> | 1.285 |
| <i>Blattabacterium</i> | 1.197 |
| <i>Aeropyrum pernix</i> | 1.212 |
| <i>Acidianus hospitalis</i> | 1.231 |
| <i>Acidilobus saccharouorans</i> | 1.296 |
| Pseudorandom | 1.295 |

Nucleotide distribution in primitive biosequences is more likely random than pseudodeterministic. Moreover, the evolution reduces the complexity of the sequence.

4.3. *Fractal Dimension.* The fractal dimension is computed on the dot plot, by the box counting algorithm [34, 35], as

the average of the number $p(n)$ of 1's in the randomly taken $n \times n$ minors of the $N \times N$ indicator matrix u_{hk} or equivalently the number $p(n)$ of black dots in the randomly taken $n \times n$ squares over the dot plot

$$D = \frac{1}{2N} \sum_{n=2}^N \frac{\log p(n)}{\log n}. \quad (38)$$

The explicit computation enables us to compare the fractal dimension on the first 100-length segments of DNA chains, with an approximation up to 10^{-3} (see Table 5).

If we compare the fractal dimensions of the bacteria with pseudorandom and pseudoperiodic we can see that the fractal dimension of nucleotide distribution ranges, for all variants, in the interval [1.28–1.30]. As expected, the more “random” sequences have higher fractal dimension.

4.4. *Entropy.* Another fundamental parameter, related to the information content of a sequence which measures the heterogeneity of data, is the information entropy (or Shannon entropy) [36–42]. Based on the axiom that less information

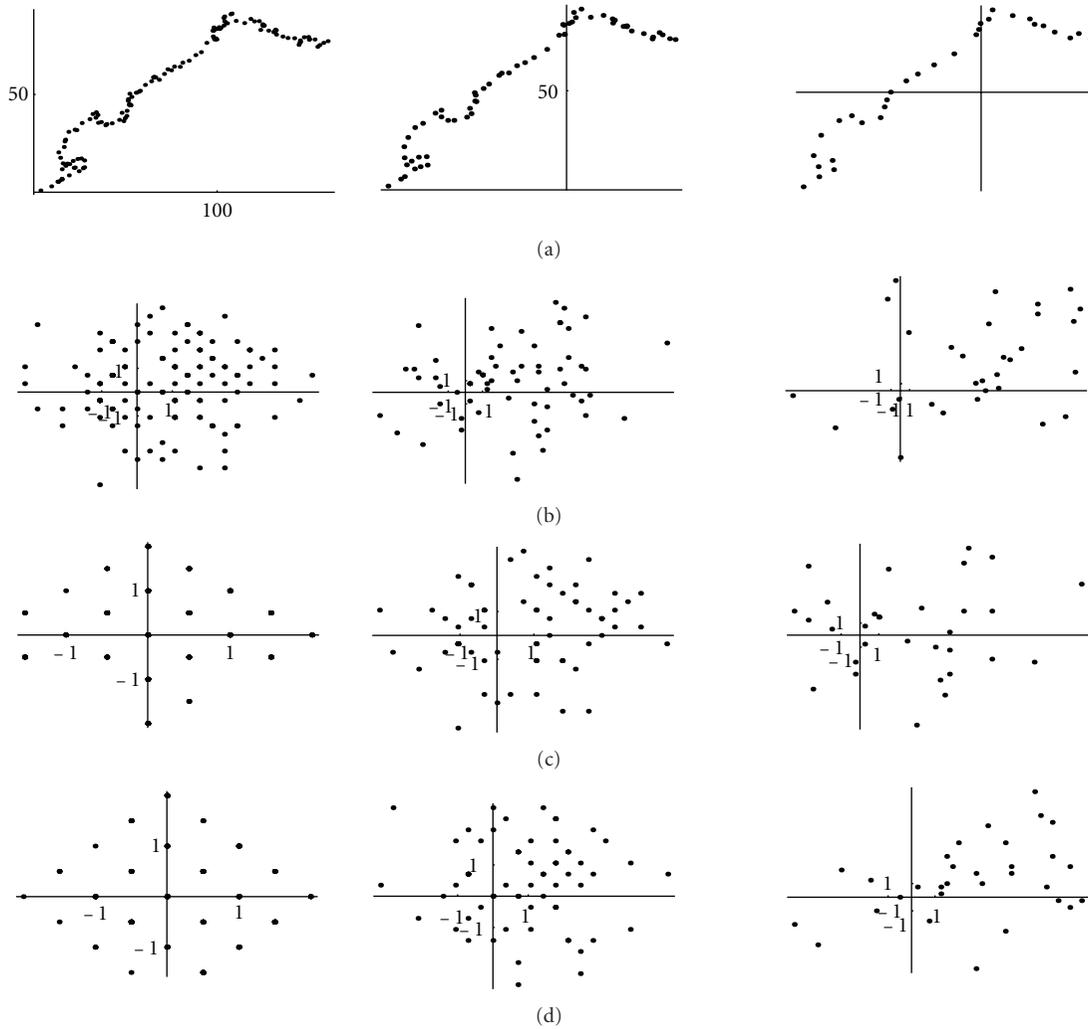


FIGURE 15: Cluster analysis of the 8th (left), 16th (middle column), 32th (right) short Haar wavelet transform of the DNA walk on the first 1000 nucleotides of h1 (*Aeropyrum*) in the planes: (a) (α, α^*) ; (b) (β_0^0, β_0^1) ; (c) (β_0^1, β_0^2) ; (d) (β_1^1, β_1^2) .

TABLE 5: Fractal dimensions.

| | |
|----------------------------------|-------|
| <i>Mycoplasma putrefaciens</i> | 1.283 |
| <i>Mortierella verticillata</i> | 1.296 |
| <i>Blattabacterium</i> | 1.287 |
| <i>Aeropyrum pernix</i> | 1.288 |
| <i>Acidianus hospitalis</i> | 1.290 |
| <i>Acidilobus saccharovorans</i> | 1.297 |
| pseudorandom | 1.298 |
| pseudoperiodic | 1.285 |

implies a larger uncertainty and vice versa that more information leads us to a more deterministic model, the entropy concept has been recently offering some interesting

interpretations about uncertainty in DNA. In fact, DNA as any other signal has been considered as a sequence of symbols carrying chemical-functional information.

The normalized Shannon entropy [39, 40, 42] is defined, over the alphabet \mathcal{A}_ℓ , as

$$H(n) = -\frac{1}{\log \ell} \sum_{x \in \mathcal{A}_\ell} p_x(n) \times \begin{cases} \log p_x(n) & \text{if } p_x(n) \neq 0, \\ 0 & \text{if } p_x(n) = 0, \end{cases} \quad (39)$$

where $p_x(n)$ should be computed for large sequences. According to (32), (34), we will approximate its value with

$$p_x(n) \cong \frac{1}{n} \sum_{i=1}^n u_{xi}, \quad (x \in \mathcal{A}_\ell, 1 \leq n \leq N). \quad (40)$$

However, the entropy is a parameter very similar to the complexity. In fact, it can be easily seen that (for the proof see [29]) the entropy H and the measure of complexity K differ for a factor. There follows that the entropy does not give any

TABLE 6: Shannon entropy.

| | |
|----------------------------------|-------|
| <i>Mycoplasma putrefaciens</i> | 0.877 |
| <i>Mortierella verticillata</i> | 0.976 |
| <i>Blattabacterium</i> | 0.911 |
| <i>Aeropyrum pernix</i> | 0.922 |
| <i>Acidianus hospitalis</i> | 0.937 |
| <i>Acidilobus saccharovorans</i> | 0.984 |
| pseudorandom | 0.984 |

new information comparing with the previous parameters. As expected also the table of entropies classifies bacteria and archaea in the same way (Table 6).

5. Complex Root Representation of DNA Words

The complex (digital) representation of a DNA sequence of words is the map of the symbolic sequence of words into a set of complex numbers and it is defined as

$$\mathcal{D}_\ell(S_N) \xrightarrow{\rho} \mathbb{C} \quad (41)$$

such that for each $x_h \in \mathcal{D}_\ell(S_N)$ it is $\rho(x_h) \in \mathbb{C}$.

The complex root representation of the sequence S_N is the sequence $\mathcal{D}_\ell(S_N)$ of complex numbers $\{y_h\}_{h=1,\dots,N}$ defined as

$$y_h = \rho(x_h) \stackrel{\text{def}}{=} e^{2\pi i(j-1)/|\mathcal{A}_\ell|}, \quad (j = 1, \dots, |\mathcal{A}_\ell|, h = 1, \dots, N) \quad (42)$$

with $i = \sqrt{-1}$ being the imaginary unit. There follows that, independently on the alphabet, it is

$$|y_h| = \left| e^{2\pi i(j-1)/|\mathcal{A}_\ell|} \right| = 1, \quad (\forall \ell; h = 1, \dots, N) \quad (43)$$

being all complex roots, of the unit, located on the unit circle of the complex plane \mathbb{C}^1 .

For instance, with $\mathcal{A}_1 = \{A, C, G, T\}$, the cardinality of the alphabet is $|\mathcal{A}_1| = 4$ and

$$\begin{aligned} \rho(A) &= e^{0/4} = 1, & j = 1, \\ \rho(C) &= e^{\pi i/2} = i, & j = 2, \\ \rho(G) &= e^{\pi i} = -1, & j = 3, \\ \rho(T) &= e^{\pi i 3/2} = -i, & j = 4. \end{aligned} \quad (44)$$

Analogously, with $\mathcal{A}_3 = \{M, E, \dots, W\}$ it is $|\mathcal{A}_3| = 20$ and the 20 complex roots of unit

$$\rho(x_n) = e^{2\pi i(n-1)/20}, \quad (n = 1, \dots, 20; x_n \in \mathcal{A}_3) \quad (45)$$

so that explicitly is

$$\begin{aligned} \rho(M) &= e^{2\pi i 0/20} = 1, & j = 1, \\ \rho(E) &= e^{\pi i/10} = \frac{1}{4} \left[\sqrt{2(5 + \sqrt{5})} + i(\sqrt{5} - 1) \right], & j = 2, \\ \rho(Q) &= e^{\pi i/5} = \frac{1}{4} \left[1 + \sqrt{5} + i\sqrt{2(5 - \sqrt{5})} \right], & j = 3, \\ &\vdots & \vdots \\ \rho(W) &= e^{\pi i 19/10} = \frac{1}{4} \left[\sqrt{2(5 + \sqrt{5})} - i(\sqrt{5} - 1) \right], & j = 20. \end{aligned} \quad (46)$$

Therefore the complex representation of a DNA sequence is a sequence of complex numbers

$$y_h = \xi_h + \eta_h i, \quad \xi_h = \Re(y_h), \quad \eta_h = \Im(y_h) \quad (47)$$

with y_h given by (42).

An n -length pseudorandom (white noise) complex sequence belonging to the unit circle can be defined directly by using some random exponents

$$R_n \stackrel{\text{def}}{=} (-1)^{r_n} i^{s_n}, \quad |R_n| = 1, \quad (48)$$

with r_n, s_n being random values in the set $\{0, \mathbb{N}\}$.

5.1. Random Walks. Random walk on the complex sequence \mathbf{Y}_N is defined as the series $\mathbf{Z}_N = \{z_n\}_{n=1,\dots,N}$

$$z_n \stackrel{\text{def}}{=} \sum_{k=1,\dots,n} y_k, \quad n = 1, \dots, N \quad (49)$$

which is the cumulative sum

$$\left\{ y_1, y_1 + y_2, \dots, \sum_{s=1}^n y_s, \dots, \sum_{s=1}^N y_s \right\}. \quad (50)$$

When $y_k = \rho(x_k)$ with $x_k \in \mathcal{A}_\ell$ and $\mathbf{X}_k \in S_N$ we will properly call these walks as DNA walk. When the y_k are randomly generated we will call them random walks.

By remembering the definition of frequencies, DNA walk is the complex value signal $\{Z_n\}_{n=0,\dots,N-1}$ with

$$z_n = (\Re[z_n], \Im[z_n]) = (a_n - g_n) + (t_n - c_n)i, \quad z_n \in \mathbb{C}_1, \quad (51)$$

where the coefficients a_n, g_n, t_n, c_n given by (12) fulfill the condition (31).

If we compare the DNA walks (Figure 11) some primitive archaea such as h3 are very similar to a random walk (Figure 13). In particular archaea seem to grow less than other bacteria (with the exception of b2).

It is interesting also to notice that the random walks on amino acids (Figure 12) show that more evolved organisms have some ‘‘periodic’’ behavior, while the absolute value of walks on archaea is growing fast.

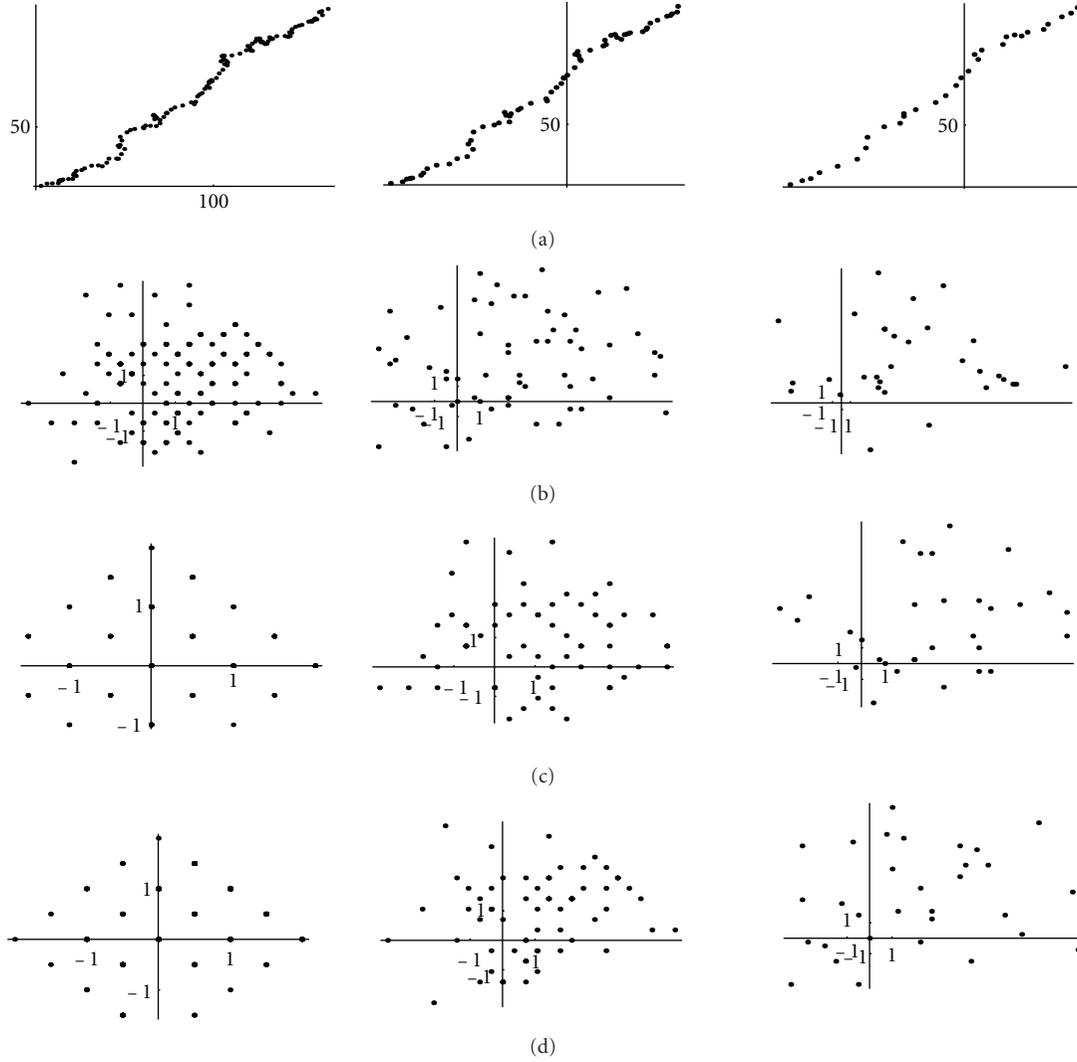


FIGURE 16: Cluster analysis of the 8th (left), 16th (middle column), 32th (right) short Haar wavelet transform of the DNA walk on the first 1000 nucleotides of h2 (*Acidianus*) in the planes: (a) (α, α^*) ; (b) (β_0^0, β_0^*0) ; (c) (β_0^1, β_0^*1) ; (d) (β_1^1, β_1^*1) .

6. Wavelet Analysis

Wavelet analysis is a powerful method extensively applied to the analysis of biological signals [12, 43–45] aiming to single out the most significant parameters of complexity and heterogeneity in a time series and, in particular, in a DNA sequence. This method is based on the analysis of wavelet coefficients which are obtained by the wavelet transform.

We will consider in the following the Haar wavelet basis (see, e.g., [3, 4, 29]) made by scaling functions:

$$\varphi_k^n(x) \stackrel{\text{def}}{=} 2^{n/2} \varphi(2^n x - k), \quad (0 \leq n, 0 \leq k \leq 2^n - 1),$$

$$\varphi(2^n x - k) = \begin{cases} 1, & x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right), \\ 0, & x \notin \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right), \end{cases} \quad (52)$$

and the *Haar wavelets*:

$$\psi_k^n(x) \stackrel{\text{def}}{=} 2^{n/2} \psi(2^n x - k), \quad \|\psi_k^n(x)\|_{L^2} = 1,$$

$$\psi(2^n x - k) \stackrel{\text{def}}{=} \begin{cases} -1, & x \in \left[\frac{k}{2^n}, \frac{k+1/2}{2^n} \right), \\ 1, & x \in \left[\frac{k+1/2}{2^n}, \frac{k+1}{2^n} \right), \\ 0, & \text{elsewhere.} \end{cases} \quad (53)$$

The *discrete Haar wavelet transform* is the $N \times N$ matrix $\mathcal{W}^N : \mathbb{K}^N \subset \ell^2 \rightarrow \mathbb{K}^N \subset \ell^2$ which maps the vector

$$\mathbf{Y} \equiv \{Y_i\}, \quad (i = 0, \dots, 2^M - 1, 2^M = N < \infty, M \in \mathbb{N}) \quad (54)$$

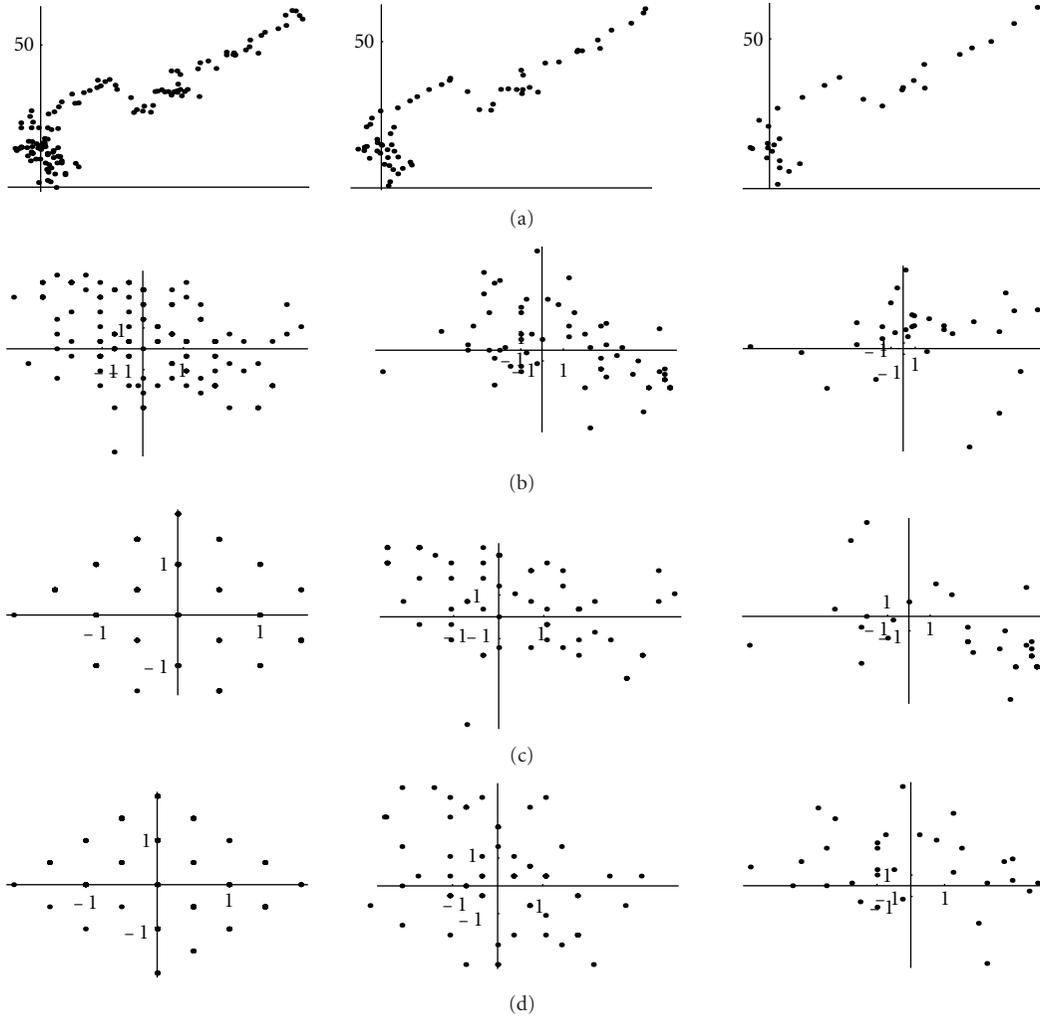


FIGURE 17: Cluster analysis of the 8th (left), 16th (middle column), 32th (right) short Haar wavelet transform of the DNA walk on the first 1000 nucleotides of h3 (*Acidilobus saccharovorans*) in the planes: (a) (α, α^*) ; (b) (β_0^0, β_0^*) ; (c) (β_0^1, β_0^*) ; (d) (β_1^1, β_1^*) .

into the vector of *wavelet coefficients* $\mathbf{N} = \{\alpha, \beta_k^n\}$:

$$\begin{aligned} \mathcal{W}_N \mathbf{Y} &= \mathbf{N}, \\ \mathbf{N} &\stackrel{\text{def}}{=} \{\alpha, \beta_0^0, \dots, \beta_{2^M-1}^{M-1}\}, \\ \mathbf{Y} &\stackrel{\text{def}}{=} \{Y_0, Y_1, \dots, Y_{N-1}\}, \quad (2^M = N). \end{aligned} \quad (55) \quad \text{and [1-3, 14]}$$

The matrix \mathcal{W}_N can be easily computed by some recursive product [3, 4, 13, 29, 46] so that with $N = 4$, $M = 2$, we have [3, 4, 29]

$$\mathcal{W}_4 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}. \quad (56)$$

From (55) with $M = 2$, $N = 4$, by explicit computation, we have

$$\alpha = \frac{1}{4}(Y_0 + Y_1 + Y_2 + Y_3) \quad (57)$$

$$\beta_0^0 = \frac{1}{2}(Y_2 - Y_0 + Y_3 - Y_1),$$

$$\beta_0^1 = \frac{1}{\sqrt{2}}(Y_0 - Y_1),$$

$$\beta_1^1 = \frac{1}{\sqrt{2}}(Y_3 - Y_2). \quad (58)$$

Thus the first wavelet coefficient α represents the average value of the sequence and the other coefficients β the finite differences. The wavelet coefficients β 's, also called details coefficients, are strictly connected with the first-order properties of the discrete time series.

In the following we will consider the short wavelet transform which consists in the subdivision of the DNA sequence

into 4-length segments and apply the wavelet transform to each segment. As a result, from the $N = 2^M$ -length complex vector \mathbf{Y} , which is subdivided into 2^{M-2} segments, the 4-parameter short Haar wavelet transform gives the cluster of points

$$(\mathcal{W}^p \Re(\mathbf{Y}^s), \mathcal{W}^p \Im(\mathbf{Y}^s)), \quad s = 0, \dots, \sigma = \frac{N}{p}, \quad p = 4 \quad (59)$$

in the 8-dimensional space $\mathbb{R}^4 \times \mathbb{R}^4$, that is,

$$(\alpha, \alpha^*), (\beta_0^0, \beta_{*0}^0), \dots, (\beta_{2^{p-1}-1}^{p-1}, \beta_{*2^{p-1}-1}^{p-1}), \quad p = 4. \quad (60)$$

This algorithm enables us to construct clusters of wavelet coefficients and to study the correlation between the real and imaginary coefficients of the DNA representation and DNA walk. It has been observed [3, 4, 29] that some symmetry arises from the plots of wavelet coefficients of DNA walks.

6.1. Cluster Analysis of the Wavelet Coefficients of the Complex DNA Representation. Let us first compute the clusters of wavelet coefficients for the random sequence (48). As can be seen the wavelet coefficients both for the sequence and for its series range in some discrete set of values (see Figure 13).

The cluster algorithm applied to the complex representation sequence shows that the values of the wavelet coefficients belong to some discrete finite sets (Figure 14).

It should be noticed that this symmetry on detail coefficients is lost for wavelet transform on longer segments (Figures 15, 16 and 17).

There follows that DNA sequences have to be considered as Markov chain with short range dependence; in other words any acid nucleic is attached to the chain on the base of a correlation of the previous acid nucleic. In other words, if we look for a dependence rule on the DNA nucleotides this dependence might be summarized by a function as

$$x_{n+1} = f(x_n), \quad (n = 1, \dots, N). \quad (61)$$

7. Conclusions

In this paper archaea DNAs have been studied by focussing on the main parameters for complexity. It has been shown that more or less the main indices for complexity and heterogeneity, such as entropy, fractal dimension, and complexity do not differ too much when we have to classify the complexity of the sequence. However, some DNA sequences look more close to random sequences than others, thus suggesting that the evolution involves a process of complexity reduction: the more evolved a sequence is, the more far from a random distribution it is. In any case seems to be apparently impossible to distinguish between a random sequence and a DNA chain. By using the short wavelet transform instead we have shown that on short range (4-nucleotides) a DNA sequence shows some symmetries that slowly disappear by increasing the length of the analysed segment. Moreover, more evolved organisms have a more symmetrical distribution of wavelet coefficients.

References

- [1] C. Cattani, "Complex representation of DNA sequences," in *Proceedings of the Bioinformatics Research and Development Second International Conference*, M. Elloumi et al., Ed., Springer, Vienna, Austria, July 2008.
- [2] C. Cattani, "Complex representation of DNA sequences," *Communications in Computer and Information Science*, vol. 13, pp. 528–537, 2008.
- [3] C. Cattani, "Wavelet Algorithms for DNA Analysis," in *Algorithms in Computational Molecular Biology: Techniques, Approaches and Applications*, M. Elloumi and A. Y. Zomaya, Eds., Wiley Series in Bioinformatics, chapter 35, pp. 799–842, John Wiley & Sons, New York, NY, USA, 2010.
- [4] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, pp. 1–31, 2010.
- [5] C. Cattani and G. Pierro, "Complexity on acute myeloid leukemia mRNA transcript variant," *Mathematical Problems in Engineering*, vol. 2011, pp. 1–16, 2011.
- [6] C. Cattani, G. Pierro, and G. Altieri, "Entropy and multi-fractality for the myeloma multiple TET 2 gene," *Mathematical Problems in Engineering*, vol. 2011, pp. 1–17, 2011.
- [7] C. Cattani and J. J. Rushchitsky, *Wavelet and Wave Analysis as applied to Materials with Micro or Nanostructure, Series on Advances in Mathematics for Applied Sciences*, vol. 74, World Scientific, Singapore, 2007.
- [8] K. B. Murray, D. Gorse, and J. M. Thornton, "Wavelet transforms for the characterization and detection of repeating motifs," *Journal of Molecular Biology*, vol. 316, no. 2, pp. 341–363, 2002.
- [9] A. A. Tsonis, P. Kumar, J. B. Elsner, and P. A. Tsonis, "Wavelet analysis of DNA sequences," *Physical Review E*, vol. 53, no. 2, pp. 1828–1834, 1996.
- [10] M. Altaiski, O. Mornev, and R. Polozov, "Wavelet analysis of DNA sequences," *Genetic Analysis—Biomolecular Engineering*, vol. 12, no. 5-6, pp. 165–168, 1996.
- [11] A. Arneodo, Y. D'Aubenton-Carafa, E. Bacry, P. V. Graves, J. F. Muzy, and C. Thermes, "Wavelet based fractal analysis of DNA sequences," *Physica D*, vol. 96, no. 1-4, pp. 291–320, 1996.
- [12] M. Zhang, "Exploratory analysis of long genomic DNA sequences using the wavelet transform: examples using polyomavirus genomes," in *Proceedings of the 6th Genome Sequencing and Analysis Conference*, pp. 72–85, 1995.
- [13] C. Cattani, "Haar wavelet-based technique for sharp jumps classification," *Mathematical and Computer Modelling*, vol. 39, no. 2-3, pp. 255–278, 2004.
- [14] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.
- [15] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, pp. 1–26, 2010.
- [16] M. Li and J. Y. Li, "On the predictability of long-range dependent series," *Mathematical Problems in Engineering*, vol. 2010, Article ID 397454, pp. 1–9, 2010.
- [17] M. Li and S. C. Lim, "Power spectrum of generalized Cauchy process," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 219–222, 2010.
- [18] National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/genbank>.
- [19] Genome Browser, <http://genome.ucsc.edu>.
- [20] European Informatics Institute, <http://www.ebi.ac.uk>.
- [21] Ensembl, <http://www.ensembl.org>.

- [22] J. L. Howland, *The Surprising Archaea*, Oxford University Press, New York, NY, USA, 2000.
- [23] C. R. Woese and G. E. Fox, "Phylogenetic structure of the prokaryotic domain: the primary kingdoms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 11, pp. 5088–5090, 1977.
- [24] M. T. Madigan and B. L. Marrs, "Extremophiles," *Scientific American*, vol. 276, no. 4, pp. 82–87, 1997.
- [25] 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.
- [26] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters*, vol. 5, pp. 973–977, 1987.
- [27] J. Szczepański and T. Michałek, "Random fields approach to the study of DNA chains," *Journal of Biological Physics*, vol. 29, no. 1, pp. 39–54, 2003.
- [28] M. Stein and S. M. Ulam, "An observation on the distribution of primes," *American Mathematical Monthly*, vol. 74, no. 1, p. 4344, 1967.
- [29] C. Cattani, "Complexity and Simmetries in DNA sequences," in *Handbook of Biological Discovery*, M. Elloumi and A. Y. Zomaya, Eds., Wiley Series in Bioinformatics, chapter 22, pp. 700–742, John Wiley & Sons, New York, NY, USA, 2012.
- [30] M. A. Gates, "Simpler DNA sequence representations," *Nature*, vol. 316, no. 6025, p. 219, 1985.
- [31] M. A. Gates, "A simple way to look at DNA," *Journal of Theoretical Biology*, vol. 119, no. 3, pp. 319–328, 1986.
- [32] E. Hamori and J. Ruskin, "H curves, a novel method of representation of nucleotide series especially suited for long DNA sequences," *Journal of Biological Chemistry*, vol. 258, no. 2, pp. 1318–1327, 1983.
- [33] J. A. Berger, S. K. Mitra, M. Carli, and A. Neri, "Visualization and analysis of DNA sequences using DNA walks," *Journal of the Franklin Institute*, vol. 341, no. 1-2, pp. 37–53, 2004.
- [34] P. Bernaola-Galván, R. Román-Roldán, and J. L. Oliver, "Compositional segmentation and long-range fractal correlations in DNA sequences," *Physical Review E*, vol. 55, no. 5, pp. 5181–5189, 1996.
- [35] C. L. Berthelsen, J. A. Glazier, and M. H. Skolnick, "Global fractal dimension of human DNA sequences treated as pseudorandom walks," *Physical Review A*, vol. 45, no. 12, pp. 8902–8913, 1992.
- [36] P. R. Aldrich, R. K. Horsley, and S. M. Turcic, "Symmetry in the language of gene expression: a survey of gene promoter networks in multiple bacterial species and non- σ regulons," *Symmetry*, vol. 3, pp. 1–20, 2011.
- [37] R. Ferrer-I-Cancho and N. Forns, "The self-organization of genomes," *Complexity*, vol. 15, no. 5, pp. 34–36, 2010.
- [38] T. Misteli, "Self-organization in the genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 17, pp. 6885–6886, 2009.
- [39] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [40] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 623–656, 1948.
- [41] R. V. Solé, "Genome size, self-organization and DNA's dark matter," *Complexity*, vol. 16, no. 1, pp. 20–23, 2010.
- [42] R. M. Yulmetyev, N. A. Emelyanova, and F. M. Gafarov, "Dynamical Shannon entropy and information Tsallis entropy in complex systems," *Physica A*, vol. 341, no. 1–4, pp. 649–676, 2004.
- [43] A. Arneodo, E. Bacry, P. V. Graves, and J. F. Muzy, "Characterizing long-range correlations in DNA sequences from wavelet analysis," *Physical Review Letters*, vol. 74, no. 16, pp. 3293–3296, 1995.
- [44] A. Arneodo, Y. D'Aubenton-Carafa, B. Audit, E. Bacry, J. F. Muzy, and C. Thermes, "What can we learn with wavelets about DNA sequences?" *Physica A*, vol. 249, no. 1–4, pp. 439–448, 1998.
- [45] W. Li, "The study of correlation structures of DNA sequences: a critical review," *Computers and Chemistry*, vol. 21, no. 4, pp. 257–271, 1997.
- [46] C. Cattani, "Haar wavelets based technique in evolution problems," *Proceedings of the Estonian Academy of Sciences: Physics & Mathematics*, vol. 53, no. 1, pp. 45–63, 2004.

Research Article

The Stochastic Evolution of a Protocell: The Gillespie Algorithm in a Dynamically Varying Volume

T. Carletti¹ and A. Filisetti²

¹Namur Center for Complex Systems and University of Namur, 5000 Namur, Belgium

²CIRI, Energy and Environment Interdepartmental Center for Industrial Research, University of Bologna and European Centre for Living Technology, Cà Foscari University of Venice, 30124 Venice, Italy

Correspondence should be addressed to T. Carletti, timoteo.carletti@fundp.ac.be

Received 30 September 2011; Accepted 28 November 2011

Academic Editor: Carlo Cattani

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

We propose an improvement of the Gillespie algorithm allowing us to study the time evolution of an ensemble of chemical reactions occurring in a varying volume, whose growth is directly related to the amount of some specific molecules, belonging to the reactions set. This allows us to study the stochastic evolution of a protocell, whose volume increases because of the production of container molecules. Several protocell models are considered and compared with the deterministic models.

1. Introduction

All known life forms are composed of basic units called *cells*; this holds true from the single-cell prokaryote bacterium to the highly sophisticated eukaryotes, whose existence is the result of the coordination, in term of self-organization and emergence, of the behavior of each single basic unit.

While present-day cells are endowed with highly sophisticated regulatory mechanisms, which represent the outcome of almost four billion-years of evolution, it is generally believed that the first life forms were much simpler. Such primordial life-bricks, the *protocells*, were most probably exhibiting only few simplified functionalities, that required a primitive embodiment structure, a protometabolism, and rudimentary genetics, so to guarantee that offsprings were “similar” to their parents [1–3].

Intense research programs are being established aiming at obtaining protocells capable of growth and duplication, endowed with some limited form of genetics [3–6]. Despite all efforts, artificial protocells have not yet been reproduced in laboratory and it is thus extremely important to develop reference models [6–9] that capture the essence of the first protocells appeared on Earth and enable to monitor their subsequent evolution. Due to the uncertainties about the

details, high-level abstract models are particularly relevant. Quoting Kaneko [10], it is necessary to “consider simplified models able to capture universal behaviors, without carefully adding complicating details.”

Most of the models present in the literature are based on deterministic differential equations governing the evolution of the concentrations of the involved reacting molecules. Even if the results are worth discussing and provide important insights, it should be stressed that the former assumptions are rarely satisfied in a cell [11]. Firstly, the number of involved molecules is small and should be counted by integer numbers, hence the use of the concentrations can be questioned; secondly, the presence of the thermal noise introduces in the system a degree of stochasticity than cannot be trivially encoded by a differential equation, mostly because this makes the time evolution a stochastic process. One possible way to overcome such difficulties is to use the Chemical Master equation: given the present state of the system, namely, the number of available molecules for each species, and the possible reactions among them, one can compute the transition probabilities to reach and leave the given state and thus get a partial differential equation describing the time evolution of the probability distribution of having a given number of molecules at any future times [11, 12].

Analytically solving the resulting equation is normally a very hard task, one should thus resort to use numerical methods. A particularly suitable one is the algorithm presented by Gillespie [11, 12], allowing to determine, as a function of the present state of the system, the most probable reaction and the most probable reaction time, that is, the time at which such reaction will occur.

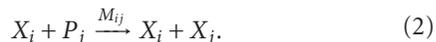
Let us however observe that in the setting we are hereby interested in, the chemical reactions occur in a varying volume, because of the protocell growth; we thus need to adapt the Gillespie method to account for this factor. To the best of our knowledge, there are in the literature very few papers dealing with the Gillespie algorithm in a varying volume [13, 14]. Moreover in all these papers, the volume variation can be considered as an exogenous factor, not being directly related to the number of lipids forming the protocell membrane. So our main contribution is to improve the Gillespie algorithm taking into account the protocell varying volume which is moreover consistent with the increase of the number of lipids constituting the protocell membrane.

The paper is organized as follow. In Section 2 we briefly recall the Surface Reaction Models of protocell that would be used to compare our stochastic numerical scheme. Then in Sections 3 and 4 we will present our implementation of the Gillespie algorithm in a dynamically varying volume. Finally in Section 5 we will present some applications of our method.

2. Surface Reaction Models

Among the available models for protocells, a particularly interesting one is the Surface Reactions Model [7, 9], SRM for short, and its applicability to the synchronization problem. Such model is roughly inspired by the Los Alamos bug hypothesis [2, 6] but which, due to its abstraction level, the SRM can be applied to a wider set of protocell hypotheses.

The SRM is built on the assumption that a protocell should comprise at least one kind of “container” molecule (typically a lipid or amphiphile), hereby called C molecule, and one kind of replicator molecule—loosely speaking “genetic material,” hereafter called Genetic Memory Molecule, GMM for short, and named with the letter X . There are therefore two kinds of reactions which are crucial for the working of the protocell: those which synthesize the container molecules and those which synthesize the GMM replicators, respectively,



In both cases L_i and P_j are the buffered precursors, respectively, of container molecules and of the j th GMM, while α_i and M_{ij} are the reaction kinetic constants.

A second main assumption of the SRM, is that such reactions occur on the *surface* of the protocell, exposed to the external medium where precursors are free to move. Hence, as long as container molecules are produced, they are incorporated in the membrane that thus increases its size,

until a critical point at which, due to physical instabilities, the membrane splits and two offsprings are obtained, each one getting half of the mother’s GMMs and whose size is roughly half that of the mother just before the division.

Under the previous assumptions and in the deterministic setting, one can prove [7, 9] that the number of membrane molecules and the number of GMMs evolve in time according to

$$\begin{aligned} \dot{C} &= \left(\frac{C}{\rho}\right)^{\beta-1} \vec{\alpha} \cdot \vec{X}, \\ \dot{X} &= C^{\beta-1} M \cdot \vec{X}, \end{aligned} \quad (3)$$

where $\vec{X} = (X_1, \dots, X_N)$ represents the amount of each GMM, $\vec{\alpha} = (\alpha_1, \dots, \alpha_N)$ is the vector of the reaction constants responsible for the production of C molecules from the X molecules plus some appropriate precursor. (M_{ij}) denotes the reaction constant at which X_i is produced by X_j plus some precursor. $\beta \in [2/3, 1]$ is a geometrical shape factor that relates the surface to the volume of the protocell and ρ is the lipid density (for more details the interested reader can consult [7, 9]). Let us observe that in this setting the precursors are assumed to be buffered and thus their amount to be constant, hence the latter can be incorporated into the constants α and M .

So starting with an initial value of container molecules, $C(t_0) = C_0$, and of GMMs, $\vec{X}(t_0) = \vec{X}_0$, the protocell will grow until some time $t_0 + \Delta T_1$ at which the amount of C molecules has doubled with respect to the initial value, $C(t_0 + \Delta T_1) = 2C_0$ and thus the protocell undergoes a division. Each offspring will get half of the GMMs the mother protocell had just before the division, $\vec{X}^{(1)} = \vec{X}(t_0 + \Delta T_1)/2$. And the protocell cycle starts once again. One can prove [7, 9] that under suitable conditions $\vec{X}^{(n)}$ tends to a constant value once n goes to infinity, implying thus the emergence of synchronization of growth and information production.

3. The Method

Let us now improve the previous scheme by introducing a probabilistic setting *à la Gillespie*. We thus consider a protocell made by a lipidic vesicle and containing a well-stirred mixture of N GMMs, X_1, \dots, X_N , that may react through m elementary reaction channels R_μ , $\mu = 1, \dots, m$, running within the volume $V(t)$ of the protocell.

Let us observe that because of the protocell growth the volume is an increasing function of time. Actually one can relate the volume to the amount of container molecules via their density $V = C/\rho$ where C denotes the integer number of molecules forming the lipidic membrane. We will hereby use the same symbol X_i to denote both the i th GMM and the integer number of molecules of type X_i in the system.

For each reaction channel R_μ assume that there exists a scalar rate c_μ such that $c_\mu dt + o(dt)$ is the probability that a random combination of molecules from channel R_μ will react in the interval $[t, t + dt)$ within the volume $V(t)$.

Let $h_\mu(Y)$ be the total number of possible distinct combinations of molecules for a channel R_μ when the system is in

state $Y = (X_1, \dots, X_N, C)$, then we can define the *propensity* [14] of the reaction R_μ to be $a_\mu(Y) = h_\mu(Y)c_\mu$.

One can prove [11] that for a binary reaction the rate c_μ can be written in the form $c_\mu = k_\mu/V$, where k_μ is a fixed constant. Similarly one can prove that for a reaction involving n different species, we get $c_\mu = k_\mu/V^{n-1}$. And thus for a single molecule reaction, that is, a decay, we get $c_\mu = k_\mu$, namely, independently from the volume.

Let us now assume that among the m reactions, Q_1 involve one single molecule, Q_2 are binary reactions, Q_3 are ternary reactions, and so on. Of course $Q_1 + Q_2 + \dots + Q_{N+1} = m$. We recall that we have N GMMs and the container-type molecule C , hence $N + 1$ species. For short we will denote \mathcal{Q}_1 the set of indices μ for monomolecule reactions, and by \mathcal{Q} the remaining ones. Let us observe that in this way some coefficient a_μ , will depend both on the system state Y and on the time via the volume $V(t)$: $a_\mu(Y, t)$ for $\mu \in \mathcal{Q}$.

More precisely to study the time evolution of the system we need to determine the probability $P_\mu(\tau | Y, t)d\tau$, that given the system in the state $Y = (X_1, \dots, X_n, C)$ at time t , then the next reaction will occur in the infinitesimal time interval $(t + \tau, t + \tau + d\tau)$ and it will be the reaction R_μ . This probability will be computed as

$$P_\mu(\tau | Y, t)d\tau = P_{\text{not}}(\tau | Y, t) \times a_\mu(Y, t + \tau)d\tau, \quad (4)$$

where $P_{\text{not}}(\tau | Y, t)$ is the probability that no reaction occurs in $(t, t + \tau)$ given the state Y at time t whereas the rightmost term denotes the probability to have a reaction R_μ in $(t + \tau, t + \tau + d\tau)$ given the state Y at time $t + \tau$.

To compute the first term P_{not} , let us take $s \in [t, t + \tau]$ and observe that:

$$\begin{aligned} P_{\text{not}}(s + ds | Y, t) &= P_{\text{not}}(s | Y, t)P_{\text{not}}(ds | Y, t + s) \\ &= P_{\text{not}}(s | Y, t) \left(1 - \sum_{\mu} a_\mu(Y, t + s)ds \right), \end{aligned} \quad (5)$$

being $1 - \sum_{\mu} a_\mu(Y, t + s)ds$ the probability that no reaction will occur in $(t + s, t + s + ds)$ once we are in state Y at time $t + s$. Thus rewriting the previous difference equation as a differential equation, passing to the limit $ds \rightarrow 0$, and observing that $P_{\text{not}}(0 | Y, t) = 1$, we get the solution:

$$P_{\text{not}}(\tau | Y, t) = \exp \left[-A_{Q_1}(Y)\tau - \int_0^\tau A_Q(Y, s + t)ds \right], \quad (6)$$

where

$$A_{Q_1}(Y) = \sum_{\mu \in \mathcal{Q}_1} a_\mu(Y), \quad A_Q(Y, s + t) = \sum_{\mu \in \mathcal{Q}} a_\mu(Y, s + t). \quad (7)$$

The apparent asymmetry in the exponential term in (6) is easily recovered by observing that $A_{Q_1}(Y)\tau = \int_0^\tau A_{Q_1}(Y)ds$.

We can thus conclude that

$$\begin{aligned} P_\mu(\tau | Y, t)d\tau &= \exp \left[-A_{Q_1}(Y)\tau - \int_t^{t+\tau} A_Q(Y, s)ds \right] a_\mu(Y, t + \tau)d\tau. \end{aligned} \quad (8)$$

Let us observe that the rightmost term is correctly $a_\mu(Y, t + \tau)$, namely the system is still in the state Y at time $t + \tau$, because no reaction has been produced in $(t, t + \tau)$.

Let us recall that the volume enters in the previous relation via the function A_Q , more explicitly one has

$$\begin{aligned} A_Q(Y, s) &= \sum_{\mu \in \mathcal{Q}_2} \frac{h_\mu(Y)k_\mu}{V(s)} + \sum_{\mu \in \mathcal{Q}_3} \frac{h_\mu(Y)k_\mu}{(V(s))^2} \\ &+ \dots + \sum_{\mu \in \mathcal{Q}_{N+1}} \frac{h_\mu(Y)k_\mu}{(V(s))^N}, \end{aligned} \quad (9)$$

that can be rewritten in terms of C molecules using the relation $C = \rho V$. So our method applies to a different problem with respect to the one considered in [14], in fact in our case the volume growth is not imposed a priori but dynamically evolves according to the reaction scheme, if C is produced then V increases otherwise it will keep a constant value, while in [14] the volume growth is an exogenous variable.

4. The Stochastic Simulation Algorithm in a Growing Volume

Once we have the probability function $P_\mu(\tau | Y, t)$ we can build an algorithm that reproduces the time evolution given by the model defined above.

Given the system in some state Y at time t , we must determine the interval of time τ and the reaction channel R_μ according to the probability distribution function $P_\mu(\tau | Y, t)$, and finally update the state $Y \rightarrow Y + \nu_\mu$, where ν_μ is a stoichiometric vector representing the increase and decrease of molecular abundance due to the reaction R_μ . This will be accomplished following the standard approach by Gillespie [11] but taking care of the time dependence of the propensities. We will thus need to compute the cumulative probability distribution function and then make use of the inversion method [12], to determine the channel μ and the next reaction time τ , distributed according to $P_\mu(\tau | Y, t)$.

From (8) we can compute the *cumulative distribution function*

$$F(\tau | Y, t) = \int_0^\tau \sum_{\mu} P_\mu(s | Y, t)ds, \quad (10)$$

providing the probability that any reaction will occur in $(t, t + \tau)$ starting from the state Y at time t . The function $F(\tau | Y, t)$ can be explicitly computed by the following.

Proposition 1. *Under the above assumptions we have*

$$F(\tau | Y, t) = 1 - \exp \left[-A_{Q_1}(Y)\tau - \int_t^{t+\tau} A_Q(Y, s) ds \right]. \quad (11)$$

Proof. The first step is to use (8) and perform a sum over all the channels μ to rewrite (10) as

$$F(\tau | Y, t) = \int_0^\tau (A_{Q_1}(Y) + A_Q(Y, t+s)) \times \exp \left[-A_{Q_1}(Y)s - \int_t^{t+s} A_Q(Y, r) dr \right] ds. \quad (12)$$

Then we can observe that

$$\begin{aligned} & \frac{\partial}{\partial s} \left(\exp \left[-A_{Q_1}(Y)s - \int_t^{t+s} A_Q(Y, r) dr \right] \right) \\ &= -(A_{Q_1}(Y) + A_Q(Y, t+s)) \\ & \times \exp \left[-A_{Q_1}(Y)s - \int_t^{t+s} A_Q(Y, r) dr \right], \end{aligned} \quad (13)$$

and thus

$$\begin{aligned} F(\tau | Y, t) &= - \int_0^\tau \frac{\partial}{\partial s} \left(\exp \left[-A_{Q_1}(Y)s - \int_t^{t+s} A_Q(Y, r) dr \right] \right) ds \\ &= 1 - \exp \left[-A_{Q_1}(Y)\tau - \int_t^{t+\tau} A_Q(Y, r) dr \right]. \end{aligned} \quad (14)$$

Once we have the cumulative distribution function we can obtain the value τ by drawing a random number u_1 from an uniform distribution in $[0, 1]$ and then solve with respect to τ the implicit equation:

$$u_1 = 1 - \exp \left[-A_{Q_1}(Y)\tau - \int_t^{t+\tau} A_Q(Y, s) ds \right]. \quad (15)$$

Let us stress once again that this is not as straightforward as for the original Gillespie [11] scheme, or the simplified one presented in [14], because of the time dependence of A_Q via the volume. One can nevertheless find suitable approximation for the integral, this will be the goal of the next sections.

4.1. The Adiabatic Assumption. Let us assume that τ is very small, or which is equivalent, that the time scale of the chemical reactions involving the GMMs is much faster than the production of container molecules, hence the volume growth is very slow compared with the production of the chemicals X_i .

Under this hypothesis one can assume that in the interval $(t, t+\tau)$ the volume does not vary and thus one can make the following approximation

$$\int_t^{t+\tau} A_Q(Y, s) ds \sim A_Q(Y, t)\tau. \quad (16)$$

One can thus explicitly solve (15) to get

$$\tau_{\text{Gill}} = -\frac{1}{A_{Q_1}(Y) + A_Q(Y, t)} \log(1 - u_1), \quad (17)$$

that is the standard Gillespie result except now that $A_Q(Y, t)$ depends on time and as long the volume increases, then the contribution arising from $A_Q(Y, t)$ might become smaller because $A_Q \sim 1/V$.

4.2. The Next Order Correction. One can obtain a somehow better estimate valid in the case of comparable time scales for the reactions involving GMM and the container growth. The idea is to compute the integral in (15) using the following approximation:

$$\begin{aligned} \int_t^{t+\tau} A_Q(Y, s) ds &= \int_0^\tau A_Q(Y, t+s) ds \\ &= \int_0^\tau \left(A_Q(Y, t) + \frac{\partial A_Q(Y, t)}{\partial t} s + \dots \right) ds \\ &= A_Q(Y, t)\tau + \frac{\partial A_Q(Y, t)}{\partial t} \frac{\tau^2}{2} + \mathcal{O}(\tau^3), \end{aligned} \quad (18)$$

where $\partial A_Q(Y, t)/\partial t$ can be obtained using the definition (9) and expressing the volume in terms of $C = V(t)\rho$, namely,

$$\begin{aligned} & \frac{\partial A_Q(Y, t)}{\partial t} \\ &= -\frac{\dot{C}}{C} \left(\sum_{\mu \in \mathcal{Q}_2} \frac{h_\mu(Y)k_\mu}{C(t)} + 2 \sum_{\mu \in \mathcal{Q}_3} \frac{h_\mu(Y)k_\mu}{(C(t))^2} \right. \\ & \quad \left. + \dots + N \sum_{\mu \in \mathcal{Q}_{N+1}} \frac{h_\mu(Y)k_\mu}{(C(t))^N} \right). \end{aligned} \quad (19)$$

To compute \dot{C}/C we make the assumption that in a very short time interval, as the one we are interested in, the deterministic growth of the container is a good approximation for the stochastic underlying mechanism; this implies that we can use (3)

$$\frac{\dot{C}}{C} = \left(\frac{C(t)}{\rho} \right)^{\beta-1} \frac{\vec{\alpha} \cdot \vec{X}(t)}{C(t)}. \quad (20)$$

Inserting the previous result into (18) and finally solving (15) with respect to τ , we can compute the next reaction time up to correction of the order of τ^3 , as follows

$$\begin{aligned} \tau_{\text{Gill}} &= \frac{-(A_{Q_1}(Y) + A_Q(Y, t))}{\dot{A}_Q(Y, t)} \\ & \quad + \frac{\sqrt{(A_{Q_1}(Y) + A_Q(Y, t))^2 - 2 \log(1 - u_1) \dot{A}_Q(Y, t)}}{\dot{A}_Q(Y, t)}, \end{aligned} \quad (21)$$

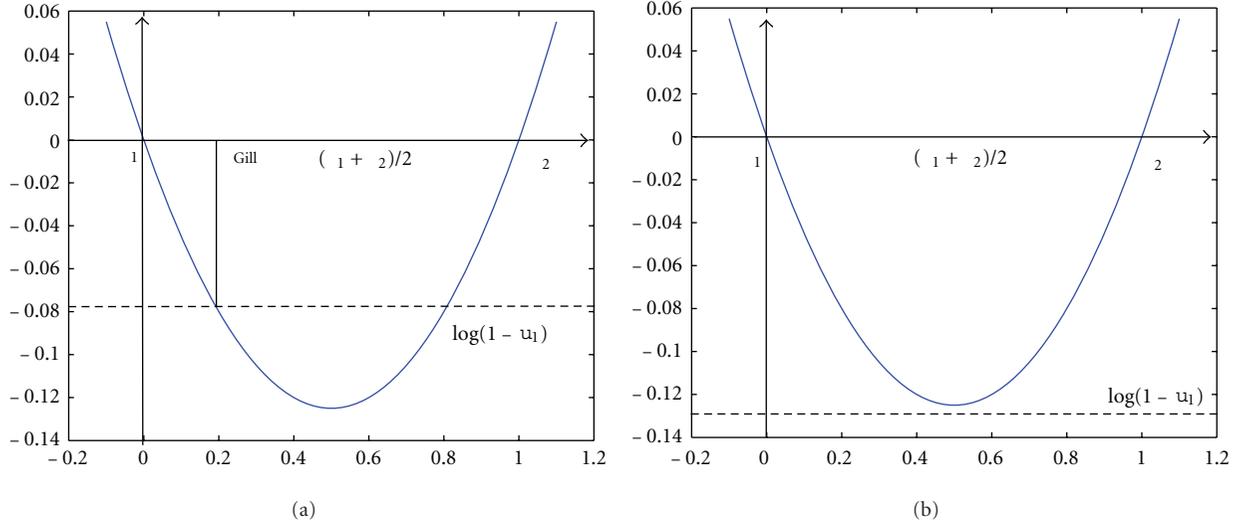


FIGURE 1: Geometrical interpretation of the existence of the next reaction time τ_{Gill} . (a) τ_{Gill} is the smallest intersection between the parabola and the horizontal line $\log(1 - u_1)$. (b) τ_{Gill} does not exist, the horizontal line is located below the minimum of the parabola.

where we wrote for short $\dot{A}_Q(Y, t) = \partial A_Q(Y, t)/\partial t$ and we selected the positive square root in such a way in the limit $\dot{A}_Q(Y, t) \rightarrow 0$ we recover the previous solution (17).

Remark 2 (On the existence of τ_{Gill}). In the case of variable volume a new phenomenon can arise: the volume growth can be so fast that no reaction can occur in the interval $(t, t + \tau + d\tau)$ for any τ . Mathematically this translates into a sign condition for the term under square root in (21), if

$$\log(1 - u_1) < \frac{(A_{Q_1}(Y) + A_Q(Y, t))^2}{(2\dot{A}_Q(Y, t))}, \quad (22)$$

then (15) has no real solution.

This can be geometrically interpreted as follows. The relation (15) determines τ_{Gill} as the intersection of the parabola $-A_{Q_1}(Y) - A_Q(Y, t)\tau - \dot{A}_Q(Y, t)\tau^2/2$ with the horizontal line $\log(1 - u_1)$, which is negative because $u_1 \in (0, 1)$. Such parabola intersect the y -axis at $\tau_1 = 0$ and $\tau_2 = -2(A_{Q_1}(Y) + A_Q(Y, t))/\dot{A}_Q(Y, t) > 0$ and it is concave. Then its absolute (negative) minimum is reached at the vertex $\tau_V = (t_1 + t_2)/2$ and its value is $(A_{Q_1}(Y) + A_Q(Y, t))^2/(2\dot{A}_Q(Y, t))$ and it is negative because $\dot{A}_Q(Y, t)$ is negative. Hence if the horizontal line is below this value, that is, condition (22) is verified, the parabola and the line do not have any real intersections (see Figure 1).

Let us also observe that, whenever it exists, τ_{Gill} is always positive as it should be. In the case of a protocell the nonexistence of such next reaction time could be translated into the death by dilution of the protocell.

4.3. The Next Reaction Channel. Whenever the next reaction time does exist, the next reaction channel is determined using the classical Gillespie method, namely, by drawing a second

uniformly distributed random number $u_2 \in [0, 1]$ and fix the channel μ such that

$$\sum_{\nu=1}^{\mu-1} a_\nu(Y, t + \tau) \leq u_2 a_0(Y, t + \tau) \leq \sum_{\nu=1}^{\mu} a_\nu(Y, t + \tau), \quad (23)$$

where $a_0(Y, t + \tau) = A_{Q_1}(Y) + A_Q(Y, t + \tau) = \sum_{\nu=1}^m a_\nu(Y, t + \tau)$.

Remark 3. Let us observe that if all the reactions involve the same number of chemicals, then the determination of which reaction channel μ will be activated in the next reaction does not depend on the volume which factorizes out from (23). In fact assuming all the reactions to involve p chemical, we obtain by definition

$$a_\nu(Y, t + \tau) = \frac{h_\nu(Y)k_\nu}{[V(t + \tau)]^p} \quad \forall \nu \in \{1, \dots, m\}, \quad (24)$$

and thus (23) rewrites

$$\sum_{\nu=1}^{\mu-1} \frac{h_\nu(Y)k_\nu}{[V(t + \tau)]^p} \leq u_2 \sum_{\nu=1}^m \frac{h_\nu(Y)k_\nu}{[V(t + \tau)]^p} \leq \sum_{\nu=1}^{\mu} \frac{h_\nu(Y)k_\nu}{[V(t + \tau)]^p}, \quad (25)$$

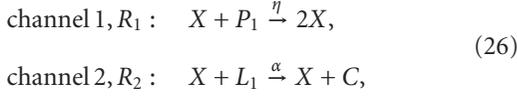
which is clearly independent of the volume value V .

5. Some Applications

The aim of this section is to provide some applications of the previous algorithm to the study of the evolution of a protocell.

5.1. One Single Genetic Memory Molecule. The simplest model is the one where only one GMM specie is present in

the protocell [9] and thus only two chemical channels are active:



where P_1 and L_1 are, respectively, precursors of GMM, that is, nucleotide, and precursors of amphiphiles.

One can thus compute the propensities in the state $Y = (X, C)$ at time t :

$$\begin{aligned} a_1(X, C, t) &= h_1(X, C) \frac{\eta}{V(t)} = \eta \frac{P_1 X}{V(t)}, \\ a_2(X, C, t) &= h_2(X, C) \frac{\alpha}{V(t)} = \alpha \frac{L_1 X}{V(t)}, \end{aligned} \quad (27)$$

let us observe that we assume that precursors are buffered and thus they are constant.

Because system (26) contains only bimolecular reactions, all the propensities are time dependent, hence $A_{Q_1} = 0$ and $A_Q = a_1(X, C, t) + a_2(X, C, t) = (P_1\eta + L_1\alpha)X/V(t)$, thus (15) simplifies into

$$u_1 = 1 - \exp\left[-\int_t^{t+\tau} A_Q(Y, s) ds\right], \quad (28)$$

whose second-order solution (21) is given by

$$\begin{aligned} \tau_{\text{Gill}} &= \frac{-A_Q(Y, t)}{\dot{A}_Q(Y, t)} \\ &\quad + \frac{\sqrt{(A_Q(Y, t))^2 - 2 \log(1 - u_1) \dot{A}_Q(Y, t)}}{\dot{A}_Q(Y, t)}, \\ \frac{\partial A_Q(X, C, t)}{\partial t} &= -\frac{\dot{V}(t)}{V(t)} \left(\frac{P_1\eta X}{V(t)} + \frac{L_1\alpha X}{V(t)} \right) \Big|_{V(t)=C(t)/\rho} \\ &= -\left(\frac{C}{\rho}\right)^{\beta-1} \frac{\rho L_1\alpha X^2}{C^2} (P_1\eta + L_1\alpha). \end{aligned} \quad (29)$$

So we can finally obtain

$$\begin{aligned} \tau_{\text{Gill}} &= \frac{C}{L_1\alpha X} \left(\frac{\rho}{C}\right)^{\beta-1} \\ &\quad - \sqrt{\left[\frac{C}{L_1\alpha X} \left(\frac{\rho}{C}\right)^{\beta-1}\right]^2 + 2 \frac{C^2}{L_1\alpha\rho X^2 (P_1\eta + L_1\alpha)} \log(1 - u_1)}, \end{aligned} \quad (30)$$

provided

$$\log(1 - u_1) \geq -\frac{\rho}{2\alpha} \left(\frac{\rho}{c}\right)^{2(\beta-1)} (P_1\eta + L_1\alpha). \quad (31)$$

Which reaction channel μ will be active in the time interval $[t, t + \tau]$ can be obtained according to

$$\begin{aligned} \text{if } u_2 \frac{(P_1\eta + L_1\alpha)X}{V} &\leq \frac{P_1\eta X}{V} \\ \text{namely } 0 \leq u_2 &\leq \frac{P_1\eta}{P_1\eta + L_1\alpha} \quad \text{then } \mu = 1 \\ \text{if } \frac{P_1\eta X}{V} < u_2 \frac{(P_1\eta + L_1\alpha)X}{V} &\leq \frac{(P_1\eta + L_1\alpha)X}{V} \\ \text{namely } \frac{P_1\eta}{P_1\eta + L_1\alpha} < u_2 &\leq 1 \quad \text{then } \mu = 2. \end{aligned} \quad (32)$$

Let us observe that according to Remark 3, the choice of μ does not depend on the volume, because only binary reactions are present.

Let C_0 be the initial amount of container molecules, then we assume that once $C(\bar{t}) = 2C_0$ the protocell splits into two offspring, almost halving the GMM amount. More precisely we assume that the first offspring will get a number of GMMs drawn according to a binomial distribution with parameter $p = 1/2$ and $n = X(\bar{t})$. From this step, for technical reason, only one randomly chosen offspring will be studied during each generation.

In Figure 2 we report a comparison between the deterministic (3) and the stochastic dynamics, under the adiabatic assumption for τ_{Gill} , corresponding to the continuous growth phase of the container between two successive divisions. As one should expect, a system composed by a large number of molecules exhibits small stochastic fluctuations whose average is not too far from the dynamics described by the deterministic model.

In Figure 3 we report the amount of GMM, $X^{(k)}$ (Figure 3(a)), at the beginning of each protocell cycle and the duplication time (Figure 3(b)), namely, the interval of time needed to double the amount of C molecules, for both the stochastic and deterministic models. Once again one can clearly observe the small fluctuations of the stochastic system around the value obtained by the numerical integration of the deterministic description (3). Let us observe that these fluctuations are due to the stochastic integrator scheme and also on the division mechanism.

We are now interested in studying the fluctuations dependence on the amount of molecules. We already know that for a sufficiently large number of molecules the stochastic dynamics follows closely the deterministic one and thus the fluctuations are small. On the other hand, one should expect that when the number of molecules decreases, then the fluctuation will rise and the system behavior could not be completely described by means of a deterministic approach. This is confirmed by Figures 4 and 5, where we can observe that a model composed by a small number of initial molecules, 20-times lesser than in the model presented in Figure 2 exhibits larger stochastic fluctuations.

In Figure 6 we summarize the results of several protocell models each one with a different amount of initial molecules, in order to appreciate the influence of the latter on the stochastic fluctuations. To compare with, we also report the case of the deterministic model. Because the kinetic constants

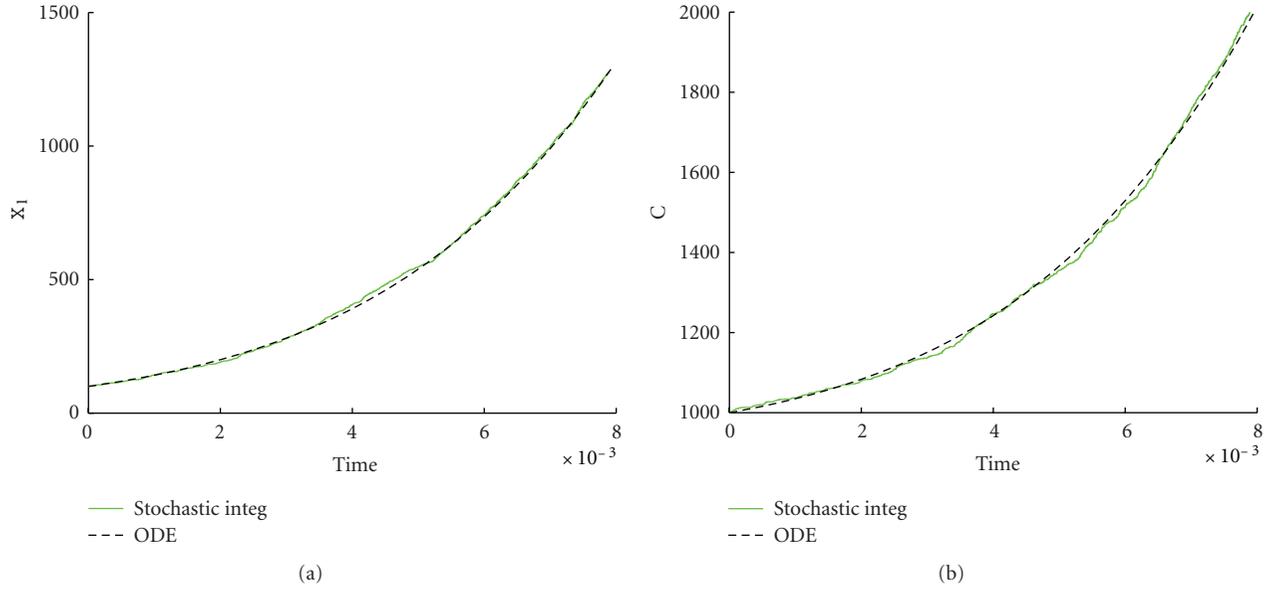


FIGURE 2: Stochastic versus ODE SRM protocell (3). Case of one GMM, (a) the time evolution of the amount of GMM, (b) the time evolution of the amount of C . Parameters are $\eta = 1$, $\alpha = 1$, $L1 = 500$, $P1 = 600$, $X_1(0) = 100$, $C(0) = 1000$, $\rho = 200$, and $\beta = 2/3$.

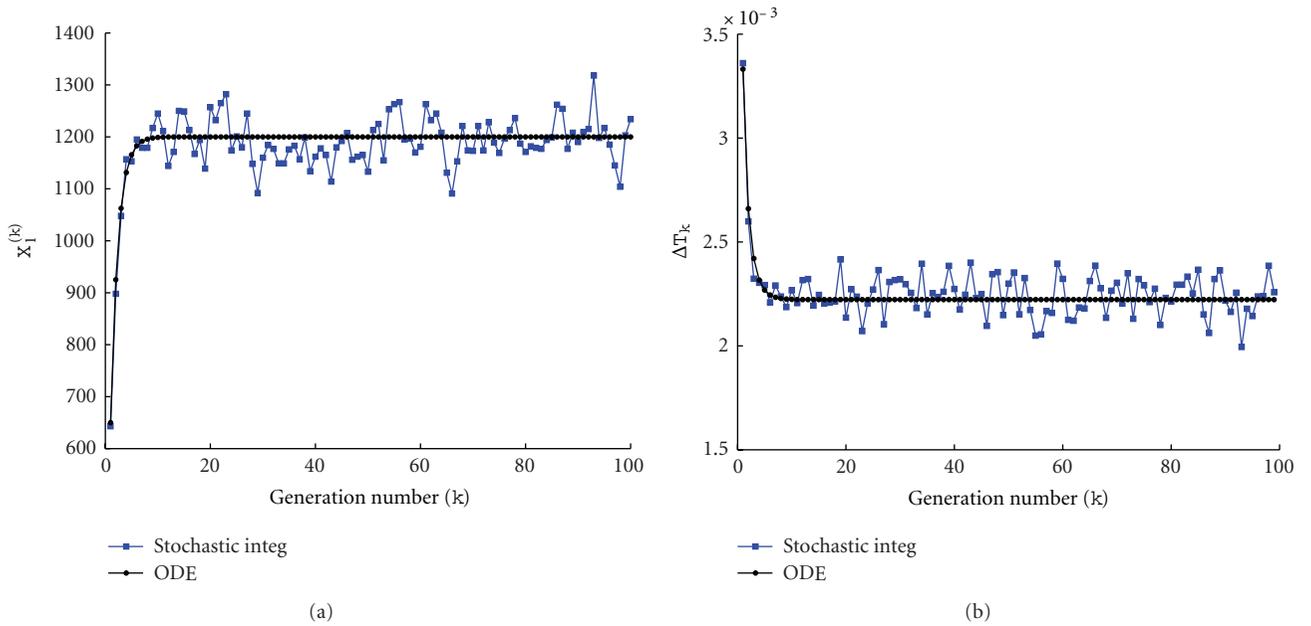


FIGURE 3: Stochastic versus ODE SRM protocell (3). Case of one GMM, (a) the amount of GMM at the beginning of each division cycle, (b) the division time as a function of the number of elapsed divisions. Parameters are $\eta = 1$, $\alpha = 1$, $L1 = 500$, $P1 = 600$, $X_1(0) = 100$, $C(0) = 1000$, $\rho = 200$, and $\beta = 2/3$.

are kept constant, the analytical theory for the deterministic model ensures that the division time does not vary [7]. Nevertheless the fewer the initial amount of X_0 and C_0 is, the larger the fluctuations present in the stochastic integration are.

To get a more complete understanding of the fluctuations dependence, we decided to measure them using the standard deviation of the protocell division time (after a sufficiently

long transient phase). In Figure 7 we report the standard deviation of the division time ΔT as a function of the initial amount of molecules. As expected the fluctuations strength decreases rapidly as soon as the number of molecules increases and the relation can be very well approximated by a power law distribution with exponent -0.54 ± 0.03 (linear best fit).

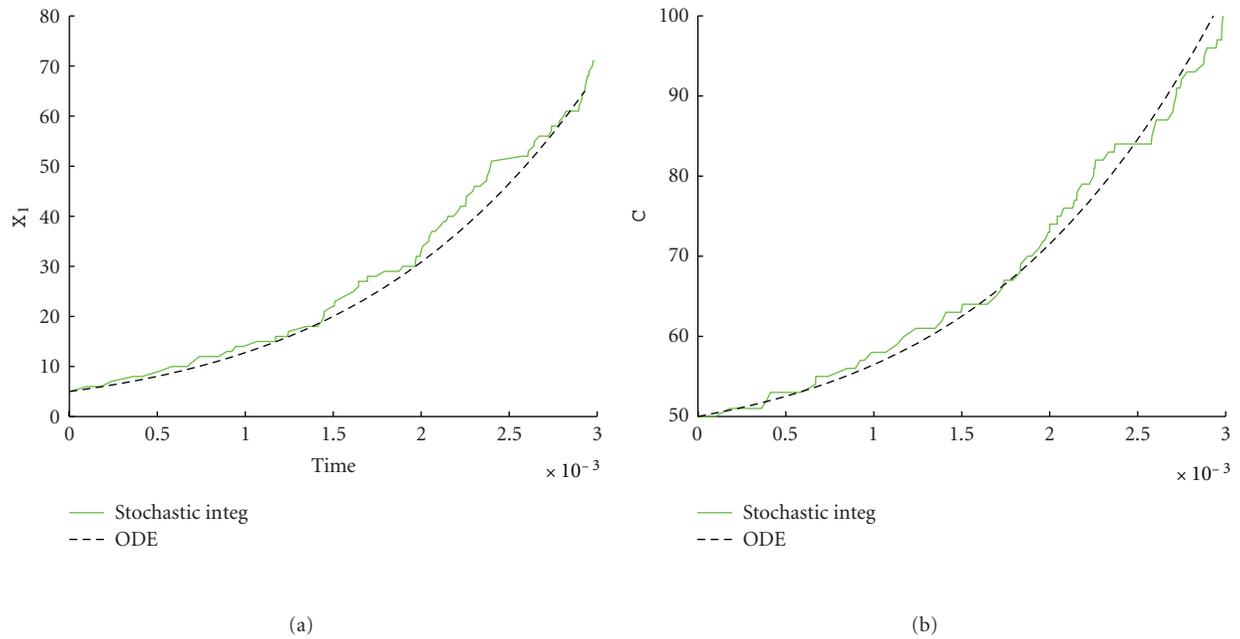


FIGURE 4: Stochastic versus ODE SRM protocell (3). Case of one GMM, (a) the time evolution of the amount of GMM, (b) the time evolution of the amount of C. Parameters are $\eta = 1$, $\alpha = 1$, $L1 = 500$, $P1 = 600$, $X_1(0) = 5$, $C(0) = 50$, $\rho = 200$, and $\beta = 2/3$.

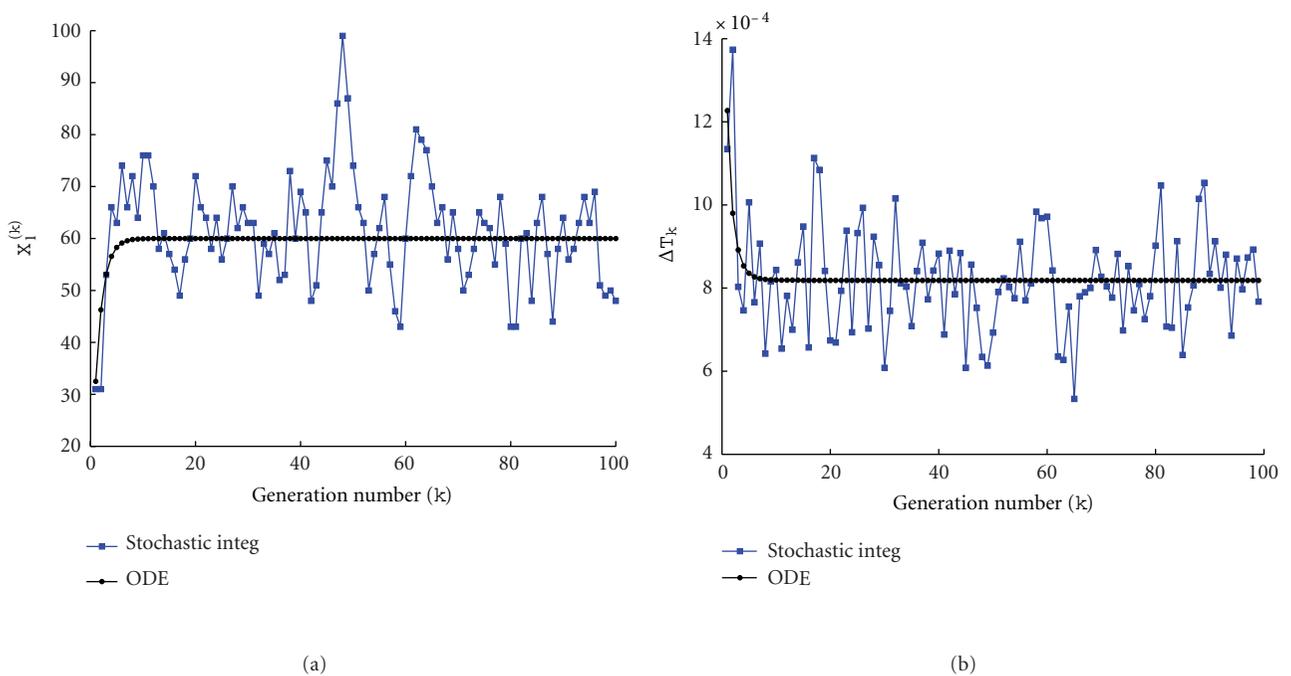


FIGURE 5: Stochastic versus ODE SRM protocell (3). Case of one GMM, (a) the amount of GMM at the beginning of each division cycle, (b) the division time as a function of the number of elapsed divisions. Parameters are $\eta = 1$, $\alpha = 1$, $L1 = 500$, $P1 = 600$, $X_1(0) = 5$, $C(0) = 50$, $\rho = 200$, and $\beta = 2/3$.

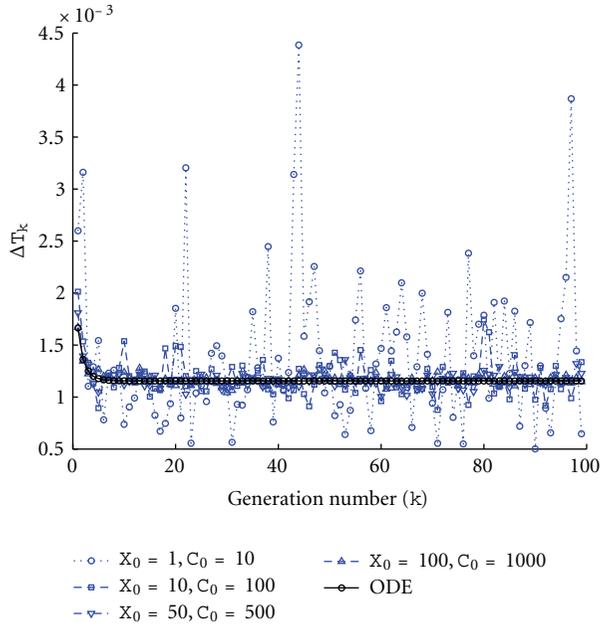


FIGURE 6: Fluctuation dependence on the initial conditions. We report the division times as a function of the number of elapsed divisions, for 5 different protocells models. Protocell \circ : $X_1(0) = 5$, $C(0) = 10$, protocell \square : $X_1(0) = 10$, $C(0) = 100$, protocell ∇ : $X_1(0) = 50$, $C(0) = 500$, protocell \triangle : $X_1(0) = 100$, $C(0) = 1000$. The black line denotes the deterministic protocell. All the remaining parameters have been fixed to $\eta = 1$, $\alpha = 1$, $L_1 = 500$, $P_1 = 600$, $\rho = 100$, and $\beta = 1$.

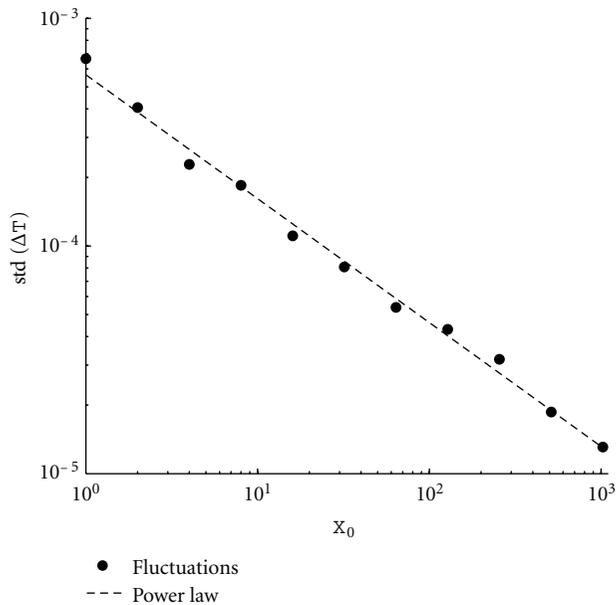
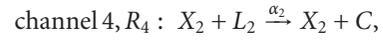
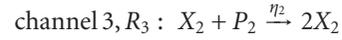
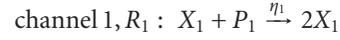


FIGURE 7: Fluctuation dependence on the initial conditions. We report the standard deviation of the protocell division time as a function of the initial amount of molecules X_0 (\bullet) and a linear best fit, whose slope is -0.54 ± 0.03 . Parameters are $X(0) = 2^n$ with $n = 0, \dots, 10$, $C(0) = 10X(0)$, $\eta = 1$, $\alpha = 1$, $L_1 = 500$, $P_1 = 600$, $\rho = 100$, and $\beta = 1$.

5.2. *Two Noninteracting Genetic Memory Molecules.* A slightly more sophisticated model can be obtained by considering two linear non interacting GMMs. The system can be described by the following chemical reactions:



(33)

where P_i and L_i are, respectively, precursors of the i th GMM, that is, nucleotide, and precursors of amphiphiles used by the i th GMM to build a C molecule.

As previously done, we compare the stochastic and the deterministic models. Results are reported in Figure 8 and one can still observe that in presence of a large number of molecules the deterministic dynamics well approximates the stochastic model. On the other hand, the protocell division time exhibits large fluctuations around the deterministic value even in presence of quite large number of molecules (see Figure 9(b)).

The parameters have been set in such a way that only one GMM will survive according to the analytical theory for the deterministic model. One can observe that, despite the fluctuations, the same fate is obtained for the stochastic model (see Figure 9(b)).

Once we reduce the number of involved molecules, the stochastic fluctuations dramatically increase (see Figures 10 and 11).

As in the case of only one GMM, when two non interacting linear GMMs are present the size of the stochastic fluctuations as a function of the initial number of molecules follows a power law distribution with exponent -0.51 ± 0.05 (linear best fit), see Figure 12: the fewer the molecules in the system are, the larger the fluctuations around the deterministic dynamics are.

A new phenomenon arises in the case of two GMMs modeled by a stochastic process. There can be a *breaking of the symmetry* emerging in systems composed of two identical GMMs (i.e., equal kinetic constants, equal initial amounts, and availability of precursors) present with a few initial amounts of each one. Although adopting a deterministic approach the dynamics of the two replicators would be perfectly the same, a small fluctuation in the very first instants of the protocell evolution entails the dilution of one of the two replicators and thus a different fate for the protocell. Let us observe that the probability to have a large fluctuation is never zero, thus waiting for a sufficiently long time, a specie can always disappear from the system, thus giving rise to the breaking of the symmetry phenomenon. See Figure 13 where we report, as a function of the initial amount of molecules $X_i(0)$, $i = 1, 2$, the proportion of simulations where the symmetry breaking has been observed repeating 50 times each simulation with the same set of parameters and initial conditions during 100 generations.

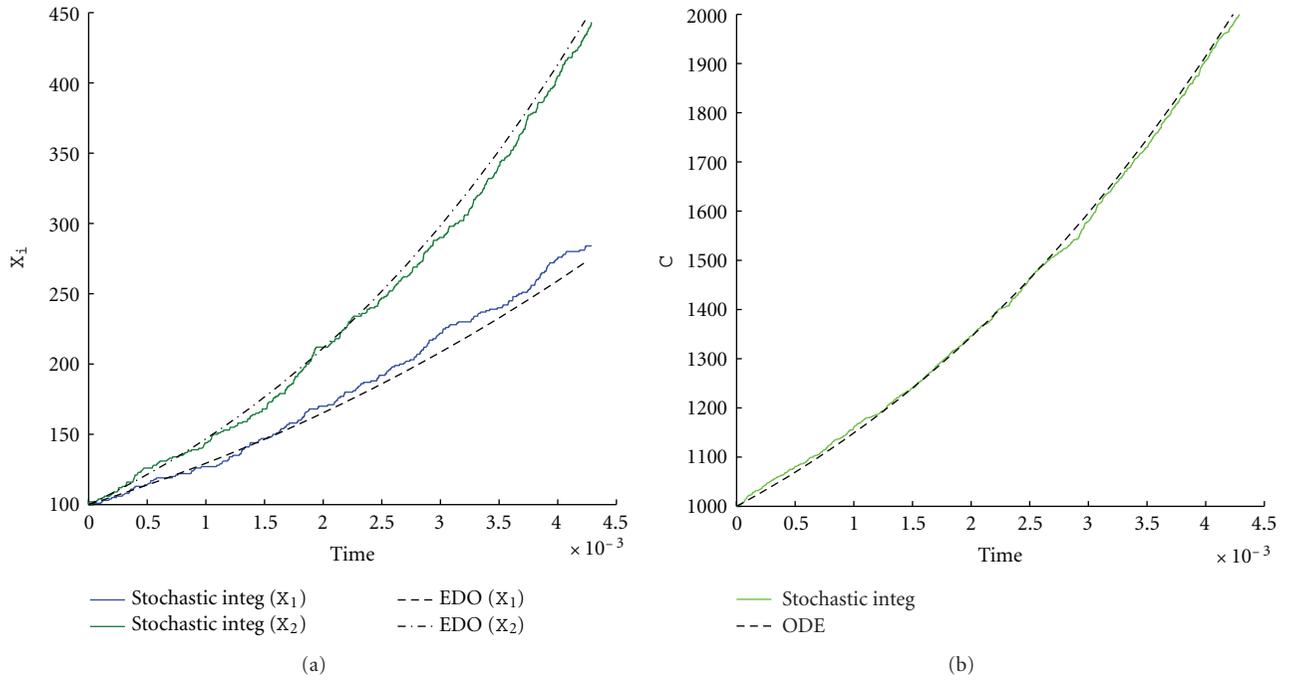


FIGURE 8: Stochastic versus ODE SRM protocell (3). Case of two GMMs, (a) the time evolution of the amount of GMM during a division cycle, (b) the time evolution of the amount of C molecules. Parameters are $\eta_1 = \eta_2 = 1$, $\alpha_1 = \alpha_2 = 2$, $L_1 = 500$, $L_2 = 600$, $P_1 = 600$, $P_2 = 670$, $X_1(0) = X_2(0) = 100$, $C(0) = 1000$, $\rho = 200$, and $\beta = 2/3$.

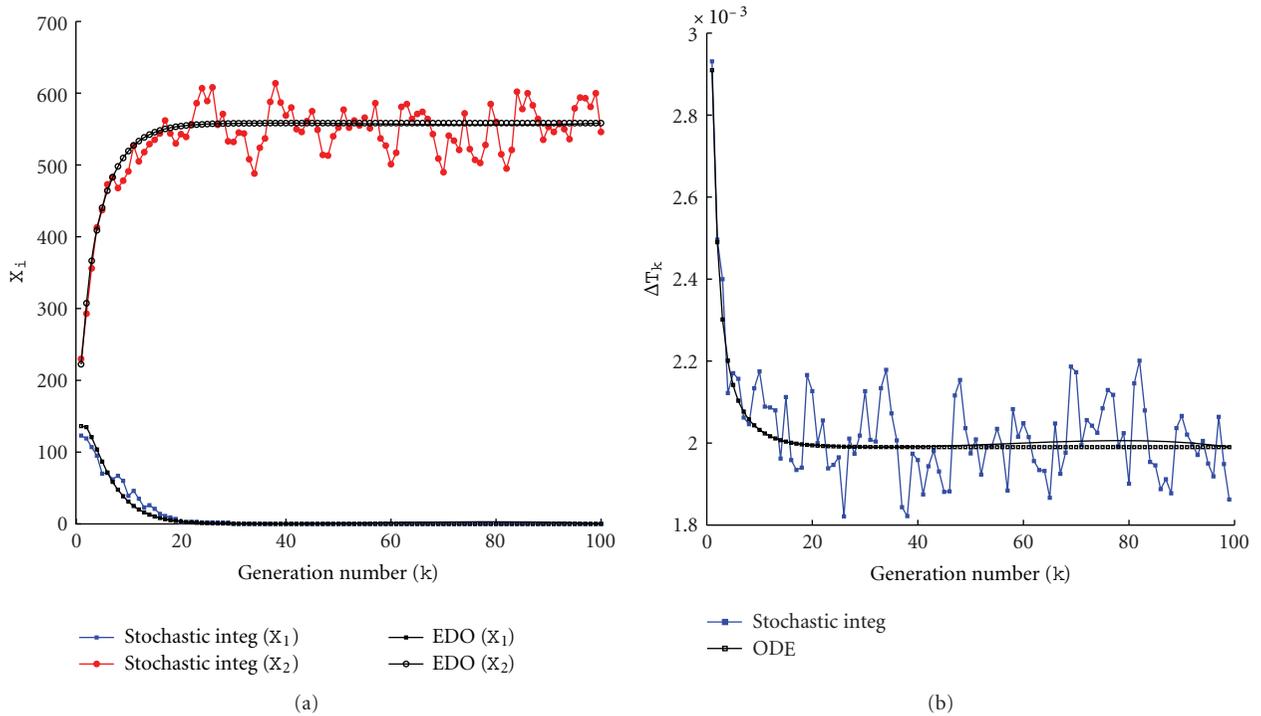


FIGURE 9: Stochastic versus ODE SRM protocell (3). Case of two GMMs, (a) the amount of GMM at the beginning of each division cycle, (b) the division time as a function of the number of elapsed divisions. Parameters are $\eta_1 = \eta_2 = 1$, $\alpha_1 = \alpha_2 = 2$, $L_1 = 500$, $L_2 = 600$, $P_1 = 600$, $P_2 = 670$, $X_1(0) = X_2(0) = 100$, $C(0) = 1000$, $\rho = 200$, and $\beta = 2/3$.

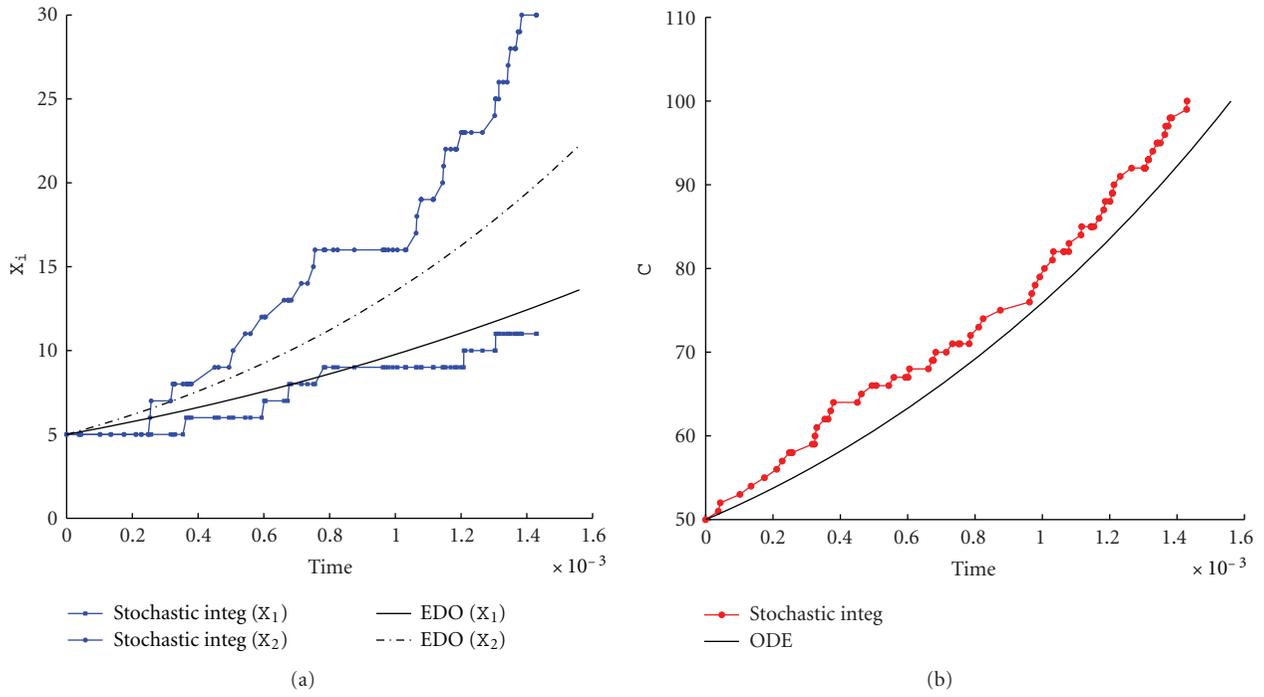


FIGURE 10: Stochastic versus ODE SRM protocell (3). Case of two GMMs, (a) the time evolution of the amount of GMM during a division cycle, (b) the time evolution of the amount of C molecules. Parameters are $\eta_1 = \eta_2 = 1$, $\alpha_1 = \alpha_2 = 2$, $L_1 = 500$, $L_2 = 600$, $P_1 = 450$, $P_2 = 670$, $X_1(0) = X_2(0) = 5$, $C(0) = 50$, $\rho = 200$, and $\beta = 2/3$.

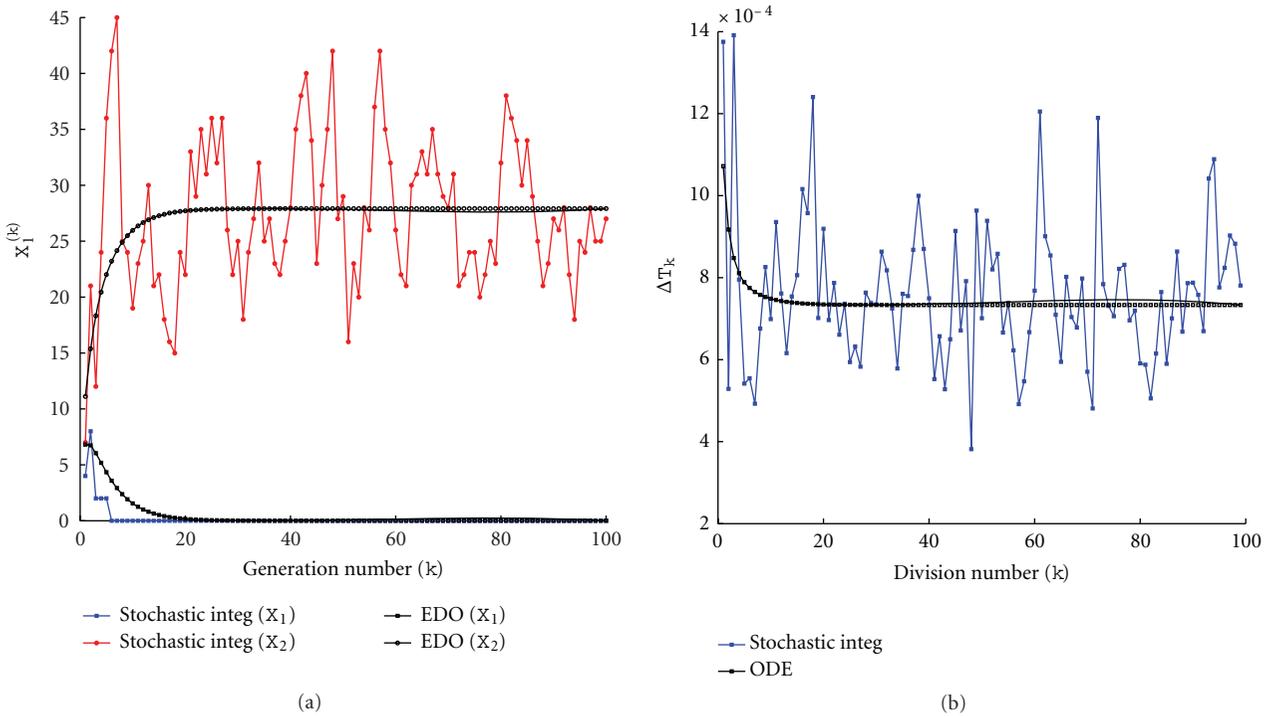


FIGURE 11: Stochastic versus ODE SRM protocell (3). Case of two GMMs, (a) the amount of GMM at the beginning of each division cycle, (b) the division time as a function of the number of elapsed divisions. Parameters are $\eta_1 = \eta_2 = 1$, $\alpha_1 = \alpha_2 = 2$, $L_1 = 500$, $L_2 = 600$, $P_1 = 450$, $P_2 = 670$, $X_1(0) = X_2(0) = 5$, $C(0) = 50$, $\rho = 200$, and $\beta = 2/3$.

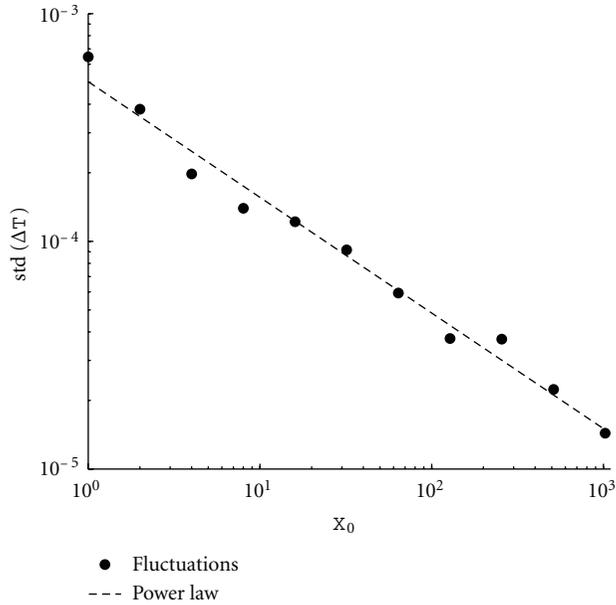


FIGURE 12: Fluctuation dependence on the initial conditions. We report the standard deviation of the protocell division time as a function of the initial amount of molecules $X_i(0)$, $i = 1, 2$, (\bullet) and a linear best fit, whose slope is $= -0.51 \pm 0.05$. Parameters are $X_1(0) = X_2(0) = 2^n$ with $n = 0, \dots, 10$, $C(0) = 10X_1(0)$, $\eta_1 = \eta_2 = 1$, $\alpha_1 = \alpha_2 = 2$, $L1 = 500$, $L2 = 500$, $P1 = 500$, $P_2 = 600$, $\rho = 100$, and $\beta = 1$.

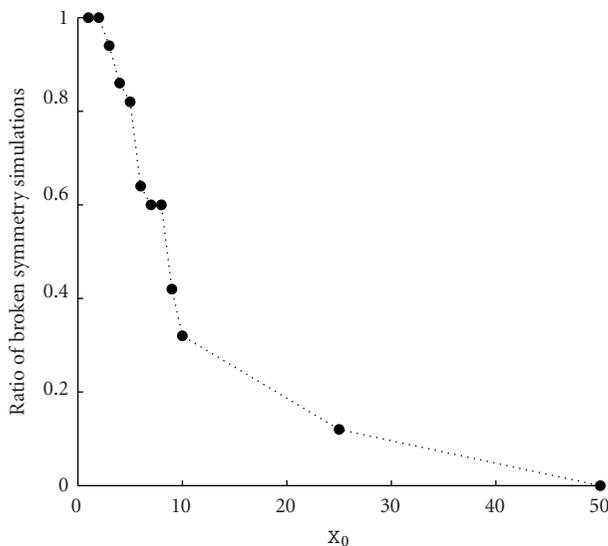


FIGURE 13: Symmetry breaking phenomenon. Each point denotes the fraction of runs exhibiting the symmetry breaking phenomenon, during 100 generations, over 50 identical replicas. Parameters are $X_1(0) = X_2(0) = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 25, 50]$, $C(0) = 10X$, $\eta_1 = \eta_2 = 1$, $\alpha_1 = \alpha_2 = 2$, $L1 = 500$, $L2 = 500$, $P1 = 600$, $P_2 = 600$, $\rho = 100$, and $\beta = 1$.

6. Conclusion

In this paper we presented a new stochastic integration algorithm based on the one introduced by Gillespie. Our contribution is devoted to the explicit introduction of the volume variation in the algorithm, which moreover is directly related to the amount of contained molecules, and thus it evolves in a self-consistent way.

This algorithm straightforwardly adapts to the study of the evolution of a protocell, simplified form of cells, where an ensemble of chemical reactions occurs in a varying volume, the volume of the protocell, that in turn increases because of the production of container molecules.

We presented several protocell models and we compare them with the analogous deterministic protocell models, namely, solved using the ODE. In this preliminary study, we emphasized the role of the fluctuations and their dependence on the initial amount of molecules. The dynamics is richer than the deterministic one and thus it is worth studying, in particular we deserve to future investigations the case where the interactions among the molecules can be modeled by a linear system, whose interaction matrix is not diagonal; the off diagonal terms representing the cross-catalysis. Also the case of nonlinear interactions will be deferred to a forthcoming paper. Also the study of the emergence of time-periodic patterns due to the fluctuations, will be analyzed. An analytical treatment of the latter case could be possible using some recent technics developed by [15, 16], see also [17] where the space is also taken into account.

Acknowledgments

The work of T. Carletti has been partially supported by the FNRS Grant Mission ‘‘Scientifique 2010-2011.’’ T. Carletti would like also to thank the European Center for Living Technology in Venice (Italy) for the warm hospitality during a short visiting stay that originated this collaboration.

References

- [1] B. Alberts et al., *Molecular Biology of the Cell*, Garland, New York, NY, USA, 2002.
- [2] S. Rasmussen, L. Chen, D. Deamer et al., ‘‘Transitions from nonliving to living matter,’’ *Science*, vol. 303, no. 5660, pp. 963–965, 2004.
- [3] J. W. Szostak, D. P. Bartel, and P. L. Luisi, ‘‘Synthesizing life,’’ *Nature*, vol. 409, no. 6818, pp. 387–390, 2001.
- [4] S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. A. Treco, and J. W. Szostak, ‘‘Template-directed synthesis of a genetic polymer in a model protocell,’’ *Nature*, vol. 454, no. 7200, pp. 122–125, 2008.
- [5] T. Oberholzer, R. Wick, P. L. Luisi, and C. K. Biebricher, ‘‘Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell,’’ *Biochemical and Biophysical Research Communications*, vol. 207, no. 1, pp. 250–257, 1995.
- [6] S. Rasmussen, L. Chen, B. M. R. Stadler, and P. F. Stadler, ‘‘Photo-organism kinetics: evolutionary dynamics of lipid aggregates with genes and metabolism,’’ *Origins of Life and Evolution of the Biosphere*, vol. 34, no. 1-2, pp. 171–180, 2004.

- [7] T. Carletti, R. Serra, I. Poli, M. Villani, and A. Filisetti, "Sufficient conditions for emergent synchronization in protocell models," *Journal of Theoretical Biology*, vol. 254, no. 4, pp. 741–751, 2008.
- [8] P. L. Luisi, F. Ferri, and P. Stano, "Approaches to semi-synthetic minimal cells: a review," *Naturwissenschaften*, vol. 93, no. 1, pp. 1–13, 2006.
- [9] R. Serra, T. Carletti, and I. Poli, "Synchronization phenomena in surface-reaction models of protocells," *Artificial Life*, vol. 13, no. 2, pp. 123–138, 2007.
- [10] K. Kaneko, *Life : An Introduction to Complex System Biology*, Springer, The Netherlands, 2006.
- [11] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.
- [12] D. T. Gillespie, *Markov Processes: An Introduction for Physical Scientists*, Academic Press, Boston, Mass, USA, 1992.
- [13] A. M. Kierzek, "STOCKS: STOChastic kinetic simulations of biochemical systems with Gillespie algorithm," *Bioinformatics*, vol. 18, no. 3, pp. 470–481, 2002.
- [14] T. Lu, D. Volfson, L. Tsimring, and J. Hasty, "Cellular growth and division in the Gillespie algorithm," *Systems biology*, vol. 1, no. 1, pp. 121–128, 2004.
- [15] A. J. McKane and T. J. Newman, "Predator-prey cycles from resonant amplification of demographic stochasticity," *Physical Review Letters*, vol. 94, no. 21, Article ID 218102, 4 pages, 2005.
- [16] T. Dauxois, F. Di Patti, D. Fanelli, and A. J. McKane, "Enhanced stochastic oscillations in autocatalytic reactions," *Physical Review E*, vol. 79, no. 3, 2009.
- [17] P. De Anna, F. Di Patti, D. Fanelli, A. J. McKane, and T. Dauxois, "Spatial model of autocatalytic reactions," *Physical Review E*, vol. 81, no. 5, Article ID 056110, 2010.

Research Article

Detecting Epileptic Seizure from Scalp EEG Using Lyapunov Spectrum

Truong Quang Dang Khoa,¹ Nguyen Thi Minh Huong,² and Vo Van Toi¹

¹Biomedical Engineering Department, International University of Vietnam National Universities, Ho Chi Minh City, Vietnam

²Faculty of Applied Science, University of Technology of Vietnam National Universities, Ho Chi Minh City, Vietnam

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

Received 30 September 2011; Accepted 28 November 2011

Academic Editor: Carlo Cattani

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

One of the inherent weaknesses of the EEG signal processing is noises and artifacts. To overcome it, some methods for prediction of epilepsy recently reported in the literature are based on the evaluation of chaotic behavior of intracranial electroencephalographic (EEG) recordings. These methods reduced noises, but they were hazardous to patients. In this study, we propose using Lyapunov spectrum to filter noise and detect epilepsy on scalp EEG signals only. We determined that the Lyapunov spectrum can be considered as the most expected method to evaluate chaotic behavior of scalp EEG recordings and to be robust within noises. Obtained results are compared to the independent component analysis (ICA) and largest Lyapunov exponent. The results of detecting epilepsy are compared to diagnosis from medical doctors in case of typical general epilepsy.

1. Introduction

Since the discovery of the human electroencephalographic (EEG) signals by Hans Berger in 1923, the EEG has been the most commonly used instrument for clinical evaluation of brain activity, classification epileptic seizures or no epilepsy, schizophrenia, sleep disorder, mental fatigue, and coma.

There were many researches on EEG in the world. An EEG signal is a measurement of currents that flow during synaptic excitations of the dendrites of many pyramidal neurons in the cerebral cortex. When brain cells (neurons) are activated, the synaptic currents are produced within the dendrites. This current generates a secondary electrical field over the scalp measurable by EEG systems. They are captured by multiple-electrode EEG machines either from inside the brain, over the cortex under the skull, or certain locations over the scalp and can be recorded in different formats.

Today, the epilepsy is important problem in the public healthy and everyone should be specially interested in it because its effects are influenced on the life's qualities, study, and working abilities, falling in line with society badly. Epilepsy is the most common neurological disorder, second

only to stroke. Nearly 60 million people worldwide are diagnosed with epilepsy whose hallmark is recurrent seizures [1]. Some 35 million have no access to appropriate treatment. This is either because services are nonexistent or because epilepsy is not viewed as a medical problem or a treatable brain disorder.

Most traditional analyses of epilepsy, based on the EEG, are focused on the detection and classification of epileptic seizures. Among them, the best method of analysis is still the visual inspection of the EEG by a highly skilled electroencephalographer. However, with the advent of new signal processing methodologies based on the mathematical theory, there has been an increased interest in the analysis of the EEG for prediction of epileptic seizures.

To detect spike, Gotman and Wang [2] defined 5 states (active wakefulness, quiet wakefulness, desynchronized EEG, phasic EEG, and slow EEG) and designed one method for automatic state classification. Then, they designed procedures for identification of nonepileptic transients (eye blinks, EMG, alpha, spindles, vertex sharp waves) by measuring parameters such as relative amplitude, sharpness, and duration of EEG waves. This method is sensitive to various

artifacts. In attempts to overcome that artifacts, Dingle et al. [3] developed a multistage system to detect the epileptiform activities from the EEGs. In another approach, Glover et al. [4] used a correlation-based algorithm that was attempted to reduce the muscle artefacts in multichannel EEGs. So, approximately 67% of the spikes can be detected. By incorporating both multichannel temporal and spatial information, and including the electrocardiogram, electromyogram, and electrooculogram information into a rule-based system [5], a higher detection rate was achieved. Artificial neural networks (ANNs) have been used for seizure detection by many researchers [6, 7]. To predict epilepsy, Zhu and Jiang [8] tracks the time evolution of the slow wave energy bigger than some preset threshold from scalp EEGs. The results from four generalized epileptic patients demonstrate that pre-seizure transition phases of several minutes can be identified clearly by their linear predictor. Among recent works, time-frequency (TF) approaches effectively use the fact that the seizure sources are localized in the time-frequency domain. Most of these methods are mainly for detection of neural spikes [9] of different types. The methods are especially useful since the EEG signals are statistically nonstationary.

One of tendencies to predict seizure is nonlinear methods. The brain is assumed to be a dynamical system, since epileptic neuronal networks are essentially complex nonlinear structures and a nonlinear behavior of their interactions is, thus, expected. So, these methods have substantiated the hypothesis that quantification of the brain's dynamical changes from the EEG might enable prediction of epileptic seizures, while traditional methods of analysis have failed to recognize specific changes prior to seizure [1]. These include reduction in correlation integrals during the ictal state (Lerner, 1996) [10] and decrease in signal complexity during seizures. In 1998, Le Van Quyen et al. [11] contributed a new measurement in prediction seizure which they called "correlation density". Then, this group has introduced newer nonlinear techniques, such as the "dynamical similarity index" [12, 13], which measures similarity of EEG dynamics between recordings taken at distant moments in time. Jerger et al. [14] and Jouny et al. [15] used two methods, one of which, Gabor atom density, estimates intracranial EEGs in terms of synchrony and complexity. In another other approach, Esteller et al. [16] introduced parameter of average energy of EEG signal. They demonstrated that when seizure happens, there were bursts of complex epileptiform activity, delta slowing, subclinical seizures, and gradual increases in energy in the epileptic focus. Harrison et al. [17] measured the amount of energy in EEG signal and its averaged power within moving windows.

Iasemidis introduced ideas of chaotic in predicting seizure. In 1988 and 1990, Iasemidis et al. [18] published the first of a number of prominent articles describing another nonlinear measure for predicting seizures, primarily the largest Lyapunov exponent, for characterizing intracranial EEG recordings [19]. The lowest values of Lyapunov occur during the seizure but they are still positive denoting the presence of a chaotic attractor. Then, this group introduced an efficient version of the Largest Lyapunov Exponent (L_{max}) named *Short-Term Maximum Lyapunov Exponent*

(STL $_{max}$) and proved the relationship between the temporal evolution of L_{max} and the development of epileptic seizures [20].

Most of these studies for prediction of epilepsy are based on intracranial EEG recordings. These methods faced main challenge. This is hazardous to the patient, especially the children. The scalp EEG is the most popular recording in Hospitals. But the scalp signals are more subject to environmental noise and artifacts than the intracranial EEG, and the meaningful signals are attenuated and mixed in their way via soft tissue and bone. So, the tradition methods such as the Kolmogorov entropy or the Lyapunov exponents, may be affected by the after mentioned two difficulties and, therefore, they may not distinguish between slightly different chaotic regimes of the scalp EEG [21]. There are many researchers to be interested in this problem. They tried to applied tradition nonlinear measurement to scalp EEG. This is the approach followed by Hively and Protopopescu [22]. They proposed a method based on the *phase-space dissimilarity measures* (PSDMs) for forewarning of epileptic events from scalp EEG. Iasemidis et al. [23], using the spatiotemporal evolution of the short-term largest Lyapunov exponent, demonstrated that minutes or even hours before seizure, multiple regions of the cerebral cortex progressively approach a similar degree of chaoticity of their dynamical states. They called it *dynamical entrainment*. This method has also been shown to work well on scalp-unfiltered EEG data for seizure predictability. In 2006, a research group of Saeid Sanei developed a novel approach to quantify the dynamical changes of the brain using the scalp EEG by means of an effective block-based blind source separation (BSS) technique to separate the underlying sources within the brain to overcome problems of noises and artifacts. Their methods are promising but their results also faced noises and artifact [1].

Here, we are only interested in applying the Lyapunov exponent for scalp EEG to predict epilepsy. Like previous methods, the main problem to apply the Lyapunov exponents for scalp EEG is noises. We executed combined ICA method and Lyapunov exponent by Rosenstein. In addition, we also find improvements of Lyapunov spectrum in estimating the Lyapunov exponent so that it can be more robust, especially with respect to the presence of noise in the EEG.

This paper is organized as follows. In Section 2, we describe the algorithms for filtering, estimating that the Lyapunov exponent, especially Lyapunov spectrum, considered as an optimization model for estimating Lyapunov exponents is presented. In Section 3, the EEG recording procedure is explained and the results are compared with the other methods. Conclusions are provided in Section 4.

2. Materials and Methods

2.1. Materials. The experimental data were derived from the Hospital 115 in Ho Chi Minh City, Vietnam using a Galileo EEG machine (EBNEURO, Italy) and divided into three groups: seizures (8 files), brain function disorder due to epilepsy or transform (7 files), nonseizure (15 files).

2.2. *Preprocessing.* Frequencies of EEG signals are less than 100 Hz. In addition, most recordings present a 50-Hz frequency component contaminating several electrodes. Therefore, the signals need to be lowpass filtered to eliminate this frequency component and other high-frequency components generally produced by muscular activity. A Butterworth filter of order 10 with a cutoff frequency of 45 Hz is used [1]. Within this range of frequencies, we still have the complete information about the signals.

2.3. *Independent Component Analysis (ICA) [24, 25].* After the preprocessing step, the scalp EEG is still contaminated by noise and artifacts such as eye blinks. Independent component analysis (ICA) is an effective method for removing artifacts, especially eye blinks, and separating sources of the brain signals from these recordings. ICA methods are based on the assumptions that the signals recorded on the scalp are mixtures of time courses of temporally independent cerebral and artifactual sources, that potentials arising from different parts of the brain, scalp, and body are summed linearly at the electrodes, and that propagation delays are negligible.

2.4. *Lyapunov Exponents.* The EEG recorded from one site is inherently related to the activity at other sites. This makes the EEG a multivariable time series. Generally, an EEG signal can be considered as the output of a nonlinear system, which may be characterized deterministically. Methods for calculating these dynamical measures from experimental data have been published [26]. Among them, Lyapunov exponent is one of parameters to measure chaos of a nonlinear dynamical system and characterizes the spatiotemporal dynamics in electroencephalogram (EEGs) time series recorded from patients with temporal lobe epilepsy. Wolf et al. [27] proposed the first algorithm for calculating the largest Lyapunov exponent. But the Wolf algorithm only estimates the largest Lyapunov exponent and the first few nonnegative Lyapunov exponents, not the whole spectrum of exponents. It is sensitive to noises of time series as well as to the degree of measurement or unreliable for small data sets. So, Iasemidis et al. presented algorithm of estimating the short-term largest Lyapunov exponent, which is a modified version of the program proposed of Wolf. This modification is necessary for predicting seizure (small data segments of epileptic data). Besides, there were many improvements in estimating the Lyapunov exponent of many researchers in the world such as Eckmann et al. [28], Brown et al. [29], and Rosenstein et al. [30]. Here, we also used the algorithm of Rosenstein because of its advantages. The Rosenstein algorithm is fast, easy to implement, and robust to changes in the following quantities: embedding dimension, size of data set, reconstruction delay, and noise level. Furthermore, one may use the algorithm to calculate simultaneously the correlation dimension. Thus, one sequence of computations will yield an estimate of both the level of chaos and the system complexity.

2.5. *The Rosenstein Algorithm [30].* The first step of our approach involves reconstructing the attractor dynamics

from a single time series. We use the method of delays since one goal of our work is to develop a fast and easily implemented algorithm. The reconstructed trajectory, X , can be expressed as a matrix where each row is a phase-space vector. That is,

(i) vector X_i in phase space:

$$x_i = (x(t_i), x(t_i + \tau), \dots, x(t_i + (p - 1) * \tau)), \quad (1)$$

where τ is the *lag* or *reconstruction delay*, p is the *embedding dimension*, and $t_i \in [1, T - (p - 1)\tau]$

(i) from the definition of λ_1 given in theory $d(t) = C \exp(\lambda_1 t)$, we assume that the j th pair of nearest neighbors diverge proximately at a rate given by the largest Lyapunov exponent:

$$d_j(i) \approx C_j e^{(i \cdot \Delta t)}, \quad (2)$$

where C_j is the initial separation. By taking the logarithm of both sides of (2), we have

$$\ln d_j(i) \approx \ln C_j + \lambda_1 (i \cdot \Delta t). \quad (3)$$

Equation (3) represents a set of approximately parallel lines (for $j = 1, 2, \dots, m$), each with a slope roughly proportional to λ_1 . The largest Lyapunov exponent is easily and accurately calculated using a least-squares fit to the “average” line defined by

$$y(i) = \frac{1}{\Delta t} \langle \ln d_j(i) \rangle, \quad (4)$$

where $\langle \cdot \rangle$ denotes the average over all values of j . This process of averaging is the key to calculating accurate values of λ_1 using small, noisy data sets. Note that in (3), C_j performs the function of normalizing the separation of the neighbors, but as shown in (4), this normalization is unnecessary for estimating λ_1 . By avoiding the normalization, the current approach gains a slight computational advantage over the method by Sato et al. [31].

2.5.1. *The Lyapunov Spectrum [32].* Another way to view Lyapunov exponents is the loss of predictability as we look forward in time. If we assume that the true starting point x_0 of a time series is possibly displaced by an ε , we know only the information area I_0 about the starting point. After some steps, the time series is in the information area at time t , I_t . The information about the true position of the data decreases due to the increase of the information area. Consequently, we get a bad predictability. The largest Lyapunov exponent can be used for the description of the average information loss; $\lambda_1 > 0$ leads to bad predictability [32]. While there is a method which is applicable to many dimensional chaos to extract physical quantities from experimentally obtained irregular signals is Lyapunov spectrum [33]. It estimates the spectrum of several Lyapunov exponents (including positive, zeros, and even negative ones). This is necessary

for quantifying many physical quantities, especially for complicating EEG signals. Besides, in EEG processing, a main problem is noises and artifacts. There are many researches about processing EEG, especially removing noises to predict epilepsy. But most of reports only solved part of problems and faced some difficulties. Here, we will describe a method of Lyapunov spectrum which is shown to behave well in the perturbation of certain parameter values, but slightly sensitive in the presence of noise, good accuracy with great easy. It is suitable to prediction seizure.

Let us consider an observed trajectory $x(t)$, which can be considered as a solution of a certain dynamical system:

$$\dot{x} = F(x) \quad (5)$$

defined in a d -dimensional phase space. On the other hand, the evolution of a tangent vector ξ in a tangent space at $x(t)$ is represented by linearizing (5):

$$\dot{\xi} = T(x(t)) \cdot \xi, \quad (6)$$

where $T = DF = \partial F/\partial x$ is the Jacobian matrix of F . The solution of the linear nonautonomous (6) can be obtained as

$$\xi(t) = A^t \xi(0), \quad (7)$$

where A^t is the linear operator which maps tangent vector $\xi(0)$ to $\xi(t)$. The mean exponential rate of divergence of the tangent vector ξ is defined as follows:

$$\lambda(x(0), \xi(0)) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{\|\xi(t)\|}{\|\xi(0)\|}, \quad (8)$$

where $\|\cdot\|$ denotes a norm with respect to some Riemannian metric. Furthermore, there is a d -dimensional basis $\{e_i\}$ of $\xi(0)$, for which λ takes values $\lambda_i(x(0)) = \lambda(x(0), e_i)$. These can be ordered by their magnitudes $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, and are the spectrum of Lyapunov characteristic exponents. These exponents are independent of $x(0)$ if the system is ergodic.

Algorithm 1. Let $\{x_j\}$ ($j = 1, 2, \dots$) denote a time series of some physical quantity measured at the discrete time interval Δt , that is, $x_j = x(t_0 + (j - 1)\Delta t)$. Consider a small ball of radius ε centered at the orbital point x_j , and find any set of points $\{x_{k_i}\}$ ($i = 1, 2, \dots, N$) included in this ball, that is,

$$\{y^i = \{x_{k_i} - x_j \mid \|x_{k_i} - x_j\| \leq \varepsilon\}\}, \quad (9)$$

where y^i is the displacement vector between x_{k_i} and x_j . We used a usual Euclidean norm defined as follows: $\|w\| = (w_1^2 + w_2^2 + \dots + w_d^2)^{1/2}$ for some vectors $w = (w_1, w_2, \dots, w_d)$. After the evolution of a time interval $\tau = m\Delta t$, the orbital point x_j will proceed to x_{j+m} and neighboring points $\{x_{k_i}\}$ to $\{x_{k_i+m}\}$. The displacement vector $y^i = x_{k_i} - x_j$ is thereby mapped to

$$\{z^i\} = \{x_{k_i+m} - x_{j+m} \mid \|x_{k_i} - x_j\| \leq \varepsilon\}. \quad (10)$$

If the radius ε is small enough for the displacement vectors $\{y^i\}$ and $\{z^i\}$ to be regarded as good approximation

of tangent vectors in the tangent space, evolution of y^i to z^i can be represented by some matrix A_j , as follows:

$$z^i = A_j y^i. \quad (11)$$

The matrix A_j is an approximation of the flow map A^t at x_j in (7). Let us proceed to the optimal estimation of the linearized flow map A_j from the data sets $\{y^i\}$ and $\{z^i\}$. A plausible procedure for optimal estimation is the least-square-error algorithm, which minimizes the average of the squared error norm between z^i and $A_j y^i$ with respect to all components of the matrix A_j as follows:

$$\min_{A_j} S = \min_{A_j} \frac{1}{N} \sum_{i=1}^N \|z^i - A_j y^i\|^2. \quad (12)$$

Denoting the (k, l) component of matrix A_j and applying condition (12), one obtains $d \times d$ equations to solve $\partial S/\partial a_{kl}(j) = 0$. One will easily obtain the following expression for A_j :

$$A_j V = C, \quad (V_{kl}) = \frac{1}{N} \sum_{i=1}^N y^{ik} y^{il}, \quad (13)$$

$$(C_{kl}) = \frac{1}{N} \sum_{i=1}^N z^{ik} y^{il},$$

where V and C are $d \times d$ matrices, called covariance matrices, and y^{ik} and z^{ik} are the k components of vectors y^i and z^i , respectively. If $N \geq d$ and there is no degeneracy, (13) has a solution for $a_{kl}(j)$.

Now that we have the variational equation in the tangent space along the experimentally obtained orbit; the Lyapunov exponents can be computed as

$$\lambda_i = \lim_{n \rightarrow \infty} \frac{1}{n\tau} \sum_{j=1}^n \ln \|A_j e_i^j\|, \quad (14)$$

for $i = 1, 2, \dots, d$, where A_j is the solution of (13), and $\{e_i^j\}$ ($i = 1, 2, \dots, d$) is a set of basis vectors of the tangent space at x_j . In the numerical procedure, choose an arbitrary set $\{e_i^j\}$. Operate with the matrix A_j on $\{e_i^j\}$, and renormalize $A_j e_i^j$ to have length 1. Using the Gram-Schmidt procedure, maintain mutual orthogonality of the basis. Repeat this procedure for n iterations and compute (14). The advantage of the present method is now clear, since we can deal with arbitrary vectors in a tangent space and trace the evolution of these vectors. In this method, these vectors are not restricted to observed data points, in contrast with the conventional methods. The feature allows us to compute all exponents to good accuracy with great easy.

3. Results and Discussions

Signals are firstly preprocessed by Butterworth filter of order 10 with a cutoff frequency of 45 Hz to remove noise 50 Hz and high-frequency components. Filtered signals were then

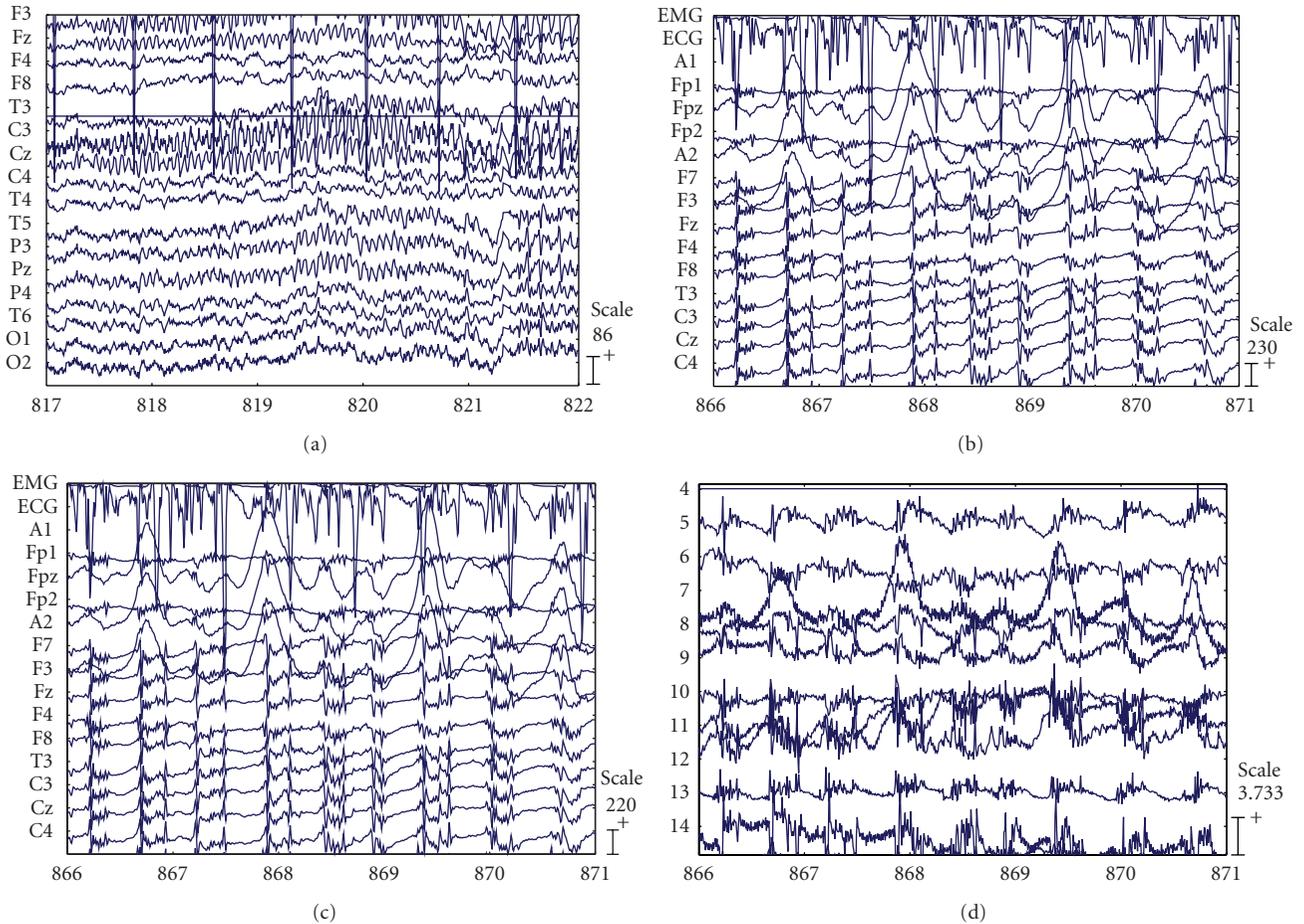


FIGURE 1: A scalp EEG recording of 21 minutes containing a general epilepsy. (a) The 5-second EEG segment at the preictal of frontal seizure was recorded by the scalp electrodes before removing noises. (b) EEG signal (5 s) during the seizure. (c) The result of (b) after being filtered by Butterworth filter of order 10 with a cutoff frequency of 45 Hz. (d) The signals obtained after applying the proposed ICA algorithm to the same segment (c).

analyzed by Independent Component Analysis (ICA) to get main components for comparison purposes. Quantifying the changes in the brain dynamics was carried out by nonlinear methods such as estimating the largest Lyapunov exponent λ_1 and the Lyapunov spectrum was also used to evaluate chaotic behavior of scalp EEG recordings.

Figures 1(a)–1(d) show the results obtained for a scalp EEG recording of 21 minutes containing a general epilepsy. In Figure 1(a), the 5-second EEG segment at the preictal of frontal seizure was recorded by the scalp electrodes before removing noises. At second 817, there are series of high-frequency, repetitive spikes, polyspike-slow waves. The pre-seizure was clearly discernible in the scalp electrodes, around second 817, and the seizure state lasted until the second 871 (Figure 1(b)). The signals are contaminated by noises and artifacts but the seizure is discernible. Figure 1(c) is result of Figure 1(b) after being filtered by Butterworth filter of order 10 with a cutoff frequency of 45 Hz. Figure 1(d) shows the signals obtained after applying the proposed ICA algorithm to the same segment in Figure 1(c). The IC4, IC9, and IC10 are sources of noise EEG while the seizure components are in remaining ICs.

Figures 2(a) and 2(b) are Lyapunov profiles over time of IC6 and IC7. Both these ICs showed that, during the seizure from second 817 to 871, the Lyapunov exponents start decreasing, and at about second 847, Lyapunov exponents drop to minimum. The seizure can easily be detected from the lowest values of Lyapunov exponent. It is period of second 817 to 871. These results are suitable to points recorded in Figure 1. Besides, Figures 2(c) and 2(d), the Lyapunov profiles of IC6 and IC7 obtained by observing the Lyapunov profiles from second 500 to second 1000, show that λ_1 starts decreasing approximately 2 mins before the onset of seizure. Therefore, the Lyapunov profiles of ICs after be analyzed Independent Component can help doctors not only to detect but also to predict early seizures for 2 minutes before the seizure occurs.

Figures 3(a) and 3(b) are the Lyapunov spectrum profiles of IC6 and IC7 of the same data. The maximum drop of Lyapunov coefficients occurs around 847, where seizure happens. It means that Lyapunov spectrum can be used to detect seizure accurately. Moreover, observing a period of 5 minutes of the pre-seizure-seizure, we can see that all the Lyapunov coefficients decrease approximately two minutes

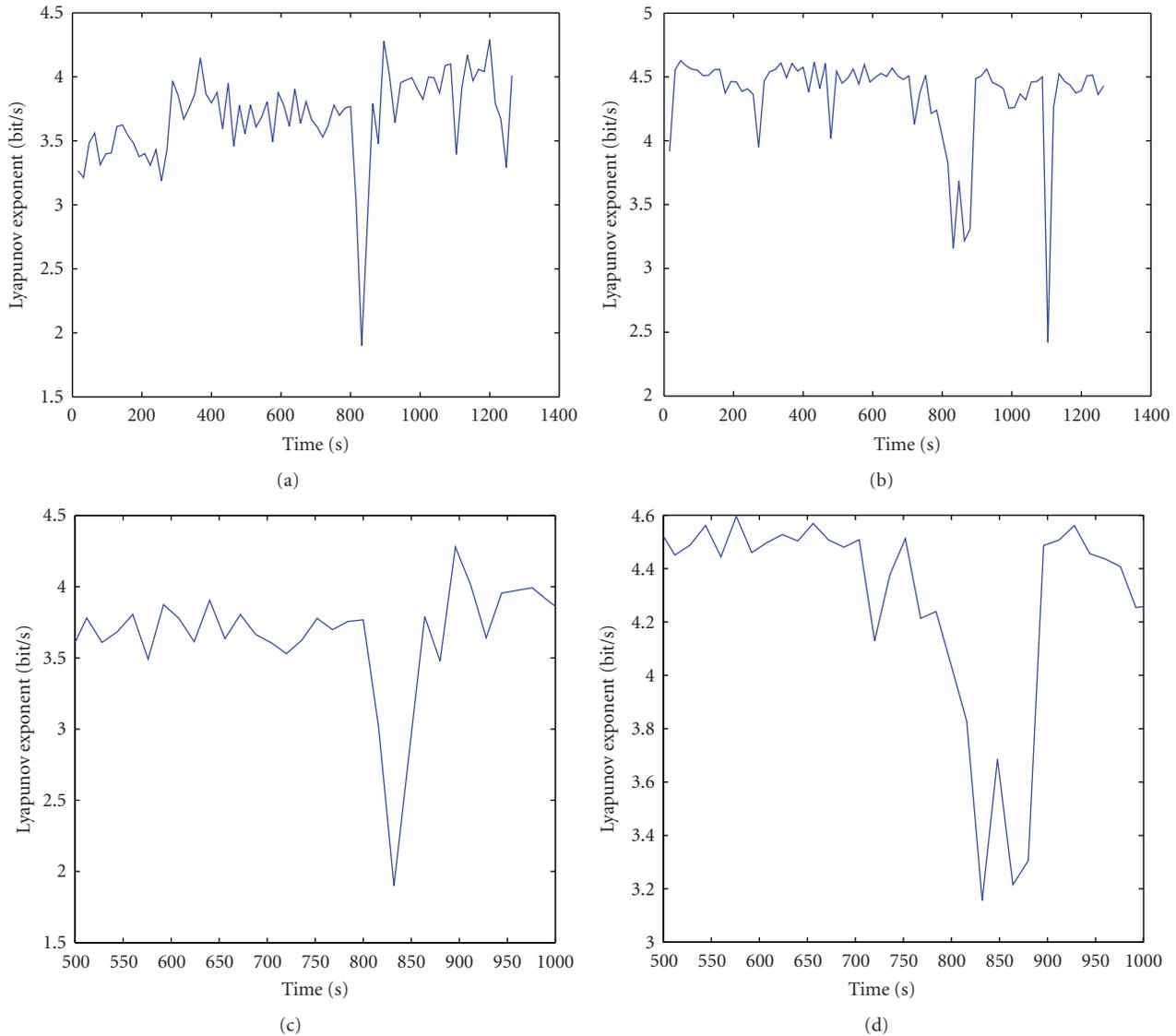


FIGURE 2: The Lyapunov exponent's profiles over time of IC6 and IC7. (a) and (b) are the largest Lyapunov exponent's profiles over time of IC6 and IC7. (c) and (d) are the largest Lyapunov exponent's profiles of IC6 and IC7 obtained by observing the Lyapunov profiles from second 500 to second 1000.

before the seizure happened. This helps the doctors to predict seizure. These results clearly show that the proposed ICA algorithm successfully separates the seizure signal (long before the seizure) from the rest of the sources, noise, and artifacts within the brain. Both the largest Lyapunov exponents and Lyapunov spectrum can be combined with ICA methods to quantify the changes in brain dynamic for diagnosing epilepsy and have brought good results.

Figures 4(a) and 4(b) are the Lyapunov profiles of the channels 8 and 11, respectively. Most channels show a minimum drop in the value of λ_1 around second 720, while pre-seizure-seizure onset interval which occurs at second 817 to second 871 has maximum peaks of the Lyapunov coefficient. Therefore, none of the channels is able to detect and predict seizure. Moreover, the scalp EEG after filtering 0.5–45 Hz was contaminated by a high-frequency activity

that causes fluctuations of for the entire recording. So, estimating only the largest Lyapunov coefficient of scalp EEG without ICA showed that mentioned features cannot detect the seizure.

The detection could be improved by examining the Lyapunov spectrum with other λ parameters. Figures 5(a) and 5(b) are Lyapunov spectrums of the channels 8 and 11 after being filtered 0–45 Hz. The Lyapunov coefficients start decreasing around second 800 and reach minimum around second 890. There is the interval in that pre-seizure and onset seizure occur. Moreover, minimum drop of λ_1 is not as clear as these of other Lyapunov coefficients. This result showed that the Lyapunov spectrum can detect seizure for noisy scalp EEG when the largest Lyapunov coefficient method cannot. This is an advantage for processing scalp EEG in practical cases in Hospital.

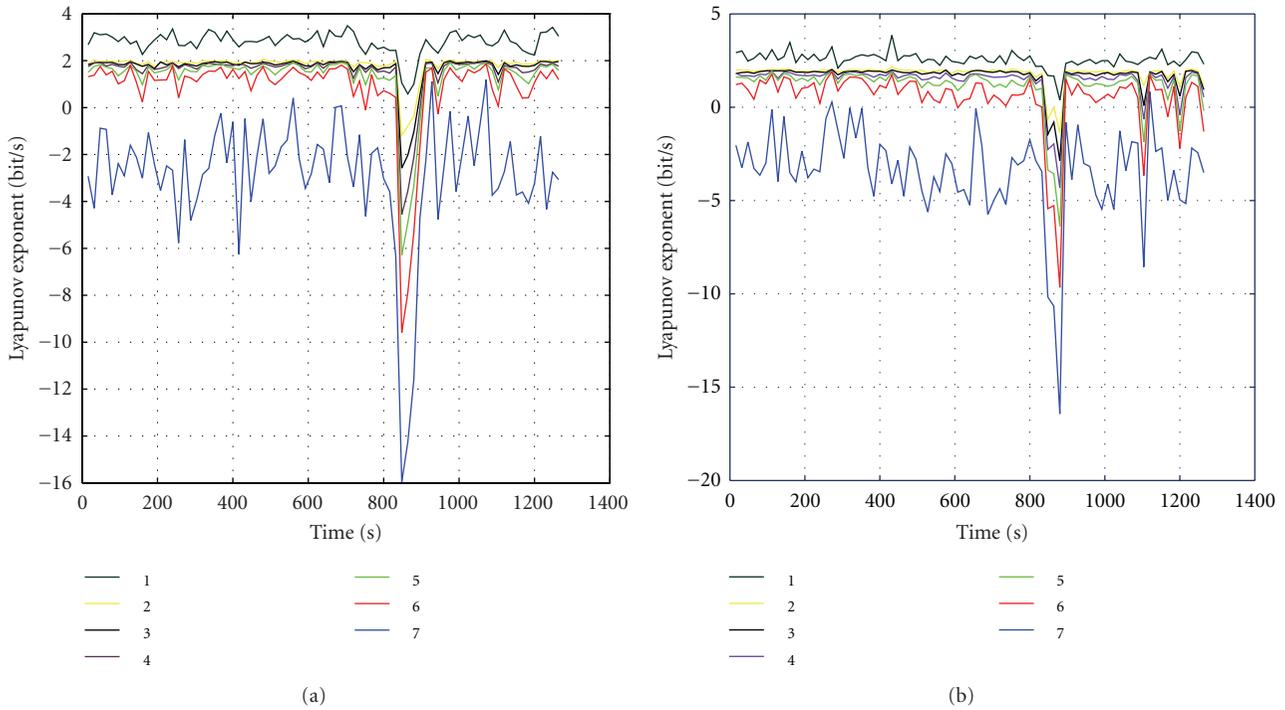


FIGURE 3: The Lyapunov spectrum profiles of IC6 and IC7 of the same data.

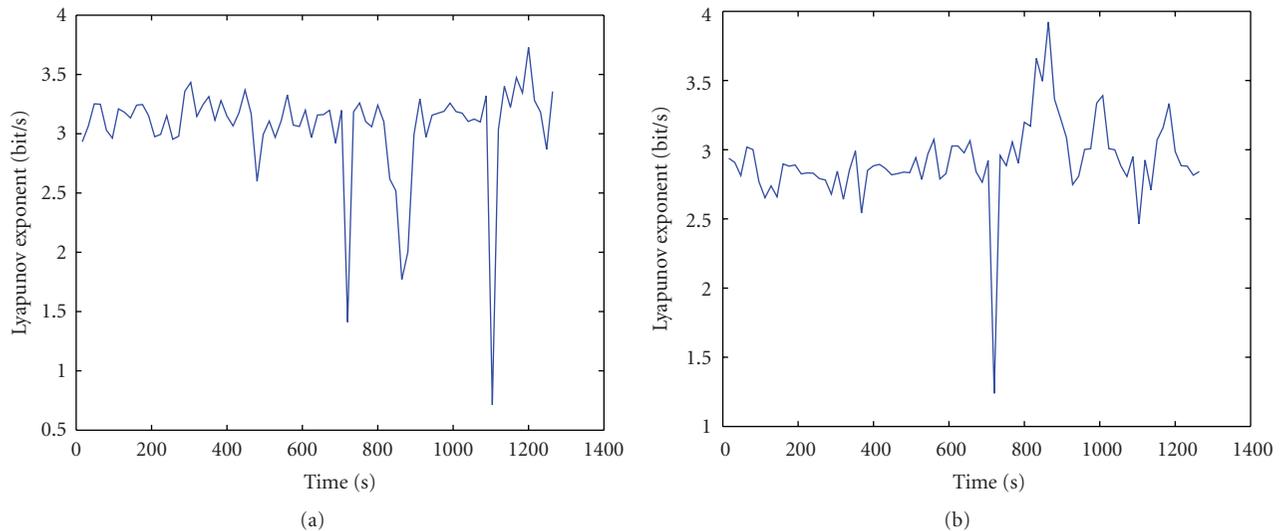


FIGURE 4: The largest Lyapunov exponent's profiles of channels 8 and 11, respectively.

Figures 6(a) and 6(b) show scalp EEG recordings of 21 minutes containing a general epilepsy. In Figure 6(a), the 5-second EEG segment at the pre-seizure of frontal seizure was recorded by the scalp electrodes before removing noises from second 679 to 684. We can see the complexity of the signal decreased and the shape of sin. Then the period of seizure occurs with signs of paroxysmal depolarization, and the waveform becomes much more complicated. Seizure ends at second 724. These signals are filtered by Butterworth filter of order 10 with a cutoff frequency of 45 Hz and then are analysed by ICA method to separate the seizure signal (long

before the seizure) from the rest of the sources, noise, and artifacts within the brain. While ICs bring seizure signs, the Lyapunov exponents are estimated.

Figures 7(a) and 7(b) illustrate the changes in the smoothed λ_1 for IC5 brings seizure signal obtained by the largest Lyapunov exponent and Lyapunov spectrum method, respectively. λ_1 starts decreasing at second 600, approximately 2 minutes before the onset of seizure, and drops minimum around second 725. The experiment results showed that ICA algorithm successfully separates the seizure signal and the combination of ICA and the Lyapunov

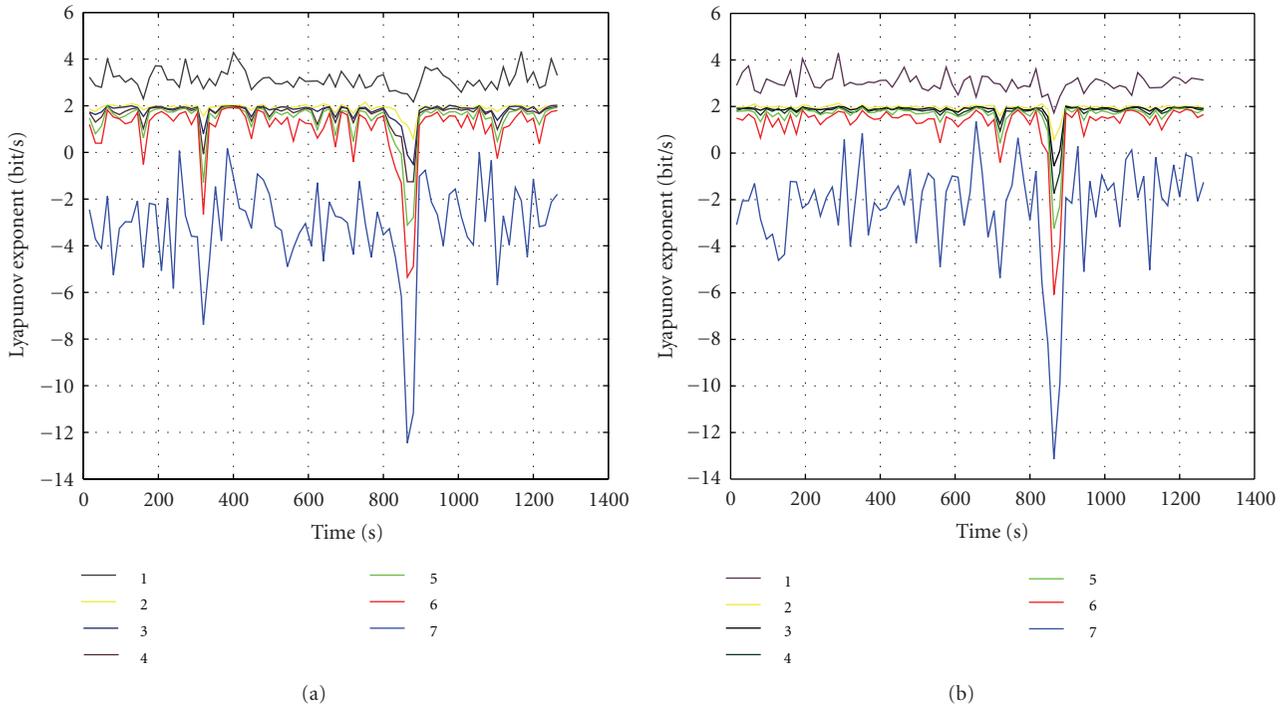


FIGURE 5: The Lyapunov spectrum of channels 8 and 11.

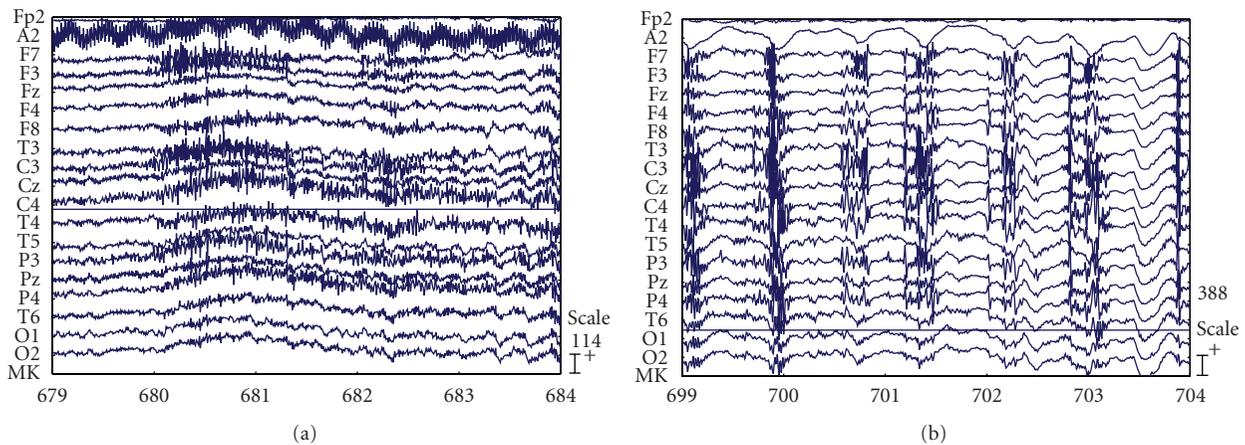


FIGURE 6: The scalp EEG recordings of 21 minutes containing a general epilepsy. (a) The 5-second EEG segment at the pre-seizure of frontal seizure. (b) EEG signal (5 s) during the seizure.

exponent method can help the doctors not only detect but also predict the epilepsy. This is an effective combination not only in removing the noises for processing the EEG signal but also quantifying the changes of brain changes as well.

Figures 8(a) and 8(b) are Lyapunov profiles of channels 9 and 10. The values λ_1 have large fluctuations that can be due to the presence of the noises and artifacts. More over, there are no clear drops of λ_1 before, in and after seizure happens. It means that the maximum Lyapunov is sensitivity to noises and it cannot detect epilepsy with quite noisy EEG. This can be caused by the description of the average information

loss of λ_1 . As mentioned previously, the detection could be improved by examining the Lyapunov spectrum with other λ parameters.

Figures 9(a) and 9(b) are Lyapunov spectrums of the channel 9 and 10 after being filtered 0–45 Hz. The Lyapunov coefficients start decreasing around second 700 and reach minimum around second 725. There is the interval in that pre-seizure and onset seizure occurs. The minimum of value λ_1 is used for detecting seizure. Moreover, values of λ_1 in both channels have peaks when time of seizure happens. This showed that estimating the spectrum of several Lyapunov

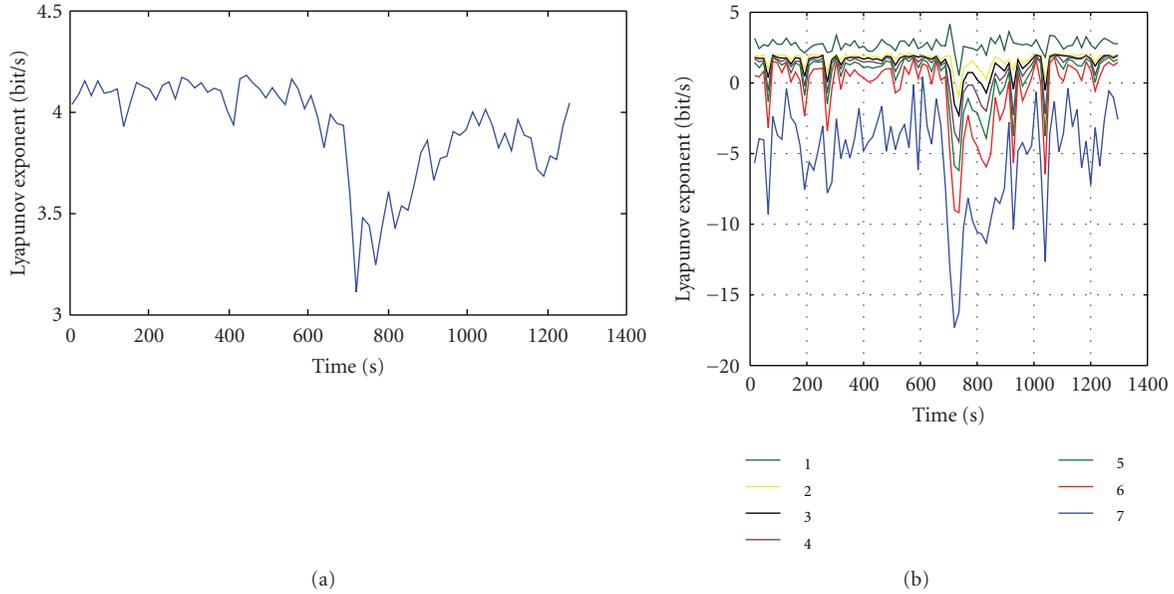


FIGURE 7: The changes in the Lyapunov exponent for IC5. (a) The smoothed λ_1 of IC5. (b) The Lyapunov spectrum of IC5.

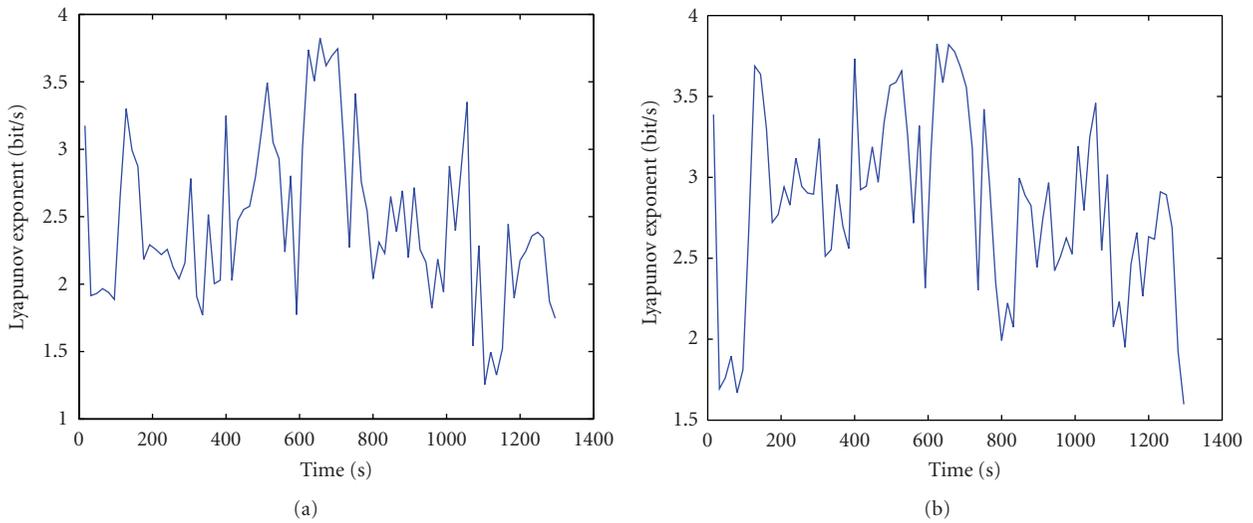


FIGURE 8: The largest Lyapunov exponent’s profiles of channels 9 and 10.

exponents (including positive, zeros, and even negative ones) is necessary for quantifying many physical quantities, especially for complicating EEG signals.

For the sets of the scalp EEG (see Table 1), 8 cases of general epilepsy were not only detected but also replaced by the combination of ICA and the Lyapunov exponent (includes the largest Lyapunov exponent and the Lyapunov spectrum) method. It means that ICA algorithm successfully separates the seizure signal within the brain. Both the largest Lyapunov exponents and Lyapunov spectrum can quantify the nonlinear changes in brain dynamic. Besides, all 8 data sets showed that the Lyapunov spectrum can detect the seizure while the largest Lyapunov exponent cannot do this for the scalp EEG without analysing ICA. This result should be an advantage for processing EEG signal.

4. Conclusions

A proposal for the estimation of Lyapunov spectrum profiles from EEG to diagnose the epilepsy has been presented. The results of the experiments clearly show that the proposal carried out advantages than the combination of ICA and the largest Lyapunov exponent method. The ICA algorithm successfully separated the seizure signal from the rest of the sources, noise, and artifacts within the brain and the largest Lyapunov exponent evaluated the chaotic behavior of the EEG signals. Lyapunov spectrum is considered as a robust and general method to process EEG signal to detect epilepsy. The results obtained for the estimated source are similar to diagnosis from medical doctors in case of typical general epilepsy.

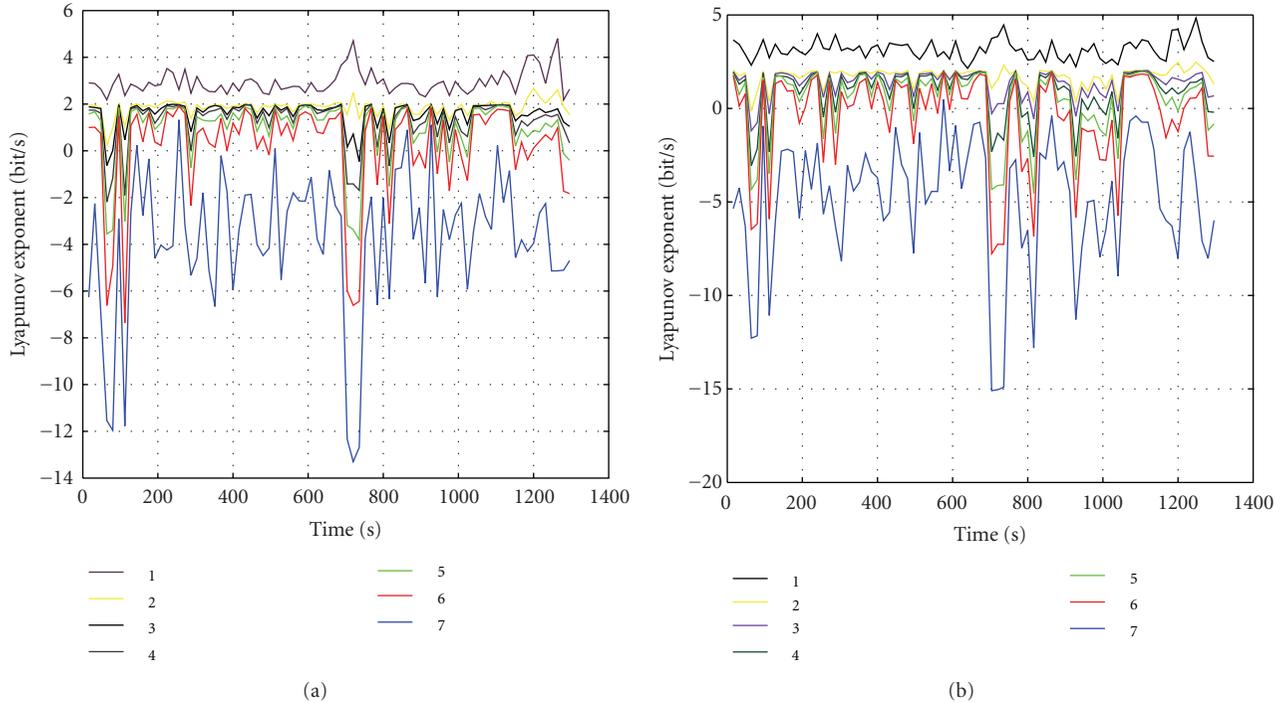


FIGURE 9: The Lyapunov spectrum of channels 9 and 10.

TABLE 1: Characteristics of the recordings (obtained in the Department of Clinical neurophysiology at Hospital 115 in Vietnam).

| Type of epilepsy | No. of patients Males/females | Age ranges | Recording length ranges (mins) | No. of electrodes |
|------------------|----------------------------------|------------|-----------------------------------|-------------------|
| General epilepsy | 7/1 | 30–45 | 20–30 | 22 |

Acknowledgments

The authors would like to thank research grant from the Department of Science and Technology, Ho Chi Minh City. Furthermore, the research was partly supported by a research fund from the Vietnam National University in Ho Chi Minh City and Vietnam National Foundation for Science and Technology Development (NAFOSTED) Research Grant No. 106.99-2010.11. They also would like to thank Dr. Cao Phi Phong, Dr. Nguyen Huu Cong, Dr. Tran Thi Mai Thy, and Dr. Nguyen Thanh Luy for their valuable advices about human physiology.

References

- [1] S. Sanei and J. A. Chambers, *EEG Signal Processing*, John Wiley & Son, New York, NY, USA, 2007.
- [2] J. Gotman and L. Y. Wang, “State-dependent spike detection: concepts and preliminary results,” *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 1, pp. 11–19, 1991.
- [3] A. A. Dingle, R. D. Jones, G. J. Carroll, and W. R. Fright, “A multistage system to detect epileptiform activity in the EEG,” *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 12, pp. 1260–1268, 1993.
- [4] J. R. Glover Jr., P. Y. Ktonas, N. Raghavan, J. M. Urnuela, S. S. Velamuri, and E. L. Reilly, “A multichannel signal processor for the detection of epileptogenic sharp transients in the EEG,” *IEEE Transactions on Biomedical Engineering*, vol. 33, no. 12, pp. 1121–1128, 1986.
- [5] J. R. Glover Jr., N. Raghavan, P. Y. Ktonas, and J. D. Frost, “Context-based automated detection of epileptogenic sharp transients in the EEG: elimination of false positives,” *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 5, pp. 519–527, 1989.
- [6] W. R. S. Webber, B. Litt, K. Wilson, and R. P. Lesser, “Practical detection of epileptiform discharges (EDs) in the EEG using an artificial neural network: a comparison of raw and parameterized EEG data,” *Electroencephalography and Clinical Neurophysiology*, vol. 91, no. 3, pp. 194–204, 1994.
- [7] C. Kurth, F. Gilliam, and B. J. Steinhoff, “EEG spike detection with a Kohonen feature map,” *Annals of Biomedical Engineering*, vol. 28, no. 11, pp. 1362–1369, 2000.
- [8] J. Zhu and D. Jiang, “A linear epileptic seizure predictor based on slow waves of scalp EEGs,” in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS ’05)*, pp. 7277–7280, September 2005.
- [9] Z. Nenadic and J. W. Burdick, “Spike detection using the continuous wavelet transform,” *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 1, pp. 74–87, 2005.
- [10] D. E. Lerner, “Monitoring changing dynamics with correlation integrals: case study of an epileptic seizure,” *Physica D*, vol. 97, no. 4, pp. 563–576, 1996.
- [11] M. Le Van Quyen, J. Martinerie, M. Baulac, and F. Varela, “Anticipating epileptic seizures in real time by a non-linear

- analysis of similarity between EEG recordings,” *NeuroReport*, vol. 10, no. 10, pp. 2149–2155, 1999.
- [12] M. Le Van Quyen, C. Adam, J. Martinerie, M. Baulac, S. Clémenceau, and F. Varela, “Spatio-temporal characterizations of non-linear changes in intracranial activities prior to human temporal lobe seizures,” *European Journal of Neuroscience*, vol. 12, no. 6, pp. 2124–2134, 2000.
- [13] M. Le Van Quyen, J. Martinerie, V. Navarro et al., “Anticipation of epileptic seizures from standard EEG recordings,” *The Lancet*, vol. 357, no. 9251, pp. 183–188, 2001.
- [14] K. K. Jerger, S. L. Weinstein, T. Sauer, and S. J. Schiff, “Multivariate linear discrimination of seizures,” *Clinical Neurophysiology*, vol. 116, no. 3, pp. 545–551, 2005.
- [15] C. C. Jouny, P. J. Franaszczuk, and G. K. Bergey, “Signal complexity and synchrony of epileptic seizures: is there an identifiable preictal period?” *Clinical Neurophysiology*, vol. 116, no. 3, pp. 552–558, 2005.
- [16] R. Esteller, J. Echauz, M. D’Alessandro et al., “Continuous energy variation during the seizure cycle: towards an on-line accumulated energy,” *Clinical Neurophysiology*, vol. 116, no. 3, pp. 517–526, 2005.
- [17] M. A. F. Harrison, M. G. Frei, and I. Osorio, “Accumulated energy revisited,” *Clinical Neurophysiology*, vol. 116, no. 3, pp. 527–531, 2005.
- [18] L. D. Iasemidis, H. P. Zaveri, J. C. Sackellares, and W. J. Williams, “Linear and nonlinear modeling of ecog in temporal lobe epilepsy,” in *Proceedings of the 5th Annual Rocky mountains Bioengineering Symposium*, pp. 187–193, 1988.
- [19] L. D. Iasemidis and J. C. Sackellares, “The temporal evolution of the largest lyapunov exponent on the human epileptic cortex,” in *Measuring Chaos in the Human Brain*, D. W. Duke and W. S. Pritchard, Eds., World Scientific, Singapore, 1991.
- [20] L. D. Iasemidis, J. C. Principe, and J. C. Sackellares, “Measurement and quantification of spatiotemporal dynamics of human epileptic seizures,” in *Nonlinear Biomedical Signal Processing*, M. Akay, Ed., pp. 296–318, IEEE Press, New York, NY, USA, 2000.
- [21] L. M. Hively, V. A. Protopopescu, and P. C. Gailey, “Timely detection of dynamical change in scalp EEG signals,” *Chaos*, vol. 10, no. 4, pp. 864–875, 2000.
- [22] L. M. Hively and V. A. Protopopescu, “Channel-consistent forewarning of epileptic events from scalp EEG,” *IEEE Transactions on Bio-Medical Engineering*, vol. 50, no. 5, pp. 584–593, 2003.
- [23] L. D. Iasemidis, J. C. Principe, J. M. Czaplewski, R. L. Gilmore, S. N. Roper, and J. C. Sackellares, “Spatiotemporal transition to epileptic seizures: a nonlinear dynamical analysis of scalp and intracranial eeg recordings,” in *Spatiotemporal Models in Biological and Artificial Systems*, F. Silva, J. C. Principe, and L. B. Almeida, Eds., pp. 81–88, IOS Press, Amsterdam, The Netherlands, 1997.
- [24] M. Ungureanu, C. Bigan, R. Strungaru, and V. Lazarescu, “Independent component analysis applied in biomedical signal processing,” *Measurement Science Review*, vol. 4, section 2, 2004.
- [25] A. Hyvärinen and E. Oja, *Independent Component Analysis: Algorithms and Applications*, Neural Networks Research Centre, 2000.
- [26] J. C. Sackellares, L. D. Iasemidis, D. S. Shiau, R. L. Gilmore, and S. N. Roper, “Detection of the preictal transition from scalp eeg recordings,” *Epilepsia*, vol. 40, supplement 7, p. 176, 1999.
- [27] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, “Determining Lyapunov exponents from a time series,” *Physica D*, vol. 16, no. 3, pp. 285–317, 1985.
- [28] J. P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, “Liapunov exponents from time series,” *Physical Review A*, vol. 34, no. 6, pp. 4971–4979, 1986.
- [29] R. Brown, P. Bryant, and H. D. I. Abarbanel, “Computing the Lyapunov spectrum of a dynamical system from an observed time series,” *Physical Review A*, vol. 43, no. 6, pp. 2787–2806, 1991.
- [30] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, “A practical method for calculating largest Lyapunov exponents from small data sets,” *Physica D*, vol. 65, no. 1-2, pp. 117–134, 1993.
- [31] S. Sato, M. Sano, and Y. Sawada, “Practical methods of measuring the generalized dimension and the largest Lyapunov exponent in high dimensional chaotic systems,” *Progress of Theoretical Physics*, vol. 77, no. 1, pp. 1–5, 1987.
- [32] M. Sano and Y. Sawada, “Measurement of the lyapunov spectrum from a chaotic time series,” *Physical Review Letters*, vol. 55, no. 10, pp. 1082–1085, 1985.
- [33] P. M. Pardalos, V. A. Yatsenko, A. Messo, A. Chinchuluun, and P. Xanthopoulos, “An optimization approach for finding a spectrum of Lyapunov exponents,” *Computational Neuroscience*, vol. 38, part 3, pp. 285–303, 2010.

Research Article

Iterative Methods for Obtaining Energy-Minimizing Parametric Snakes with Applications to Medical Imaging

Alexandru Ioan Mitrea,¹ Radu Badea,² Delia Mitrea,³ Sergiu Nedeveschi,³ Paulina Mitrea,³ Dumitru Mircea Ivan,¹ and Octavian Mircia Gurzău¹

¹ *Department of Mathematics, Technical University of Cluj-Napoca, George Baritiu Street, no. 25, 400020 Cluj-Napoca, Romania*

² *Department of Ultrasonography, University of Medicine and Pharmacy "Iuliu Hațieganu" Cluj-Napoca, Victor Babeș Street, no. 8, 400079 Cluj-Napoca, Romania*

³ *Department of Computer Science, Technical University of Cluj-Napoca, George Baritiu Street, no. 26-28, 400027 Cluj-Napoca, Romania*

Correspondence should be addressed to Delia Mitrea, delia.mitrea@cs.utcluj.ro

Received 30 September 2011; Accepted 8 November 2011

Academic Editor: Maria Crisan

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

After a brief survey on the parametric deformable models, we develop an iterative method based on the finite difference schemes in order to obtain energy-minimizing snakes. We estimate the approximation error, the residue, and the truncature error related to the corresponding algorithm, then we discuss its convergence, consistency, and stability. Some aspects regarding the prosthetic sugical methods that implement the above numerical methods are also pointed out.

1. Introduction

The deformable models represent a powerful researched model-based approach to computer-assisted medical image analysis, their applications in this framework including image segmentation, shape representation and motion tracking. The theory of deformable models is an interdisciplinary scientific domain, which has appeared and developed in the last two decades, in strong connection with practical problems of medicine, image processing, and physics. This theory joins methods, results, and techniques of various mathematical fields, physics and mechanics. The mathematical foundation of this theory represents the confluence of Functional Analysis, Approximation Theory, Differential Equations, Differential Geometry, Calculus of Variations, Numerical Analysis, Linear Algebra, and Probability Theory. The ancestors of the deformable models, in classical sense, are considered Fischler and Elschlager, with their spring-loaded templates, [1], together with Widrow [2], with its rubber mask technique [2, 3].

The theory of deformable models, in its modern form, originates from the general theory of continuous multidimensional deformable models in a Lagrangian dynamic of

Terzopoulos (1987) [4]. In fact, the deformable curves (2D models) and the deformable surfaces (3D models) gained popularity after their use in computer vision by Kaas et al. [5] and in computer graphics, by Terzopoulos and Fleischer [6] in the mid-1980s. Since then, the deformable models, known also as active contour models or snakes, have been extensively used for many applications both in 2D and 3D.

Two general types of deformable models have been developed: firstly, the parametric or variational models, which originate from the papers of Kaas et al. [5] and are based on the minimization of the energy-functional associated to the model, and secondly, the geometric models, which were introduced independently by Caselles et al. [7] and Malladi et al. [8], and are based on the front propagation theory [9].

A good survey on deformable models and their applications can be found in [10, 11]. Recent contributions on parametric deformable models have appeared in the papers [12, 13]. On the general topic of numerical methods applied in medical imaging, the recent papers [14, 15] must be mentioned.

In this paper, we deal with the deformable parametric models. The basic goal of the theory of parametric deformable models is to determine the energy-minimizing 2D or 3D models, namely, the curves or surfaces which minimize the corresponding energy functional. Two approaches will point out in order to obtain the optimal model. The first approach is based on the Euler-Lagrange-Poisson (ELP) and Euler-Gauss-Ostrogradski (EGO) equations of Calculus of Variations in order to minimize the energy-functional. The second one (the classical approach) consists of using reconstruction methods, such as the interpolation of the sparse data extracted from the image, in order to obtain a representation of the original data. In what follows we develop methods and techniques related to the first approach. Generally, the energy-functional is not convex, so it may have many local minimum. On the other hand, the analytic solution of (ELP) equation has a complicated form or it is inaccessible explicitly. Therefore, a practical and strong approach for finding local minimum of the energy functional is to construct a dynamic system that is governed by the energy functional and allow the system to evolve to the equilibrium state. Dynamic models are valuable for medical image analysis, because most anatomical structures are deformable and continually undergo nonrigid motion “in vivo.” In fact, the user is interested to find a good 2D or 3D contour in a given area. Consequently, a rough prior estimation of the 2D or 3D model is provided, then this initial model undergoes a deformation until reaching a local minimum of the energy functional. This deformation process can be achieved in one of the following ways:

- (1) in a Hamiltonian-type approach, by performing a strictly decreasing energy path, for example, via dynamic programming methods [16, 17];
- (2) in a Lagrange-type approach, by applying the mechanical principles of Lagrange [3, 18];
- (3) by using a friction force, in order to constrain the displacement of the snake [5];
- (4) by using the (ELP) evolution equation, associated to the initial (ELP) equation [19].

In this paper, we shall adopt the method of the evolution equation. So, a prior estimation of the deformable surface is provided, then it is refined step by step, based on the (EGO) equation and using discretization methods.

The paper outline is as follows. The next section is devoted to present 2D and 3D energy-minimizing models, both in their static and dynamic forms. The method for reducing the 3D problem to a 2D modeling is also pointed out, in order to minimize the computational costs of the numerical methods. The third section contains the main theoretical result of the paper. Based on finite difference schemes of explicit type, we derive an (ELP) algorithm for obtaining an energy-minimizing snake in its approximated form, then we estimate its approximation error and we discuss its consistency, convergence, and stability. The last section deals with the behaviour of prosthetic surgical methods and prosthetic medical materials, based on Software tools, which implement the iterative methods developed in the previous sections.

2. Energy-Minimizing Models

2.1. Energy-Minimizing Snakes (2D Models). From mathematical point of view, a 2D parametric deformable model (usually known as *snake*) is provided by a family \mathcal{A} of parametrized smooth curves satisfying given boundary conditions and an associated energy-functional. More exactly, denote by $C^2([0, 1], \mathbb{R}^2)$ the space of all vectorial functions $v = (x, y)^T$ so that the scalar functions $x = x(s)$ and $y = y(s)$, $0 \leq s \leq 1$ are continuous together with their derivatives up to the second-order on the standard interval $[0, 1]$, that is, $x, y \in C^2[0, 1]$; obviously, we can consider an arbitrary compact interval $[a, b]$ of the real axis instead of $[0, 1]$. The family \mathcal{A} of *admissible deformations* consists of all parametrized curves (snakes):

$$\begin{aligned} (y) : v(s) &= (x(s), y(s))^T, \\ 0 \leq s \leq 1, \quad v &\in C^2([0, 1], \mathbb{R}^2), \end{aligned} \quad (1)$$

such that the values $v(0), v(1), v'(0)$, and $v'(1)$ are given; we adopt the notation $|v|^2 = |x|^2 + |y|^2$.

In order to find the optimal position of the snake, it is necessary to characterize its state, by means of an *energy-functional*, that is associated to the class \mathcal{A} . Let us consider the following data:

- (i) the *weight-functions* $w_1(s)$ and $w_2(s)$, which control the elasticity and the rigidity of (y) , respectively; generally, these are nonnegative scalar functions of class $C^2[0, 1]$,
- (ii) the *image intensity* function $I = I(x, y)$, which is a real function of class $C^2(\mathbb{R}^2)$,
- (iii) the *potential* associated to the external forces, represented by a real function $P(v) = P(x, y)$, of class $C^2(\mathbb{R}^2)$. The simplest useful choice for the potential is $P(v) = w_3 I(v)$, where w_3 is a weight-scalar. The most used choice is $P(v) = -\lambda |\nabla I(v)|$, where $\lambda > 0$ is a given scalar and $\nabla = (\partial/\partial x, \partial/\partial y)^T$ is the Hamilton (nabla) operator; this choice will be used in this paper, too. Note that P can be defined also by $P = G_{\sigma_0} * I$, that is, the Gaussian (variance σ_0) filtered image of the input image I [10], and
- (iv) the vectorial function $k(v) = (k_1(v), k_2(v))^T$ of class $C^1(\mathbb{R}^2, \mathbb{R}^2)$ which control the local dilatation or local contraction of (y) along its normal; usually, we take $k(v) = cv$, with $c \in \mathbb{R}$.

The shape of the snake (y) subject to the image $I(v)$ is dictated by the *energy functional*:

$$E(v) = E_{\text{int}}(v) + E_{\text{ext}}(v) + E_{\text{bal}}(v), \quad (2)$$

where the terms of the right hand of (2) are defined as follows.

The *internal energy*

$$E_{\text{int}}(v) = E_{\text{cls}}(v) + E_{\text{rig}}(v) \quad (3)$$

is obtained by adding the *elastic energy*

$$E_{\text{els}}(v) = \int_0^1 \alpha(s) |v'(s)|^2 ds, \quad (4)$$

and the *rigid (bending) energy*

$$E_{\text{rig}}(v) = \int_0^1 \beta(s) |v''(s)|^2 ds. \quad (5)$$

The internal energy characterizes the deformation of a stretchy, flexible snake (contour). The values of $w_1(s)$ and $w_2(s)$ show the extent to which the snake can stretch or bend at an arbitrary point $(x(s), y(s))$ of the snake.

The *external energy*, derived from the image, is given by

$$E_{\text{ext}}(v) = \int_0^1 P(v(s)) ds = -\lambda \int_0^1 |\nabla I(x, y)|^2 ds, \quad (6)$$

and it allows to find the edges in an image so that the snake is attracted to contour with large image gradients.

The *balloon energy* is an energy of constrained-type, defined as

$$E_{\text{bal}}(v) = - \int_0^1 \det(k(v), v') ds = - \int_0^1 (k_1 y' - k_2 x') ds. \quad (7)$$

This energy can be added, optionally, by users, in order to expand (or contract) the snake.

Denoting by

$$\begin{aligned} F(s, v, v', v'') &= w_1(s) |v'^2(s)| + w_2(s) |v''^2(s)| \\ &\quad + P(v(s)) - \det(k(v), v'), \\ P(v) &= -\lambda |\nabla I(v)|^2, \end{aligned} \quad (8)$$

the following expression of energy-functional $E(v)$ is obtained from (2)–(8):

$$E(v) = \int_0^1 F(s, v, v', v'') ds. \quad (9)$$

By definition, the triple (\mathcal{A}, I, E) is said to be a *deformable 2D model (snake)*.

The basic goal of a deformable parametric model is to minimize its energy-functional $E(v)$, which leads to the energy-minimizing snake. The minimization of the snake energy gives rise to the following vectorial Euler-Lagrange-Poisson (ELP) Equation of Calculus of Variations:

$$\frac{\partial F}{\partial v} - \frac{d}{ds} \left(\frac{\partial F}{\partial v'} \right) + \frac{d^2}{ds^2} \left(\frac{\partial F}{\partial v''} \right) = 0. \quad (10)$$

Now, taking into account the relations (8) and (10), we obtain the vectorial (ELP) equation:

$$\begin{aligned} 2w_2(s)v^{iv}(s) + 4w_2'(s)v'''(s) + 2[w_2''(s) - w_1(s)]v''(s) \\ - [2w_1'(s)I_2 + \text{Tr}(\nabla k)J_2]v'(s) + \nabla P(v(s)) = 0, \end{aligned} \quad (11)$$

where $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $J_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, and $\text{Tr}(A)$ is the trace of a square matrix A .

The scalar (ELP) equations, derived from (11), have the form:

$$\begin{aligned} 2w_2x^{iv} + 4w_2'x''' + 2(w_2'' - w_1)x'' - 2w_1'x' \\ - \left(\frac{\partial k_1}{\partial x} + \frac{\partial k_2}{\partial y} \right) y' + \frac{\partial P}{\partial x} = 0, \\ 2w_2y^{iv} + 4w_2'y''' + 2(w_2'' - w_1)y'' - 2w_1'y' \\ + \left(\frac{\partial k_1}{\partial x} + \frac{\partial k_2}{\partial y} \right) x' + \frac{\partial P}{\partial y} = 0. \end{aligned} \quad (12)$$

On the other hand, we infer from (8)

$$\begin{aligned} F(s, v, v', v'') &= w_1(x'^2 + y'^2) + w_2(x''^2 + y''^2) \\ &\quad + P(x, y) - k_1 y' + k_2 x', \end{aligned} \quad (13)$$

which leads to

$$\frac{\partial^2 F}{\partial (v'')^2} = (2w_1, 2w_2)^T. \quad (14)$$

According to the Legendre conditions and the hypothesis $w_2 > 0$, the relation (14) proves that any solution of the (ELP) equations (11) or (12) provides a minimum for the energy-functional $E(v)$, namely, an energy-minimizing snake.

Example 1. If we choose in (12) $w_1 = 1$, $w_2 = 0.05$, and the boundary conditions $v(0) = v(1) = (0, 5)^T$, $v'(0) = v'(1) = (0.5, 0.5)^T$ we obtain the general solution of the (ELP) equation:

$$\begin{aligned} x(s) &= C_1 e^{3.8042s} + C_2 e^{-3.8042s} + C_3 e^{2.3511s} + C_4 e^{-2.3511s}, \\ y(s) &= C_5 e^{3.8042s} + C_6 e^{-3.8042s} + C_7 e^{2.3511s} + C_8 e^{-2.3511s}. \end{aligned} \quad (15)$$

Using boundary conditions, we obtain the graph of the curve (γ) in Figure 1.

Example 2. If we choose in (12) $w_1 = 1$, $w_2 = 1$, $k(v) = 10v$, $I(v) = I(x, y) = (2\sqrt{2}/3)(x^{3/2} + y^{3/2})$, $\lambda = 1$, and the boundary conditions $v(0) = (1, 1)^T$, $v'(0) = (3.1, -8.1)^T$, $v(\pi) = (1 - \pi/10, -2 - \pi/10 - 3 \cosh(2\pi))^T$, $v'(\pi) = (2.1 - \cosh(2\pi), -8.1 - 2 \sinh(2\pi))$, then the (ELP) equations have the form:

$$\begin{aligned} x^{iv} - x'' - 10y' - 1 &= 0, \\ y^{iv} - y'' + 10x' - 1 &= 0, \end{aligned} \quad (16)$$

with the analytical solutions:

$$\begin{aligned} x(s) &= \sin 2s + \cos 2s + \sin s \cosh 2s + 0.1s, \\ y(s) &= -4 \sin 2s + 4 \cos 2s + 3 \cos s \cosh 2s, \\ &\quad - 4 \sinh 2s \sin s - 0.1s - 6. \end{aligned} \quad (17)$$

The graph of the curve (γ) is given in Figure 2.

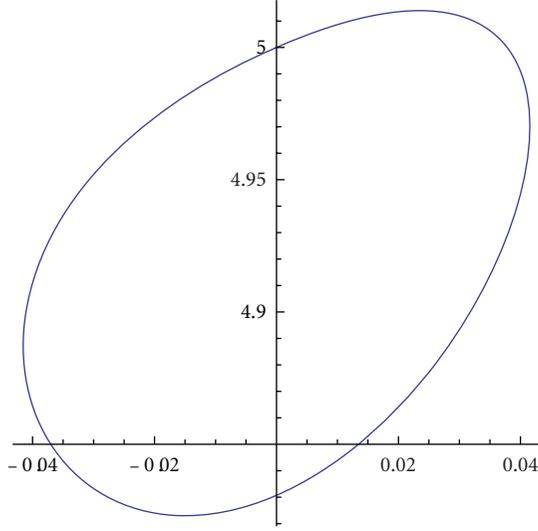


FIGURE 1

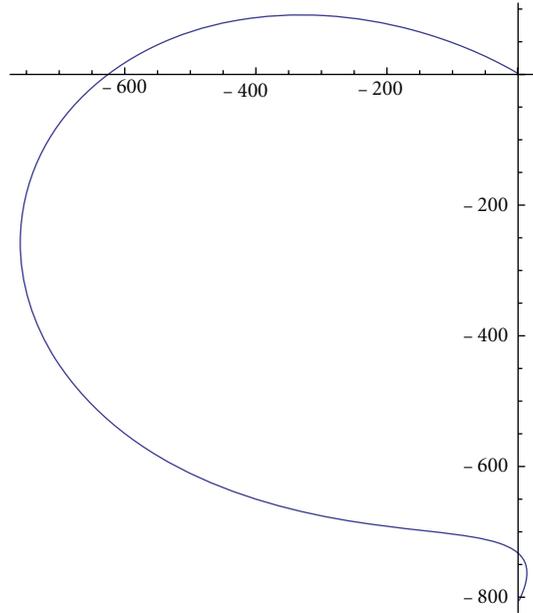


FIGURE 2

2.2. Deformable Dynamic 2D Models. Roughly speaking, the differential fourth-order vectorial equation (11) or the differential eight-order system (12) may have many solutions, which leads to many possible energy-minimizing snakes. As we have seen in the preceding examples (Section 2.1), these solutions have a complicated form; moreover, they are often inaccessible explicitly. In order to eliminate these drawbacks, we point out in this section, two approaches which lead to a practical and more simple solution of the (ELP) equation.

2.2.1. The Method of Evolution Equation [19]. Denote by

$$(\gamma) : v = v_0(s), \quad 0 \leq s \leq 1, \quad (18)$$

an initial estimate of the optimal snake and let consider a family of curves (contours)

$$(\gamma^t) : v = v(t, s), \quad v \in C^2(\mathbb{R}_+ \times [0, 1], \mathbb{R}^2), \quad (19)$$

where the parameter $t \geq 0$ describes the evolution in time of the snake and $s \in [0, 1]$ is the standard parameter of the curve. The *evolution equation* associated to the dynamic model is

$$\begin{aligned} \frac{\partial v}{\partial t} + \frac{\partial^2}{\partial s^2} \left(w_2(s) \frac{\partial^2 v}{\partial s^2} \right) - \frac{\partial}{\partial s} \left(w_1(s) \frac{\partial v}{\partial s} \right) \\ - J_2(\nabla k) \frac{\partial v}{\partial s} + \nabla P(v) - (\nabla k)(J_2 v') = 0, \end{aligned} \quad (20)$$

together with the initial condition:

$$v(0, s) = v_0(s), \quad 0 \leq s \leq 1, \quad (21)$$

and the boundary conditions

$$\begin{aligned} v(t, 0) = v_0(0), \quad v(t, 1) = v_0(1), \\ \frac{\partial v}{\partial s}(t, 0) = v'_0(0), \quad \frac{\partial v}{\partial s}(t, 1) = v'_0(1), \quad t \geq 0. \end{aligned} \quad (22)$$

A solution of the *static* problem described by (ELP) equation (11) is achieved when the solution $v(t, s)$ becomes stable with respect to the time parameter, that is, $\lim_{t \rightarrow \infty} (\partial v / \partial t)(t, s) = 0$, uniformly with respect to the parameter $s \in [0, 1]$; in this case, the evolution equation (20) provides a solution of the *static* problem (11). According to [20], we note that this approach of making the time derivative term vanish is equivalent to applying a gradient descent algorithm to find the local minimum of the energy functional $E(v)$.

2.2.2. The Method of the Lagrange Dynamics [3, 18]. A dynamic snake is represented by introducing a time-varying contour

$$v(t, s) = (x(t, s), y(t, s))^T, \quad (23)$$

see (19), with a mass density $\mu(s)$ and a damping density $\omega(s)$. The Lagrange equation for a snake defined in Section 2.1 is

$$\begin{aligned} \mu \frac{\partial^2 v}{\partial t^2} + \omega \frac{\partial v}{\partial t} + \frac{\partial^2}{\partial s^2} \left(w_2(s) \frac{\partial^2 v}{\partial s^2} \right) - \frac{\partial}{\partial s} \left(w_1(s) \frac{\partial v}{\partial s} \right) \\ + \nabla P(v) - J_2(\nabla k) \frac{\partial v}{\partial s} - (\nabla k)(J_2 v') = 0. \end{aligned} \quad (24)$$

The first two terms in the left hand side of (24) represent the inertial and damping forces, while the remaining terms, see also (11), represent the internal stretching force (the term containing $\partial v / \partial s$), the bending (rigidity) force (the term containing $\partial^2 v / \partial s^2$), the external force ($\nabla P(v)$) and the *balloon*-type force (the last two terms). Equilibrium is achieved when these forces balance and the contour comes to rest, that is,

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial t^2} = 0, \quad (25)$$

which leads to the equilibrium condition (11).

2.3. *Deformable Surfaces (3D Models)*. In this section we define briefly the notion of deformable 3D model (deformable surface), both in the static and dynamic forms, and we describe a method for reducing the problem of its optimization to a 2D modelling problem.

2.3.1. *Energy-Minimizing Surfaces*. From mathematical point of view, a 3D variational deformable model is emphasized by a family \mathcal{A} of parameterized smooth surfaces with given boundary conditions, named *admissible surfaces*, and an associated *energy functional*.

Denoting by $D = [0, 1] \times [0, 1]$ the unit square of \mathbb{R}^2 , let us consider a surface of vectorial equation:

$$(S) : v = v(s, r), \quad (s, r) \in D, \quad (26)$$

where $v \in C^2(D, \mathbb{R}^3)$, $v = (x, y, z)^T$; in this subsection we set $|v|^2 = x^2 + y^2 + z^2$, $v_s = \partial v / \partial s$, $v_{ss} = \partial^2 v / \partial s^2$, $v_{sr} = \partial^2 v / \partial s \partial r$, $v_{rr} = \partial^2 v / \partial r^2$. Given the functions $g \in C^2(\partial D, \mathbb{R}^3)$ and $h \in C^1(\partial D, \mathbb{R}^3)$, where ∂D is the boundary of D , let \mathcal{A} be the set of *admissible deformations*, which consists of all functions $v \in C^2(D, \mathbb{R}^3)$ satisfying the boundary conditions $v(s, r) = g(s, r)$ and $(\partial v / \partial n)(s, r) = h(s, r)$ on ∂D , where n is the normal vector with respect to the surface (S) defined by (26). Further, let us consider the following functions: *the image intensity function* $I \in C^2(\mathbb{R}^3)$; *the potential function associated to the external forces* $P(v) = -\lambda |\nabla I(v)|^2$, $\lambda > 0$; *the control functions* corresponding to the internal forces acting on the shape of the surface, namely, the elasticity functions $w_{10}(s; r)$ and $w_{01}(s; r)$; the *rigidity functions* $w_{20}(s; r)$ and $w_{02}(s; r)$, and *the twist resistance function* $w_{11}(s; r)$. The energy functional $E : \mathcal{A} \rightarrow \mathbb{R}$, associated to these data, is defined as follows:

$$E(v) = \iint_D F(v, v_s, v_r, v_{ss}, v_{sr}, v_{rr}) ds dr, \quad (27)$$

where

$$\begin{aligned} F(v, v_s, v_r, v_{ss}, v_{sr}, v_{rr}) &= w_{10} |v_s|^2 + w_{01} |v_r|^2 \\ &\quad + w_{20} |v_{ss}|^2 + 2w_{11} |v_{sr}|^2 \\ &\quad + w_{02} |v_{rr}|^2 + f(v, v_s, v_r), \end{aligned} \quad (28)$$

$$f(v, v_s, v_r) = P(v) + \det(c_0 v, v_s, v_r).$$

We notice that $E(v)$ represents the sum of *the internal energy* (the terms of (27) excepting $f(v, v_s, v_r)$), *the external energy* (defined by the term containing $P(v)$) and *the balloon energy*, which is added, optionally, by the users (the term including $\det(c_0 v, v_s, v_r)$).

The triple (\mathcal{A}, I, E) is said to be a *3D deformable model*, sometimes a *deformable surface*. The basic problem of the deformable model is to minimize its energy functional, namely, to obtain the optimal deformable surface. To this purpose, the Euler-Gauss-Ostrogradski (EGO) equation of Calculus of Variations, that is,

$$\begin{aligned} \frac{\partial F}{\partial v} - \frac{\partial}{\partial s} \left(\frac{\partial F}{\partial v_s} \right) - \frac{\partial}{\partial r} \left(\frac{\partial F}{\partial v_r} \right) + \frac{\partial^2}{\partial s^2} \left(\frac{\partial F}{\partial v_{ss}} \right) \\ + \frac{\partial^2}{\partial v \partial r} \left(\frac{\partial F}{\partial v_{sr}} \right) + \frac{\partial^2}{\partial r^2} \left(\frac{\partial F}{\partial v_{rr}} \right) = 0 \end{aligned} \quad (29)$$

is used.

By simple calculation, we obtain from (28) and (29):

$$\begin{aligned} \frac{\partial^2}{\partial s^2} (w_{20} v_{ss}) + \frac{\partial^2}{\partial r^2} (w_{02} v_{rr}) + 2 \frac{\partial^2}{\partial s \partial r} (w_{11} v_{sr}) \\ - \frac{\partial}{\partial s} (w_{10} v_s) - \frac{\partial}{\partial r} (w_{01} v_r) \\ + \frac{1}{2} \left(\nabla f - \frac{\partial}{\partial s} \left(\frac{\partial f}{\partial v_s} \right) - \frac{\partial}{\partial r} \left(\frac{\partial f}{\partial v_r} \right) \right) = 0. \end{aligned} \quad (30)$$

2.3.2. *Deformable Dynamic 3D Models*. Similarly to the 2D model, we can suppose that a rough prior estimate of surface is accessible, namely,

$$(S^0) : v = v_0(s, r), \quad (s, r) \in D. \quad (31)$$

Further, this surface is refined step by step, according to (EGO) equation; so, a sequence of surfaces, which leads to the energy-minimizing surface, is provided. More exactly, let

$$(S^t) : v = v(t, s, r), \quad t \geq 0, \quad (s, r) \in D, \quad (32)$$

be a family of surfaces, where the parameter t describes the evolution in time of the model. We associate to the previous static model (\mathcal{A}, I, E) the *evolution equation*

$$\frac{\partial v}{\partial t} + G(v, v_s, v_r, v_{ss}, v_{sr}, v_{rr}) = 0, \quad (33)$$

where $G(v, v_s, v_r, v_{ss}, v_{sr}, v_{rr})$ is the left hand member of (30), together with the *initial estimate (condition)*

$$v(0, s, r) = v_0(s, r), \quad (s, r) \in D, \quad (34)$$

and the boundary dynamic conditions

$$\begin{aligned} v(t, s, r) &= v_0(s, r), \quad (s, r) \in \partial D, \quad t \geq 0, \\ \frac{\partial v(t, s, r)}{\partial n} &= \frac{\partial v_0(s, r)}{\partial n}, \quad (s, r) \in \partial D, \quad t \geq 0. \end{aligned} \quad (35)$$

A solution of the “static” problem described by (30) is achieved, when the solution $v(t, s, r)$ becomes stable with respect to the time parameter, that is, $\lim_{t \rightarrow \infty} (\partial v / \partial t)(t, s, r) = 0$, uniformly, with respect to $(s, r) \in D$; in this case, the evolution equation (33) provides a solution of the static problem (30).

2.3.3. *The Simplified 2D Model*. The problem of finding directly energy-minimizing surfaces, that is, solutions of the p.d.e. (30), is not practically possible because these solutions contain long and complicated expressions or their explicit form is inaccessible. On the other hand, by using discretized schemes for solving (33), we get a system of algebraic equations with a high computational level. These drawbacks are eliminated by passing to a 2D modeling problem, [19]. More exactly, the third component z of (S) is constrained to depend only on r , by setting $z(s, r) = r$. So, the surface that we seek is given as a sequence of plane curves, named *slices*, and the parameter r of (26) becomes the index of the corresponding slice. In this approach, the surface that we seek is viewed as a sequence of the planar curves (slices), indexed by

the parameter r , so that each fixed value of r provides a closed curve, lying in a slice of the 3D-image. Consequently, let

$$(\gamma_r) : v(s) = (x(s), y(s)), \quad s \in [0, 1], \quad (36)$$

be the 2D curve obtained by applying this reconstruction method, for a given r .

Under the hypothesis that w_{ij} are positive constants, the (EGO) equation (29), which corresponds to (γ_r) , is

$$2w_{20} \frac{d^4 v}{ds^4} - 2w_{10} \frac{d^2 v}{ds^2} - c_0 J_2 \frac{dv}{ds} + \nabla P = 0, \quad (37)$$

where $J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

If we consider in (37) $c_0 = -0.02$, $w_{10} = 2.5$, $w_{20} = 0.4$, $P = r(x^2 + y^2)$ and $r = 0.1, 0.2, \dots, 1$ with boundary conditions $x(0) = x(1) = 1 + (r^2(1-r)^2)/25$, $x'(0) = x'(1) = r(1-r)/20$, $y(0) = y(1) = 0 + r^2(1-r)/25$, $y'(0) = y'(1) = 2r(1-r)/5$, we obtain the graphs of the slices and a 3D reconstruction of the surface, as we can see in Figures 3(a) and 3(b).

In what follows we shall restrict to the study of 2D deformable models.

3. An ELP-Algorithm for Obtaining Energy-Minimizing Snakes

In this section we suppose that the following hypotheses are satisfied: the control functions w_1 and w_2 are positive constants, the curves of the family (γ^t) given by (18) and (19) are closed for every $t \geq 0$ and $k(v) = c_0 v$, $c_0 \in \mathbb{R}_+$. Thus, the (ELP) evolution equation (20) becomes

$$2 \frac{\partial v}{\partial t} + 2w_2 \frac{\partial^4 v}{\partial s^4} - 2w_1 \frac{\partial^2 v}{\partial s^2} - 2c_0 J_2 \frac{\partial v}{\partial s} + \nabla P = 0, \quad v = v(t, s). \quad (38)$$

In order to solve numerically the partial differential equation (38), we focus on the method of finite differences, which is widely used in image processing [21]. Let δ and h be the time and the space discretization steps, respectively, and denote by $\mathcal{R} = \{(t_k, s_i), k \geq 0, 0 \leq i \leq N\}$ the plane net of discretization, with $N \in \mathbb{N}^*$, $Nh = 1$, $t_k = k\delta$, and $s_i = ih$. The following notations will be used, too: $v_i^k = v(t_k, s_i)$, $v^k = (x^k, y^k)^T$, $k \geq 0$; $g^k = (g_1^k, g_2^k)^T$, with $g_1^k = (-1/2)((\partial P/\partial x)(v^k))$, $g_2^k = (-1/2)((\partial P/\partial y)(v^k))$; obviously, $v_i^k = v_{i+N}^k$, $i \in \mathbb{Z}$, because (γ^t) , $t \geq 0$, is a closed curve. Also, we set

$$\alpha = \frac{w_1}{h^2}, \quad \beta = \frac{w_2}{h^4}, \quad \gamma = \frac{c_0}{h}. \quad (39)$$

3.1. Explicit Finite Difference Scheme. We approximate the partial derivatives involved in the (ELP) evolution equation (38) as follows:

$$\begin{aligned} \frac{\partial v}{\partial t}(t_k, s_i) &\approx \frac{1}{\delta} (v_i^{k+1} - v_i^k); & \frac{\partial v}{\partial s}(t_k, s_i) &\approx \frac{1}{h} (v_{i+1}^k - v_i^k), \\ \frac{\partial^2 v}{\partial s^2}(t_k, s_i) &\approx \frac{1}{h^2} (v_{i+1}^k - 2v_i^k + v_{i-1}^k), \\ \frac{\partial^4 v}{\partial s^4}(t_k, s_i) &\approx \frac{1}{h^4} (v_{i+2}^k - 4v_{i+1}^k + 6v_i^k - 4v_{i-1}^k + v_{i-2}^k). \end{aligned} \quad (40)$$

By replacing the relations (40) in the partial differential equation (38), it result a system of algebraic equations; denoting by $V^k = (X^k, Y^k)^T$ the solutions of this system (which approximate the exact values v_i^k of (38) at the nodes of \mathcal{R}), we get the vectorial formula:

$$\begin{aligned} \frac{V_i^{k+1} - V_i^k}{\delta} + \beta (V_{i+2}^k - 4V_{i+1}^k + 6V_i^k - 4V_{i-1}^k + V_{i-2}^k) \\ - \alpha (V_{i+1}^k - 2V_i^k + V_{i-1}^k) \\ - \gamma J_2 (V_{i+1}^k - V_i^k) + \frac{1}{2} \nabla P = 0, \quad 0 \leq i \leq N; k \geq 0, \end{aligned} \quad (41)$$

where

$$V_i^k = (X_i^k, Y_i^k) \quad (42)$$

and α, β, γ are given by (39). The scalar equations corresponding to (41) are the following:

$$\begin{aligned} \frac{X_i^{k+1} - X_i^k}{\delta} + \beta (X_{i+2}^k - 4X_{i+1}^k + 6X_i^k - 4X_{i-1}^k + X_{i-2}^k) \\ - \alpha (X_{i+1}^k - 2X_i^k + X_{i-1}^k) \\ - \gamma (Y_{i+1}^k - Y_i^k) \\ + \frac{1}{2} \frac{\partial P}{\partial x} (X_i^k, Y_i^k) = 0, \quad 0 \leq i \leq N-1; k \geq 0. \\ \frac{Y_i^{k+1} - Y_i^k}{\delta} + \beta (Y_{i+2}^k - 4Y_{i+1}^k + 6Y_i^k - 4Y_{i-1}^k + Y_{i-2}^k) \\ - \alpha (Y_{i+1}^k - 2Y_i^k + Y_{i-1}^k) \\ + \gamma (X_{i+1}^k - X_i^k) \\ + \frac{1}{2} \frac{\partial P}{\partial y} (X_i^k, Y_i^k) = 0, \quad 0 \leq i \leq N-1; k \geq 0. \end{aligned} \quad (43)$$

Now, let K be the *stiffness matrix* associated to the explicit finite difference scheme, defined as the circular matrix of order N , whose first row is

$$(a_1, a_2, a_3, 0, \dots, 0, a_3, a_2), \quad (44)$$

where

$$a_1 = 2\alpha + 6\beta, \quad a_2 = -\alpha - 4\beta, \quad a_3 = \beta. \quad (45)$$

Denote by L the circular (square) matrix of order N defined by the first row $(1, -1, 0, 0, \dots, 0)$ and let I_N be the identity matrix of order N . The relations (41) and (43) can be written in a matricial form as:

$$V^{k+1} = (I_N - \delta K) V^k - \gamma \delta L (J_2 V^k) + \delta g^k, \quad k \geq 0, \quad (46)$$

$$\begin{aligned} X^{k+1} &= (I_N - \delta K) X^k - \gamma \delta L Y^k + \delta g_1^k, \\ Y^{k+1} &= (I_N - \delta K) Y^k + \gamma \delta L X^k + \delta g_2^k, \end{aligned} \quad k \geq 0 \quad (47)$$

respectively.

In what follows, the formulas (41)–(47) will be referred as (ELP) *algorithm* for obtaining an energy minimizing snake (in its approximating form).

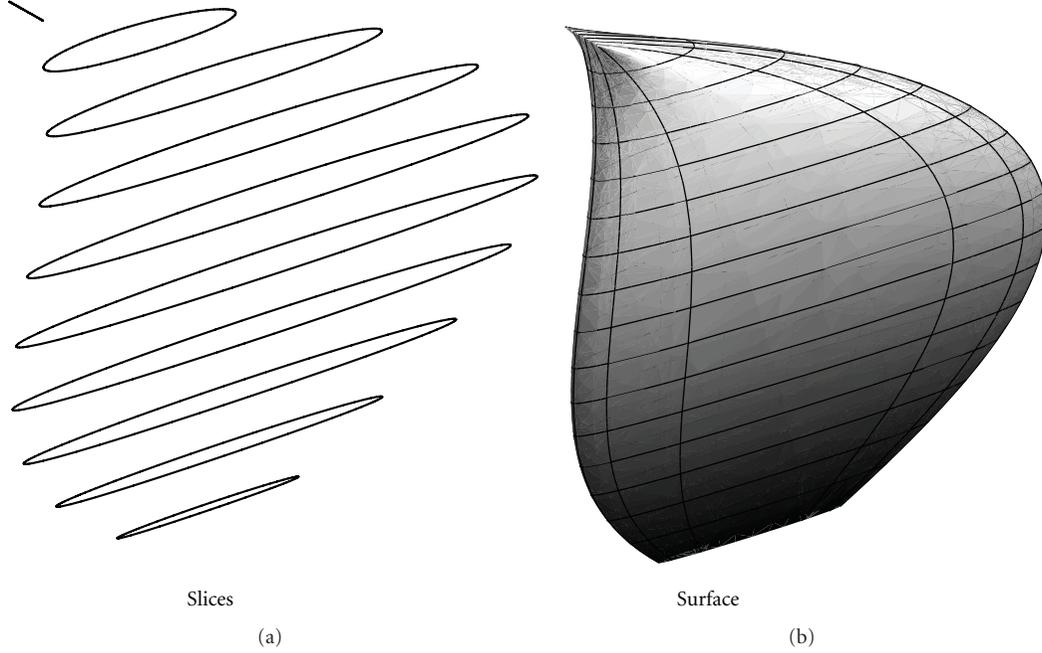


FIGURE 3

3.2. *The Residue of (ELP) algorithm.* Taking into account the relation (41), the residue associated to the (ELP) algorithm is

$$\begin{aligned}
 Rv_i = & \frac{v_i^{k+1} - v_i^k}{\delta} + \beta(v_{i+2}^k - 4v_{i+1}^k + 6v_i^k - 4v_{i-1}^k + v_{i-2}^k) \\
 & - \alpha(v_{i+1}^k - 2v_i^k + v_{i-1}^k) - \gamma J_2(v_{i+1}^k - v_i^k) \\
 & + \frac{1}{2} \nabla P(v_i^k), \quad 0 \leq i \leq N, \quad k \geq 0.
 \end{aligned} \tag{48}$$

By using Taylor expansions at the point $(t_k, s_i) \in \mathcal{R}$ we obtain

$$\begin{aligned}
 v_i^{k+1} &= v_i^k + \delta \frac{\partial v}{\partial t} + \frac{\delta^2}{2!} \frac{\partial^2 v}{\partial t^2} + \frac{\delta^3}{3!} \frac{\partial^3 v}{\partial t^3} + \dots, \\
 v_{i\pm 1}^k &= v_i^k \pm h \frac{\partial v}{\partial s} + \frac{h^2}{2!} \frac{\partial^2 v}{\partial s^2} \pm \frac{h^3}{3!} \frac{\partial^3 v}{\partial s^3} + \dots, \\
 v_{i\pm 2}^k &= v_i^k \pm 2h \frac{\partial v}{\partial s} + \frac{(2h)^2}{2!} \frac{\partial^2 v}{\partial s^2} \pm \frac{(2h)^3}{3!} \frac{\partial^3 v}{\partial s^3} + \dots,
 \end{aligned} \tag{49}$$

where the partial derivatives $\partial v / \partial t$ and $\partial^l v / \partial s^l$, $l \geq 1$ are computed at the point $(t_k, s_i) = (k\delta, ih) \in \mathcal{R}$.

By replacing the expansions (49) in the residue's formula (48) and using the relations (39), we derive

$$\begin{aligned}
 Rv_i = & \delta \left(\frac{1}{2} \frac{\partial^2 v}{\partial t^2} + \frac{\delta}{6} \frac{\partial^3 v}{\partial t^3} + \dots \right) (t_k, s_i) \\
 & + h^2 w_2 \left(\frac{1}{6} \frac{\partial^6 v}{\partial s^6} + \frac{127}{5040} h^2 \frac{\partial^8 v}{\partial s^8} + \dots \right) (t_k, s_i) \\
 & - w_1 h^2 \left(\frac{1}{12} \frac{\partial^4 v}{\partial s^4} + \frac{h^2}{60} \frac{\partial^6 v}{\partial s^6} + \dots \right) (t_k, s_i) \\
 & - c_0 J_2 h \left(\frac{1}{2} \frac{\partial^2 v}{\partial v^2} + \frac{h^2}{24} \frac{\partial^4 v}{\partial v^4} + \dots \right) (t_k, s_i), \\
 & 0 \leq i \leq N-1, \quad k \geq 0.
 \end{aligned} \tag{50}$$

If the partial derivatives of the vectorial function v are uniformly bounded on D , the relations (50) give the following estimate concerning the residue of (ELP) algorithm:

$$Rv_i = \begin{cases} O(\delta) + O(h), & \text{if } c_0 > 0, \\ O(\delta) + O(h^2), & \text{if } c_0 = 0. \end{cases} \tag{51}$$

Notice that the condition $c_0 = 0$ means that there are not existing constrains defined by the users.

3.3. *The Consistency of the ELP algorithm.* Let $\text{Tr}(v_i) = \delta Rv_i$ be the truncature error of (ELP) algorithm at the k th iteration. Under the assumption of uniform boundedness of the partial derivatives of the vectorial function v , it follows from (51):

$$\text{Tr}(v_i) = \begin{cases} O(\delta^2) + O(\delta h), & \text{if } c_0 > 0 \\ O(\delta^2) + O(\delta h^2), & \text{if } c_0 = 0. \end{cases} \tag{52}$$

The relations (52) characterize the accuracy of the discretized scheme providing the (ELP)-Algorithm.

On the other hand, the equality

$$\lim_{\substack{\delta \rightarrow 0 \\ h \rightarrow 0}} \frac{\text{Tr}(v_i)}{\delta} = \lim_{\substack{\delta \rightarrow 0 \\ h \rightarrow 0}} Rv_i = 0, \tag{53}$$

which results from (52), shows that this discretized scheme is consistent.

3.4. *Approximation Error and the Convergence.* Let us consider the approximation-error ε_i^k at the point $(t_k, s_i) \in \mathcal{R}$, namely

$$\varepsilon_i^k = v_i^k - V_i^k, \quad 0 \leq i \leq N-1, \quad k \geq 0. \tag{54}$$

By replacing $V_i^k = v_i^k - \varepsilon_i^k$ from (54) into (41) and taking into account the expressions (49) and the definition (48) of Rv_i , we get

$$\begin{aligned} \varepsilon_i^{k+1} &= \delta Rv_i - \beta\delta\varepsilon_{i+2}^k + \delta((4\alpha + \beta)I_2 + \gamma J_2)\varepsilon_{i+1}^k \\ &\quad + ((1 - 6\beta\delta - 2\alpha\delta)I_2 - \gamma J_2)\varepsilon_i^k \\ &\quad + \delta(4\beta + \alpha)\varepsilon_{i-1}^k - \beta\delta\varepsilon_{i-2}^k. \end{aligned} \quad (55)$$

Let

$$E^k = \max\left\{\left|\varepsilon_{i-2}^k\right|, \left|\varepsilon_{i-1}^k\right|, \left|\varepsilon_i^k\right|, \left|\varepsilon_{i+1}^k\right|, \left|\varepsilon_{i+2}^k\right|\right\}, \quad k \geq 0, \quad (56)$$

be the approximation error of (ELP) algorithm at k th iteration.

The relations (55) and (56) yield:

$$\begin{aligned} E^{k+1} &\leq \delta|Rv_i| \\ &\quad + \left(\beta\delta + \delta\sqrt{(4\beta + \alpha)^2 + \gamma^2}\right. \\ &\quad \left. + \sqrt{(1 - 6\beta\delta - 2\alpha\delta)^2 + \gamma^2\delta^2 + 4\beta\delta + \alpha\delta + \beta\delta}\right)E^k. \end{aligned} \quad (57)$$

On the other hand, it follows from (50):

$$|Rv_i| \leq M_1\delta + |2w_2 - w_1|M_2h^2 + c_0M_3h, \quad 0 \leq i \leq N - 1, \quad (58)$$

where M_j , $j \geq 1$ are positive constants, which do not depend on δ and h .

Now, the relations (57) and (58), combined with the classic inequality

$$\sqrt{x^2 + y^2} \leq |x| + |y|, \quad (59)$$

provide the estimate:

$$E^{k+1} \leq (10\beta\delta + 2\alpha\delta + 2\gamma\delta + |1 - 6\beta\delta - 2\alpha\delta|)E^k + A(h, \delta), \quad (60)$$

with

$$A(h, \delta) = M_1\delta^2 + M_2|2w_2 - w_1|\delta h^2 + M_3c_0\delta h. \quad (61)$$

Denote by

$$\varepsilon = \frac{\delta}{h^4} \quad (62)$$

and let us assume that the inequality

$$6\varepsilon w_2(k + 1) \leq 1 \quad (63)$$

holds. It is a simple exercise to show that the relation (63) entails the inequality

$$6\beta\delta + 2\alpha\delta \leq 1 \quad (64)$$

for N sufficiently large. Now, the relations (60) and (64) lead to:

$$E^{k+1} \leq qE^k + A(h, \delta), \quad k \geq 0, \quad (65)$$

where

$$E^0 = 0, \quad q = 1 + 4\beta\delta + 2\gamma\delta. \quad (66)$$

Writing (65) successively for $k, k - 1, \dots, 1$, we get

$$E^{k+1} \leq \frac{q^{k+1} - 1}{q - 1}A(h, \delta), \quad k \geq 0. \quad (67)$$

Taking into account that $\gamma \leq \beta$ (for N sufficiently large), the relations (66) and (39) imply $1 + 4w_2\varepsilon \leq q \leq 1 + 6w_2\varepsilon$, so that the relations (67) and (63), combined with the inequality $(1 + x)^{1/x} \leq e$, $x > 0$, yield

$$E^{k+1} \leq \frac{e - 1}{4w_2\varepsilon}A(h, \delta). \quad (68)$$

Finally, we derive from (61), (62), and (68):

$$E^{k+1} = \begin{cases} O(\delta h^4) + O(h^5), & \text{if } c_0 > 0, \\ O(\delta h^4) + O(h^6), & \text{if } c_0 = 0. \end{cases} \quad (69)$$

It follows from (69) that $E^{k+1} \rightarrow 0$ if $h \rightarrow 0$; it is easily seen that, according to the relations (62) and (63), the hypothesis $h \rightarrow 0$ implies $\delta \rightarrow 0$; consequently the following result holds.

If the inequality (63) fulfills, then the (ELP) algorithm (46) is convergent and its approximation error at the $(k + 1)$ th iteration is given by the relation (69).

3.5. The Stability. The intuitive idea regarding the stability is that small errors in the initial conditions of a partial differential equation should cause small errors in its solution. In fact, the study of the stability is useful in connection with the theorem of Lax concerning the convergence of the discretized schemes, [21].

The aim of this subsection is to examine the stability of the (ELP) algorithm (46), with $c_0 = 0$. By omitting the small terms δRv_i of (55), we get the relation:

$$\begin{aligned} \varepsilon_i^{k+1} &= (1 - 6\beta\delta - 2\alpha\delta)\varepsilon_i^k + (\alpha\delta + 4\beta\delta)(\varepsilon_{i+1}^k + \varepsilon_{i-1}^k) \\ &\quad - \beta\delta(\varepsilon_{i+2}^k + \varepsilon_{i-2}^k), \quad k \geq 0. \end{aligned} \quad (70)$$

To apply the *stability criterion of von Neumann*, [22] we set

$$\begin{aligned} r_1 &= \alpha\delta, & r_2 &= \beta\delta, \\ \eta_1 &= \omega_1 h, & \eta_2 &= \omega_2 h, \end{aligned} \quad (71)$$

$$\begin{aligned} \varepsilon_i^k &= \exp(-kl) \exp(j\omega h) \\ &= \left(\mu^k e^{j\omega_1 ih}, \mu^k e^{j\omega_2 ih}\right)^T, \\ \mu &= \exp(-l), \end{aligned} \quad (72)$$

where j and $\omega = (\omega_1, \omega_2)^T$ are complex numbers, $j^2 = -1$ and $\omega = (\omega_1, \omega_2)^T$ denotes the frequency.

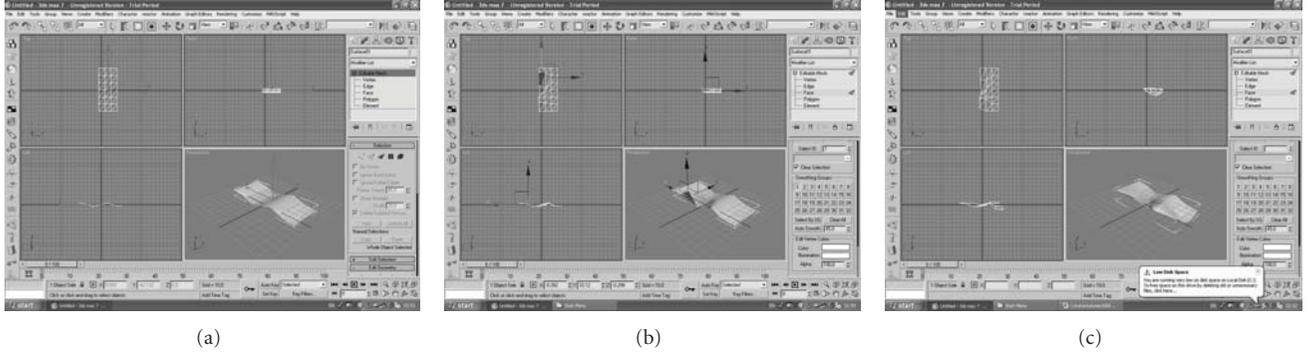


FIGURE 4: Preliminary experiments made in 3DS Max7.

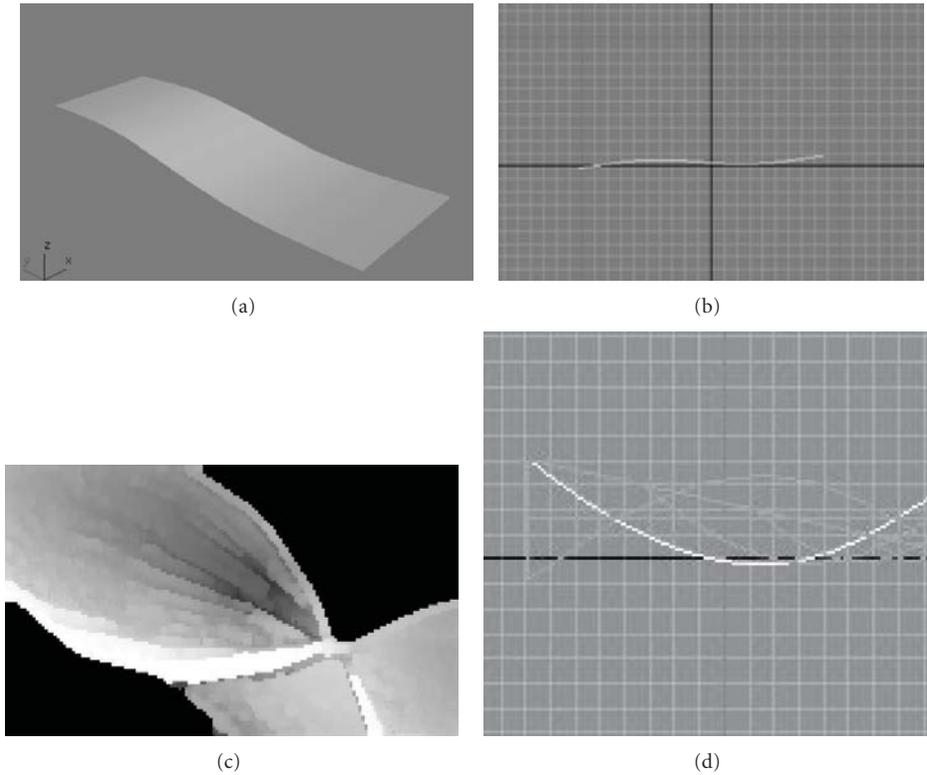


FIGURE 5: Results obtained with the 3D reconstruction component of MoDef 3D Visual environment: (a) initial 3D representation of the deformable surface of the surgical mesh, (b) curve representing a section of the surgical mesh acquired by the transducer, extracted from the context of the US image based on specific-image processing method, (c) the surface of the prosthetic mesh after the deformations produced in time due to the anatomic assimilation process, and (d) the basic set of generating curves, used to obtain the solid-view representations of the prosthetic mesh.

Now, we obtain from (70), (71), and (72):

$$\begin{aligned}
 \mu &= (1 - 2r_1 - 6r_2) + (r_1 + 4r_2)(e^{j\eta_1} + e^{-j\eta_1}) \\
 &\quad - r_2(e^{2j\eta_1} + e^{-2j\eta_1}), \\
 \mu &= (1 - 2r_1 - 6r_2) + (r_1 + 4r_2)(e^{j\eta_2} + e^{-j\eta_2}) \\
 &\quad - r_2(e^{2j\eta_2} + e^{-2j\eta_2}).
 \end{aligned} \tag{73}$$

We choose $\omega_1 = \omega_2$ and let $\eta = \eta_1/2 = \eta_2/2$. The trigonometric formulas $e^{j\alpha} + e^{-j\alpha} = 2 \cos \alpha$, $1 - \cos \alpha =$

$2 \sin^2 \alpha/2$, and $1 - \cos 2\alpha = 8 \sin^2 \alpha/2 \cos^2 \alpha/2$, $\alpha \in \mathbb{R}$ together with (73) give:

$$\mu = 1 - 4r_1 \sin^2 \eta - 16r_2 \sin^4 \eta. \tag{74}$$

On the other hand, according to the relation $|\exp(j\omega t)| = 1$, it is easy to see that the error ε_i^k of (72) does not increase in time if $|\mu| \leq 1$, so that we infer from (74):

$$2r_1 + 8r_2 \leq 1, \tag{75}$$

which represent precisely the *stability criterion* of von Neumann for the (ELP) algorithm (46).

A combination of the relations (39), (62), (71), and (75) provides the following equivalent form of the stability condition of von Neumann:

$$2\varepsilon(4w_2 + w_1h^2) \leq 1. \quad (76)$$

4. Monitoring the Behavior of Prosthetic Surgical Methods and Prosthetic Medical Materials Based on Software Implementation

In order to apply the results of the theoretical researches detailed above in the medical imaging domain, a 3D visual software environment—named MoDef—was implemented, aiming to visualize and follow up the deformation behavior of the surgical (abdominal, maxilla-facial, and orthodontic) prosthetic materials. That is performed on three distinct, but convergent, levels, as follows:

- (a) 3D reconstruction visual software component, aimed to tracks the evolution of the prosthetic materials, based on processing the US images of the anatomic context of a lot of surgical patients;
- (b) deformable prosthetic material's behavior forecasting software component, based on software tools which implements the above described mathematical methods;
- (c) quad comparative parallel tracking software component, aimed to simultaneous supervise in time both (a) and (b) levels, in comparison with the results provided by the stochastic analysis component of the 3D visual software environment MoDef.

Concerning the 3D visualizing of the prosthetic meshes by means of the MoDef software environment components, two levels of reconstruction are performed, namely

- (1) on the first level, a polynomial interpolation method is applied on each slice of the US image of the prosthetic mesh, acquired based on succeeding positions of the transducer, obtained by rotating them with a constant angle in a same preestablished direction; more exactly, the curves representing the sections of the surgical mesh acquired by the transducer are extracted from the context of the US image, based on specific image processing methods, namely, contour detection methods, that are implemented at the level of the image processing operators of the MoDef environment's image processing library. Starting with this set of basic mesh surface definition curves, extracted from the US images acquired at pre-established moments in time, a complete and consistent collection of 3D generator curve sets is obtained, by means of 3D polynomial interpolation methods, based on Lagrange, Hermite or Birkhoff operators;
- (2) on the second level, the complete collection of the 3D generator curves obtained at the first level is processed based on Blended Interpolating Methods (BIM), as well as with 3D continuous representation techniques, in order to obtain "solid-view," respect-

ively, "wired-view" representations of the prosthetic mesh.

In what follows, some preliminary experiments made in 3DS Max7, followed by some relevant results obtained with the 3D reconstruction component of MoDef 3D Visual environment are presented in Figures 4 and 5.

5. Conclusions

In this paper we considered parametric (variational) deformable models and we developed an iterative method based on finite difference schemes in order to solve numerically the (ELP) equation of Calculus of Variations, which provides the energy minimizing snake. We derived estimates concerning the approximation error related to the corresponding (ELP) algorithm and we established conditions for its convergence and stability. Some considerations about the implementation of the above numerical methods where presented, too. As future targets, we intend to consider probabilistic models which offer an alternative approach by using the Bayes technique, as well as geometric deformable models which provide an efficient alternative to address some limitation of parametric deformable models.

References

- [1] M. Fischler and R. A. Elschlager, "The representation and matching of pictorial structures," *IEEE Transactions on Computers*, vol. C-22, no. 1, pp. 67–92, 1973.
- [2] B. Widrow, "The rubber mask technique, part I," *Pattern Recognition*, vol. 5, no. 3, pp. 175–211, 1973.
- [3] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis*, vol. 1, no. 2, pp. 91–108, 1996.
- [4] D. Terzopoulos, "On matching deformable models to images," Tech. Rep. 60, Schlumberger Palo Alto Research, 1986, Reprinted in Topical Meeting on Machine Vision, Technical Digest Series, vol. 12, pp.160–167, Optical Society of America, Washington, DC, USA, 1987.
- [5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [6] D. Terzopoulos and K. Fleischer, "Deformable models," *The Visual Computer*, vol. 4, no. 6, pp. 306–331, 1988.
- [7] V. Caselles, F. Catté, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," *Numerische Mathematik*, vol. 66, no. 1, pp. 1–31, 1993.
- [8] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: a level set approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 158–175, 1995.
- [9] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988.
- [10] L. He, Z. Peng, B. Everding et al., "A comparative study of deformable contour methods on medical image segmentation," *Image and Vision Computing*, vol. 26, no. 2, pp. 141–163, 2008.
- [11] R. Hegadi, A. Kop, and A. Hangarge, "A survey on deformable model and its applications to medical imaging," in *IJCA*

Special Issue on Recent Trends in Image Processing and Pattern Recognition, pp. 64–75, 2010.

- [12] S. Y. Chen and Q. Guan, “Parametric shape representation by a deformable NURBS model for cardiac functional measurements,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3 PART 1, pp. 480–487, 2011.
- [13] 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.
- [14] 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, 2012.
- [15] S. Y. Chen, M. Zhao, G. Wu, and C. Yao, “Recent advances in morphological cell image analysis,” *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 101536, 2012.
- [16] A. A. Amini, T. E. Weymouth, and R. C. Jain, “Using dynamic programming for solving variational problems in vision,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 9, pp. 855–867, 1990.
- [17] S. Nedevschi and D. Mitrea, “Contour detection based on active contour models,” *Bulletin of Applied Mathematics and Computer Science*, vol. 2275, pp. 107–118, 2003.
- [18] N. Rougon and F. Prêteux, “Directional adaptive deformable models for segmentation,” *Journal of Electronic Imaging*, vol. 7, no. 1, pp. 231–256, 1998.
- [19] L. D. Cohen and I. Cohen, “Finite-element methods for active contour models and balloons for 2-D and 3-D images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1131–1147, 1993.
- [20] I. Cohen, L. D. Cohen, and N. Ayache, “Using deformable surfaces to segment 3-D images and infer differential structures,” *CVGIP: Image Understanding*, vol. 56, no. 2, pp. 242–263, 1992.
- [21] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing, Partial Differential Equations and Calculus of Variations*, Springer, 2002.
- [22] R. Richtmeyer and K. Morton, *Difference Methods for Initial Value Problems*, Wiley, New York, NY, USA, 1971.

Research Article

Multicriteria Optimization Model for the Study of the Efficacy of Skin Antiaging Therapy

**Maria Crișan,¹ Guido Cappare,² Luciana Neamțiu,³ Ioana Chiorean,⁴ Liana Lupșa,⁴
Diana Crișan,⁵ and Radu Badea⁶**

¹ Department of Histology, Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

² Department of Aesthetics, AMIA, Milano, Italy

³ Ion Chiricuță Cancer Institute, Cluj-Napoca, Romania

⁴ Faculty of Mathematics and Computer Science, Babeș-Bolyai University, Cluj-Napoca, Romania

⁵ Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

⁶ Department of Ultrasonography, Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

Correspondence should be addressed to Maria Crișan, mcrisan7@yahoo.com

Received 29 September 2011; Accepted 12 December 2011

Academic Editor: Carlo Cattani

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

The evolution of the cutaneous structure after topical treatment with P63 antiaging complex, assessed with high frequency ultrasound, is studied by means of multicriteria optimization model. Due to the fact that the impact of the treatment may influence the quality of life, a medical index which measures, from this point of view, the efficacy of the treatment is given, also taking into account medical and economical aspects.

1. Introduction

The basic idea of Pharmaco-Economics studies is to gain a physical and psychological comfort state for as long as possible, with the smallest amount of money. Therefore, according to these studies (see [1]) which consider a treatment in terms of results related to costs, one of the following five types of analysis is used: and cost-effects (CEA), cost-minimization (CMA), cost-utility (CUA), cost-efficiency (CEAC), and cost-benefit (CBA). Then, the data obtained after each of this analysis are used to compare two or more treatments. Cost-utility analysis was developed to help decision makers compare the value of alternative interventions that have very different health benefits, and it facilitates these comparisons without recourse to placing monetary values on different health states. The primary outcome of a cost-utility analysis is the cost per quality-adjusted life years (QALYs) or incremental cost-effectiveness ratio (ICER), which is calculated as the difference in the expected cost of two interventions, divided by the difference in the expected QALYs produced by the two interventions. QALYs measure

health as a combination of the duration of life and the health-related quality of life. Also, there is another index, denoted by NB or INB, which means *incremental net benefit*, defined by

$$\text{INB} = \lambda \Delta_e - \Delta_c, \quad (1)$$

where λ is the willingness to pay.

These indexes are largely used in the literature (see, e.g., [2–5]). Unfortunately, they may not always reflect with sufficient accuracy all aspects of medical outcomes of treatment, perception and impact of treatment on the patient's psyche, economic effects, and so forth.

Using the multicriteria optimization technique, in [6] a new index, called medicoeconomic index of a treatment (denoted by MEI), is introduced. Its construction may use all the desired aspects. It permits a simultaneous comparison of two or more medical treatments. And, in addition, due to the fact that it emphasizes the importance of every aspect in a general context, it gains an increased flexibility (see [7–10]).

There are papers in the literature which use multifactorial decisions to compare the medical treatments (see the book [11] and the studies [12, 13]).

The purpose of this paper is to show how the MEI index may best quantify the effect of applying a skin rejuvenation treatment, taking into account multiple aspects: changes in skin parameters occurred after treatment application, side effects, how the patient perceives the treatment and outcome of application, cost of the treatment.

Quantification, materialized by building the MEI index, permits comparisons among several types of treatments or comparing the results obtained with the same treatment applied to different groups of patients, groups that differ either by age or by phototype of the patients.

2. The Model on a Specific Sample

Let T be a medical treatment, applied against skin aging, whose effect we want to quantify. In particular we considered treatment with P63 antiaging complex.

From the treatment point of view, the patients were classified in two groups of study: the first group included persons to whom the new medication to be tested was applied, and the second group, the control group, to whom the old medication was given. It is important to mention that both groups were homogenous relatively to patient age, health state, and phenotype. From medical point of view, the obtained results refer to thickness of the epidermis (mm), thickness of the dermis (mm), number of low echogenity pixels (LEP), medium and high echogenity pixels (MEP and HEP), and the LEPs/LEPi ratio (low echogenic pixels in the upper dermis/low echogenic pixels in the lower dermis). All these imagistic parameters quantify both the skin aging process and the efficacy of the antiaging treatments. For each of these parameters, the mean values for every study group, age group, and phenotype group are given.

Taking this into account, for each treatment T_i , $i \in \{1, 2\}$, each age group j , $j \in \{1, 2, 3\}$, and each phenotype h , $h \in \{2, 3\}$, we indicate the following:

- (i) E_{ijh1}, E_{ijh2} , the mean values of the thickness of epidermis of the patients who take the treatment i , are in age group j and have phenotype h , before the treatment, and, respectively, after the treatment;
- (ii) D_{ijh1}, D_{ijh2} , the mean values of the thickness of dermis of the patients who take the treatment i , are in age group j and have phototype h , before the treatment, respectively, after the treatment;
- (iii) L_{ijh1}, L_{ijh2} , the mean values of the LEP of the patients who take the treatment i , are in age group j and have phototype h , before the treatment, respectively after the treatment;
- (iv) M_{ijh1}, M_{ijh2} , the mean values of the MEP of the patients who take the treatment i , are in age group j and have phototype h , before the treatment, respectively, after the treatment;
- (v) H_{ijh1}, H_{ijh2} , the mean values of the HEP of the patients who take the treatment i , are in age group

j and have phototype h , before the treatment, respectively, after the treatment;

- (vi) F_{ijh1}, F_{ijk2} , the mean values of LEPi/LEPu;
- (vii) $R_{ijh}\%$, the percent of those with no adverse effects (erythema, pruritus, ocular disturbance, etc.);
- (viii) C_{ijh} , the mean value of the cost of treatment for the period of time when the study was performed, taking into account a person in group i , of age j and a phototype h .

The patients were asked to fill in a questionnaire and give two marks, denoted by q and $s \in \{1, \dots, 10\}$, concerning the convenience of treatment application and the consequential effect of the treatment, respectively. We denote by Q_{ijh} and S_{ijk} , respectively, the mean values of the marks given by the patients in age group j for the treatment i , and which have phototype h .

Let K be the maximum accepted value for the cost of the treatment for one patient during the study period of time.

2.1. Methods. The method used for comparison was that of introducing a medicoeconomic index, by means of multicriteria optimization, which permits a rigorous observation of the evolution of our patients.

For a protocol treatment T_i , the percentage $P_{ijh}\%$ of patients from treatment group i , age group j , and phototype h , is known.

Our problem was the following: having the previous information and taking into account the cost of the new treatment and the opinions of the patients, we wanted to determine the efficacy of the new treatment, and consequently, if it was worth using it. The expected medical results are increase of the mean thickness of epidermis and dermis, decrease of the mean number of low echogenity pixels, the increase of medium and high echogenity pixels, and the increase of LEPs/LEPi ratio. From economical point of view, it is desirable that the mean value of the cost for the new treatment to be as low as possible.

Finally, from the patients' point of view, the effect of the treatment is expected to be maximum possible and its application as comfortable as it can be.

In order to solve this problem, we compared the treatments at a general level (group 1 versus group 2 of treatments), but also on each age group and phenotype group. In every case, we had to solve a mathematical multicriteria optimization problem.

We introduced the variables x_1 and x_2 , which take only the values 0 and 1. We have: $x_1 = 1$, if the first treatment protocol is chosen, $x_1 = 0$, if not; $x_2 = 1$, if the second treatment protocol is chosen, $x_2 = 0$, if not. Due to the fact that one patient has to follow one and only one treatment protocol, we have the obvious condition $x_1 + x_2 = 1$.

Then we consider the functions: $f_k : \{0, 1\}^2 \rightarrow \mathbf{R}$, $k \in \{1, \dots, 60\}$, given for every $j \in \{1, 2, 3\}$ and $h \in \{2, 3\}$, respectively, by

$$\begin{aligned}
f_{2(j-1)+(h-1)}(x_1, x_2) &= P_{1jh} \left(\frac{E_{1jh2}}{E_{1jh1}} \right) x_1 + P_{2jh} \left(\frac{E_{2jh2}}{E_{2jh1}} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2, \\
f_{6+2(j-1)+(h-1)}(x_1, x_2) &= P_{1jh} \left(\frac{D_{1jh2}}{D_{1jh1}} \right) x_1 + P_{2jh} \left(\frac{D_{2jh2}}{D_{2jh1}} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2, \\
f_{12+2(j-1)+(h-1)}(x_1, x_2) &= P_{1jh} \left(\frac{L_{1jh1}}{L_{1jh2}} \right) x_1 + P_{2jh} \left(\frac{L_{2jh1}}{L_{2jh2}} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2, \\
f_{18+2(j-1)+(h-1)}(x_1, x_2) &= P_{1jh} \left(\frac{M_{1jh2}}{M_{1jh1}} \right) x_1 + P_{2jh} \left(\frac{M_{2jh2}}{M_{2jh1}} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2, \\
f_{24+2(j-1)+(h-1)}(x_1, x_2) &= P_{1jh} \left(\frac{H_{1jh2}}{H_{1jh1}} \right) x_1 + P_{2jh} \left(\frac{H_{2jh2}}{H_{2jh1}} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2, \\
f_{30+2(j-1)+(h-1)}(x_1, x_2) &= P_{1jh} \left(\frac{F_{1jh2}}{F_{1jh1}} \right) x_1 + P_{2jh} \left(\frac{F_{2jh2}}{F_{2jh1}} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2, \\
f_{36+2(j-1)+(h-1)}(x_1, x_2) &= - \left(\frac{R_{1jh}}{P_{1jh}} \right) x_1 - \left(\frac{R_{2jh2}}{P_{2jh}} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2, \\
f_{42+2(j-1)+(h-1)}(x_1, x_2) &= P_{1jh} \left(\frac{Q_{1jh}}{10} \right) x_1 + P_{2jh} \left(\frac{Q_{2jh}}{10} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2, \\
f_{48+2(j-1)+(h-1)}(x_1, x_2) &= P_{1jh} \left(\frac{S_{1jh}}{10} \right) x_1 + P_{2jh} \left(\frac{S_{2jh}}{10} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2, \\
f_{54+2(j-1)+(h-1)}(x_1, x_2) &= -P_{1jh} \left(\frac{C_{1jh}}{K} \right) x_1 - P_{2jh} \left(\frac{C_{2jh}}{K} \right) x_2, \quad \forall (x_1, x_2) \in \{0, 1\}^2.
\end{aligned} \tag{2}$$

Obviously, our purpose is to determine, simultaneously, their maximum. Generally, this fact cannot be fulfilled; therefore, a compromised solution is to be accepted: a synthesis function is considered and each of the previous functions gets a weight, according to the importance of every criterion in the problem.

So, firstly we attribute weights to the main elements of our final purpose:

- (i) getting the best medical results (purpose indicated by the functions $f_1, \dots, f_{36}, f_{49}, \dots, f_{54}$);
- (ii) the best reception of the patient (functions f_{43}, \dots, f_{54});
- (iii) the lowest possible cost of the treatment (indicated by the functions f_{55}, \dots, f_{60}).

Each of these purposes gets a nonnegative weight denoted by t_u ($u \in \{1, 2, 3\}$), with the property that $t_1 + t_2 + t_3 = 1$. Obviously, the three weights may be or may be not equal.

Remark 1. In some other cases, we may distinguish these weights even further. More precisely, the weight t_2 , corresponding to the way in which the patient is contained with the treatment may be considered as the sum between two others weights: t_Q , for the convenience of treatment application and t_S for the degree of content after the

treatment. Also, the weight t_1 may be seen as the sum of 7 weights, $t_E, t_D, t_L, t_M, t_H, t_F, t_R$, connected with the importance given to the increase of the thickness of the epidermis, dermis, decreasing of low echogenity pixels, increasing of medium and high echogenity pixels and increasing of ratio LEPi/LEPu, respectively decreasing of the adverse reactions. We change the notation of t_3 with t_C .

3. Comparing the Results at a General Level and Constructing the MEI Index

In this case, the mathematical model of our medical problem is the following multicriterial optimization problem with the objective function $F = (f_1, \dots, f_{60}) : \{0, 1\}^2 \rightarrow \mathbf{R}^{60}$, previously given, and the set of possible solutions $X = \{(x_1, x_2) \in \{0, 1\}^2 \mid x_1 + x_2 = 1\}$.

In order to solve this problem, we apply the weight method and we construct the synthesis function $\Lambda : \{0, 1\}^2 \rightarrow \mathbf{R}$, $\Lambda(x_1, x_2) = \sum_{i=1}^{60} \lambda_i f_i(x_1, x_2)$, for every $x = (x_1, x_2) \in \{0, 1\}^2$, where λ_i is the importance (weight) given to the criteria generated by function f_i . Taking into account the previous considerations, we set $\lambda_1 = \dots = \lambda_6 = t_E$, $\lambda_7 = \dots = \lambda_{12} = t_D$, $\lambda_{13} = \dots = \lambda_{18} = t_L$, $\lambda_{19} = \dots = \lambda_{24} = t_M$, $\lambda_{25} = \dots = \lambda_{30} = t_H$, $\lambda_{31} = \dots = \lambda_{36} = t_F$, $\lambda_{37} =$

$\dots = \lambda_{42} = t_R, \lambda_{43} = \dots = \lambda_{48} = t_Q, \lambda_{49} = \dots = \lambda_{54} = t_S, \lambda_{55} = \dots = \lambda_{60} = t_C.$

Any optimum solution of the problem

$$\Lambda(x_1, x_2) \longrightarrow \max, \quad (x_1, x_2) \in X \quad (3)$$

is a max-efficient point (max-Pareto point), so is acceptable. Due to the fact that condition $x_1 + x_2 = 1$ implies $x_1 = 1 - x_2$, the solving of problem (3) reduces to the solving of the following problem:

$$\psi(u) = \sum_{i=1}^{60} \lambda_i f_i(u, 1 - u) \longrightarrow \max, \quad u \in \{0, 1\}. \quad (4)$$

The simple form of the restrictions conduces immediately to the conclusion: if $\psi(1) > \psi(0)$, then $u = 1$ is the optimal solution of problem (4); if not, $u = 0$ is the optimal solution. This mathematical result permits us to introduce the MEI index, called medicoeconomic index, attached to the protocol T_i , by

$$\begin{aligned} MEI_{T_i} = & \sum_{j=1}^3 \sum_{h=1}^2 \left[p_{ijh} \left(t_E \frac{E_{ijh2}}{E_{ijh1}} + t_D \frac{D_{ijh2}}{D_{ijh1}} + t_L \frac{L_{ijh2}}{L_{ijh1}} + t_M \frac{M_{ijh2}}{M_{ijh1}} \right. \right. \\ & \left. \left. + t_H \frac{H_{ijh2}}{H_{ijh1}} + t_F \frac{F_{ijh2}}{F_{ijh1}} \right) - t_R \frac{R_{ijk}}{p_{ijk}} \right] \\ & + \sum_{j=1}^3 \sum_{h=1}^2 \left[p_{ijh} \left(t_Q \frac{Q_{ijh2}}{Q_{ijh1}} + t_S \frac{S_{ijh2}}{S_{ijh1}} - t_C \frac{C_{ijh}}{K} \right) \right]. \end{aligned} \quad (5)$$

In case of small or homogenous samples, we may use the easier formula:

$$\begin{aligned} MEI_{T_i} = & \sum_{j=1}^3 \sum_{h=1}^2 \left(t_E \frac{E_{ijh2}}{E_{ijh1}} + t_D \frac{D_{ijh2}}{D_{ijh1}} + t_L \frac{L_{ijh2}}{L_{ijh1}} \right. \\ & \left. + t_M \frac{M_{ijh2}}{M_{ijh1}} + t_H \frac{H_{ijh2}}{H_{ijh1}} + t_F \frac{F_{ijh2}}{F_{ijh1}} - t_R R_{ijh} \right) \\ & + \sum_{j=1}^3 \sum_{h=1}^2 \left(t_Q \frac{Q_{ijh2}}{Q_{ijh1}} + t_S \frac{S_{ijh2}}{S_{ijh1}} - t_C \frac{C_{ijh}}{K} \right). \end{aligned} \quad (6)$$

Remark 2. Since for the construction of this index we consider only data related to the patient's condition before and after treatment and data referring only this treatment, the index can be used to compare several treatments, unlike ICER and NB that allow comparison of only two treatments and of only two criteria related to cost and effectiveness.

4. Special Cases

If we want to compare the overall effectiveness of treatment for two different age groups, we may follow the same steps, except that we consider only the data relative to an age group.

For example, if we want to compare the efficiency of the treatment for each age group, i.e. 40–49, 50–59 and over 60, we construct the MEI index by taking

$$\begin{aligned} MEI_{\text{group } j} = & \sum_{h=1}^2 \left[P_{1jh} \left(t_E \frac{E_{1jh2}}{E_{1jh1}} + t_D \frac{D_{1jh2}}{D_{1jh1}} + t_L \frac{L_{1jh2}}{L_{1jh1}} \right. \right. \\ & \left. \left. + t_M \frac{M_{1jh2}}{M_{1jh1}} + t_H \frac{H_{1jh2}}{H_{1jh1}} \right) - t_R \frac{R_{1jh}}{P_{1jh}} \right] \\ & + \sum_{h=1}^2 \left[P_{1jh} \left(t_F \frac{F_{1jh2}}{F_{1jh1}} + t_Q \frac{Q_{1jh2}}{Q_{1jh1}} + t_S \frac{S_{1jh2}}{S_{1jh1}} - t_C \frac{C_{1jh}}{K} \right) \right]. \end{aligned} \quad (7)$$

In case of small or homogenous samples, an easier formula for computing the MEI index is the following:

$$\begin{aligned} MEI_{\text{group } j} = & \sum_{h=1}^2 \left(t_E \frac{E_{1jh2}}{E_{1jh1}} + t_D \frac{D_{1jh2}}{D_{1jh1}} + t_L \frac{L_{1jh2}}{L_{1jh1}} + t_M \frac{M_{1jh2}}{M_{1jh1}} \right. \\ & \left. + t_H \frac{H_{1jh2}}{H_{1jh1}} + t_F \frac{F_{1jh2}}{F_{1jh1}} - t_R R_{1jh} \right) \\ & + \sum_{h=1}^2 \left(t_Q \frac{Q_{1jh2}}{Q_{1jh1}} + t_S \frac{S_{1jh2}}{S_{1jh1}} - t_C \frac{C_{1jh}}{K} \right). \end{aligned} \quad (8)$$

Another special case may be the one connected with the phototype. In this case, we have

$$\begin{aligned} MEI_{\text{phenotyp } h} = & \sum_{j=1}^3 \left[P_{1jh} \left(t_E \frac{E_{1jh2}}{E_{1jh1}} + t_D \frac{D_{1jh2}}{D_{1jh1}} + t_L \frac{L_{1jh2}}{L_{1jh1}} \right. \right. \\ & \left. \left. + t_M \frac{M_{1jh2}}{M_{1jh1}} + t_H \frac{H_{1jh2}}{H_{1jh1}} \right) - t_R \frac{R_{1jh}}{P_{1jh}} \right] \\ & + \sum_{j=1}^3 \left[P_{1jh} \left(t_F \frac{F_{1jh2}}{F_{1jh1}} + t_Q \frac{Q_{1jh2}}{Q_{1jh1}} \right. \right. \\ & \left. \left. + t_S \frac{S_{1jh2}}{S_{1jh1}} - t_C \frac{C_{1jh}}{K} \right) \right]. \end{aligned} \quad (9)$$

Or, the easier formula is as follows:

$$\begin{aligned} MEI_{\text{phenotyp } h} = & \sum_{j=1}^3 t_E \frac{E_{1jh2}}{E_{1jh1}} + t_D \frac{D_{1jh2}}{D_{1jh1}} + t_L \frac{L_{1jh2}}{L_{1jh1}} \\ & + t_M \frac{M_{1jh2}}{M_{1jh1}} + t_H \frac{H_{1jh2}}{H_{1jh1}} + t_F \frac{F_{1jh2}}{F_{1jh1}} \\ & + t_Q \frac{Q_{1jh2}}{Q_{1jh1}} + t_S \frac{S_{1jh2}}{S_{1jh1}} - t_C \frac{C_{1jh}}{K} - t_R R_{1jh}. \end{aligned} \quad (10)$$

Also, if we want a result based both on the age group and phototype, we have

$$\begin{aligned} MEI_{j,h} = & t_E \frac{E_{1jh2}}{E_{1jh1}} + t_D \frac{D_{1jh2}}{D_{1jh1}} + t_L \frac{L_{1jh2}}{L_{1jh1}} \\ & + t_M \frac{M_{1jh2}}{M_{1jh1}} + t_H \frac{H_{1jh2}}{H_{1jh1}} + t_F \frac{F_{1jh2}}{F_{1jh1}} + t_Q \frac{Q_{1jh2}}{Q_{1jh1}} \\ & + t_S \frac{S_{1jh2}}{S_{1jh1}} - t_C \frac{C_{1jh2}}{K} - t_R R_{1jh}. \end{aligned} \quad (11)$$

Table 1: Group treated with P63 antiaging treatment.

| Parameters | Mean value before treatment | Mean value after treatment |
|-----------------------------|-----------------------------|----------------------------|
| Thickness of epidermis (mm) | 0.117 | 0.135 |
| Thickness of dermis (mm) | 1.537 | 1.719 |
| LEPs/LEPi | 1.149 | 1.574 |

Table 2: Group treated with P63 complex, based on age groups.

| Parameters | Mean b.t. 40–49 | Mean a.t. 40–49 | Mean b.t. 50–59 | Mean a.t. 50–59 | Mean b.t. 60–75 | Mean a.t. 60–75 |
|-----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Thickness of epidermis (mm) | 0.114 | 0.135333 | 0.11675 | 0.1345 | 0.12025 | 0.136125 |
| Thickness of dermis (mm) | 1.8585 | 1.901 | 1.386313 | 1.740563 | 1.367625 | 1.506625 |
| LEPs/LEPi | 1.204593 | 1.457353 | 1.096062 | 1.398047 | 1.213842 | 1.510156 |

Table 3: Group treated with P63 antiaging treatment, based on phototype.

| Parameters | Mean b.t. phototype 2 | Mean a.t. phototype 2 | Mean b.t. phototype 3 | Mean a.t. phototype 3 |
|-----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Thickness of epidermis (mm) | 0.114625 | 0.129938 | 0.12 | 0.141 |
| Thickness of dermis (mm) | 1.605313 | 1.768063 | 1.460571 | 1.644214 |
| LEPs/LEPi | 1.198661 | 1.580802 | 1.092623 | 1.567998 |

5. Application of the P63 Antiaging Treatment

We exemplify the application of this index for the characterization of the treatment based on P63 antiaging treatment (a metabolic dynamiser composed of alpha hydroxyacids, retinoids, a biomimetic peptidic complex, and gluconolactone encapsulated in liposomes), in the particular case of data presented in the following table and taken from [14]. Our mathematical model was applied on a sample of 30 patients with ages between 40 and 75 years old. Due to our purpose, we divided them in three age interval groups: the first group included patients aged 40–50 years, the second group included the aged ones 50–60, and the last group contained those over 60. All patients included in the study belonged to phototype class II or III, and, therefore, following this criteria, we grouped them into two classes, denoted by 2 and 3. The treatment cost equals 250 monetary units (m.u); the K constant is taken equal to 400. For those patients who did not follow the treatment, the cost was considered 100 m.u.

The medical data concerning the group treated with P63 antiaging treatment, as well as the placebo group, are given in the following tables.

In order to compute the MEI index in case of P63 antiaging complex treatment, and placebo one, formula (6) was used. It is important to mention that no side effects were noticed.

We considered equally the weights given for the age groups. We take $t_E = 1/8, t_D = 1/8, t_F = 1/2, t_R = t_Q = t_S = 0$, and $t_C = 1/4$. Under these circumstances and using the data from Table 1, the MEI index for the treatment with

Table 4: Placebo group.

| Parameters | Mean value before treatment | Mean value after treatment |
|-----------------------------|-----------------------------|----------------------------|
| Thickness of epidermis (mm) | 0.1358 | 0.13505 |
| Thickness of dermis (mm) | 1.49635 | 1.57155 |
| LEPs/LEPi | 0.924757 | 1.042922 |

P63 antiaging complex was $T_{P63 \text{ complex}} = 0.812143$. Using the data of Table 4 to build the MEI index for the placebo treatment, we obtained $T_0 = 0.569481$. This means that the treatment is worth doing.

Using the same weights, the data from Table 2 and denoting by “Mean” the mean value, b.t. = before treatment and a.t. = after treatment and computing MEI for every age group, we also got the following results: $MEI_{40-49} = 0.724915$, $MEI_{50-59} = 0.897944$, and $MEI_{60-75} = 0.745013$.

The highest value was obtained for the 50–59 group, then for the group 60–75 and then for 40–49 group. It means that the age group 50–59 had the best response to the treatment.

Using the same weights, the data from Table 3 and computing MEI, according with the formula for phototypes 2 and 3, we get $MEI_2 = 0.782524$, respectively, $MEI_3 = 0.84888$. The greatest value is obtained for phototype 3. This result indicates that the treatment was best perceived for phototype 3.

Remark 3. We mention that a change in the weights leads to a change of the index.

6. Conclusions

As mentioned before, starting with effective data concerning the treatment with P63 complex, we studied the possibility to quantify the outcome of antiaging skin treatment, from multiple points of view: medical parameters, side effects, cost, convenience of treatment application, the way in which the patient perceive the treatment.

Using multicriteria optimization, we introduced a medico-economic index (MEI) capable to perform the required quantification. In its construction, only data connected with a specific treatment were used; therefore, it has the advantage of being able to compare more than 2 treatments.

Further, we gave an application in which we computed the MEI index, in order to compare the P63 antiaging complex treatment with a placebo one (which provided only hydration of the skin). We also constructed specific indexes, according to age groups. This permitted us to compare the effects of P63 complex treatment for different age groups and to establish, from this point of view, a hierarchy of these groups. The same constructions and comparisons were made for the groups of phototypes 2 and 3.

The application presented documents the importance and the facility of use of our index.

Acknowledgments

The authors gratefully acknowledge Cortex Technology for allowing the use of the ultrasound equipment and for their assistance with the application of the Dermascan to this project. The study is part of a national research grant CNCSIS-IDEI, no. 2624.

References

- [1] Pharmaceutical Research and Manufacturers of America, "Methodological and conduct principles for pharmaco-economic research," *Drug Benefit Trends*, vol. 8, no. 2, p. 2834, 1996.
- [2] A. Briggs, K. Claxton, and M. Sculpher, *Decision Modelling for Health Economic Evaluation*, Oxford University Press, Oxford, UK, 2006.
- [3] A. R. Willan, "Statistical analysis of cost-effectiveness data from randomized clinical trials," *Expert Review of Pharmacoeconomics and Outcomes Research*, vol. 6, no. 3, pp. 337–346, 2006.
- [4] A. O'Hagan and J. W. Stevens, "A framework for cost-effectiveness analysis from clinical trial data," *Health Economics*, vol. 10, no. 4, pp. 303–315, 2001.
- [5] A. Briggs and M. Sculpher, "An introduction to Markov modelling for economic evaluation," *Pharmacoeconomics*, vol. 13, no. 4, pp. 397–409, 1998.
- [6] L. Lupsa, "A criterion for characterizing a medical treatment that uses multicriteria programming in pharmaco-economics," in *Analysis, Functional Equations, Approximation and Convexity*, pp. 142–146, Editura Carpatica, Cluj-Napoca, Romania, 1999.
- [7] L. Lupsa, "Construction of a medico-economic effectiveness index which characterizes a vaccine," in *Séminaire de la Théorie de la Meilleure Approximation, Convexité et Optimisation*, pp. 61–65, SRIMA, Cluj-Napoca, Romania, 2001.
- [8] I. Chiorean, L. Lupsa, and L. Neamtiu, "Markov models for the simulation of cancer screening process," in *Proceedings of the International Conference on Numerical Analysis and Applied Mathematics*, T. E. Simos, G. Psihoyios, and C. Tsitouras, Eds., vol. 1048, pp. 143–146, American Institute of Physics, 2008.
- [9] L. Neamtiu, *Models and Decision Making in Health Economics*, Casa Cărții de Știință, Cluj-Napoca, Romania, 2008.
- [10] C. Grosan, A. Ajith, S. Tigan, and T. G. Chang, "How to solve a multicriterion problem for which Pareto dominance relationship cannot be applied? A case study from medicine," in *Proceedings of the 10th International Conference on Knowledge-Based Intelligent Information Engineering Systems (KES '06)*, Lecture Notes in Computer Science, Springer, 2006.
- [11] S. Tigan, A. Achimas, I. Coman, T. Drugan, and E. Iacob, *Decizii Multifactoriale*, SRIMA, Cluj-Napoca, Romania, 2001.
- [12] G. Cimoca and R. Dollinger, "Specific quantitative and qualitative attribute for medical ranking/evaluation," *Applied Medical Informatics*, vol. 8, no. 1-2, pp. 8–18, 2001.
- [13] I. Toth, L. Gherman, I. Coman, and S. Tigan, "On a multicriterial evaluation of the surgical techniques for anterior hypospadias," *Annals of the Tiberiu Popoviciu Itinerant Seminar of Functional Equations, Approximation and Convexity*, vol. 5, pp. 245–256, 2007.
- [14] M. Crisan, G. Cappare, D. Crisan et al., "Ultrasonographic assessment of the cutaneous changes induced by topical antiaging INTERACTIVE P63 complex. Farmacia ISI-Science citation index expanded. CNCSIS A, In press.

Research Article

Nonlocal Means-Based Denoising for Medical Images

Ke Lu,¹ Ning He,² and Liang Li²

¹ College of Computing & Communication Engineering, Graduate University of Chinese Academy of Sciences, Beijing 100049, China

² School of Information, Beijing Union University, Beijing 100101, China

Correspondence should be addressed to Ke Lu, luk@gucas.ac.cn

Received 20 October 2011; Accepted 29 November 2011

Academic Editor: Sheng-yong Chen

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

Medical images often consist of low-contrast objects corrupted by random noise arising in the image acquisition process. Thus, image denoising is one of the fundamental tasks required by medical imaging analysis. Nonlocal means (NL-means) method provides a powerful framework for denoising. In this work, we investigate an adaptive denoising scheme based on the patch NL-means algorithm for medical imaging denoising. In contrast with the traditional NL-means algorithm, the proposed adaptive NL-means denoising scheme has three unique features. First, we use a restricted local neighbourhood where the true intensity for each noisy pixel is estimated from a set of selected neighbouring pixels to perform the denoising process. Second, the weights used are calculated thanks to the similarity between the patch to denoise and the other patches candidates. Finally, we apply the steering kernel to preserve the details of the images. The proposed method has been compared with similar state-of-art methods over synthetic and real clinical medical images showing an improved performance in all cases analyzed.

1. Introduction

Medical images obtained from Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) and Ultrasound imaging (US) are the most common tools for diagnosis. These images are often affected by random noise arising in the image acquisition process. The presence of noise not only produces undesirable visual quality but also lowers the visibility of low-contrast objects. Image denoising is one of the classical problems in digital image processing. As a primary basis image processing procedure, noise removal has been extensively studied and many denoising schemes have been proposed, from the earlier smoothing filters and frequency domain denoising methods to the lately developed wavelet- [1–5], curvelet- [6], and ridgelet- [7] based methods, sparse representation [8] and K-SVD [9] methods, shape adaptive transform [10], bilateral filtering [11], NL-means based methods [12, 13], and more recently proposed nonlinear variational methods like the total variation minimization [14–16]. With the rapid development of modern digital imaging devices and their increasingly wide applications in our daily life, there are increasing requirements of new denoising algorithms for higher image quality. Particularly, in medical imaging, denoising is challenging since all kinds

of noise cannot be easily modeled and are known to be tissue dependent, such as ultrasound images. Although noise gives an image a generally undesirable appearance, the most significant factor is that noise can cover and reduce the visibility of certain features within the image. The presence of noise gives an image a mottled, grainy, textured, or snowy appearance. In the imaging process, the energy of the high-frequency waves is partially reflected and transmitted at the boundaries between tissues having different acoustic impedances. Nevertheless, the diagnosis quality is often low and reducing speckle while preserving anatomic information is necessary to delineate reliably and accurately the regions of interest. Recently, it has been demonstrated that image patches are relevant features for denoising images in adverse situations. The related methodology can be adapted to derive a robust filter for medical images. Accordingly, in this paper we introduce a novel restoration scheme for medical images, inspired from the NL-means approach introduced by Buades et al. [12] to denoise 2D natural images corrupted by an additive white Gaussian noise.

The rest of this paper is organized as follows. The noise distribution and estimation in medical images are depicted in Section 2.1. The brief description of NL-means algorithm is

discussed in Section 2.2 while the improved NL-means algorithm and the denoising performance under Rician noise are analyzed Section 2.3. The supporting experimental results of improved NL-means algorithm compared to other denoising methods under various conditions are illustrated in Section 3. Finally, concluding remarks are given in Section 4.

2. Improved NL-Means Denoising Method

2.1. Noise Distribution and Estimation in Medical Images. The most MR images acquired in the Fourier domain are characterized by a zero-mean Gaussian probability density function (PDF). After the inverse Fourier transform, the noise distribution in the real and imaginary components will still be Gaussian due to the linearity and the orthogonality of the Fourier transform. However, due to the subsequent transform to a magnitude image, the noise distribution will be no longer Gaussian but Rician distributed. For an MR magnitude image defined on a discrete grid Ω , $M = \{m_i \mid i \in \Omega\}$, then the PDF of m_i is

$$p(m_i \mid A, \sigma) = \frac{m_i}{\sigma^2} e^{-(m_i^2 + A^2/2\sigma^2)} I_0(Am_i/\sigma^2) \varepsilon(m_i), \quad (1)$$

where $I_0(\cdot)$ is the 0th-order modified Bessel function of the first kind and $\varepsilon(\cdot)$ is the Heaviside step function. σ^2 denotes the variance of the Gaussian noise in the complex MR data, which can be independently estimated. When the underlying intensity A equals zero, the Rician PDF simplifies to a Rayleigh distribution:

$$p(m_i \mid A, \sigma) = \frac{m_i}{\sigma^2} e^{-(m_i^2/2\sigma^2)} \varepsilon(m_i). \quad (2)$$

At high SNR, the Rician PDF approaches to a Gaussian PDF with a mean A and variance σ^2 (see Figure 1):

$$p(m_i \mid A, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-((m_i - A)^2/2\sigma^2)} \varepsilon(m_i). \quad (3)$$

That is, Rician noise in magnitude MR images behaves like Gaussian distributed when SNR is high and Rayleigh distributed for low SNR.

Now we discuss how to measure the noise variance from an MR image without the need for high SNR regions or a background region.

Let m_1, m_2, \dots, m_n be the n Rician distributed magnitude data points, and region of constant signal intensity is A . Then the joint PDF of the observations is

$$p(\{m_i\} \mid A, \sigma) = \prod_{i=1}^n \frac{m_i}{\sigma^2} e^{-(m_i^2 + A^2)/2\sigma^2} I_0\left(\frac{Am_i}{\sigma^2}\right), \quad (4)$$

where $\{m_i\}$ are the magnitude variables corresponding to the magnitude observations m_i . The maximum likelihood (ML) estimate of A and σ is then found from the global maximum of $\log L$:

$$\hat{A}_{\text{ML}} = \sum_{i=1}^n \ln\left(\frac{m_i}{\sigma^2}\right) - \sum_{i=1}^n \frac{m_i^2 + A^2}{2\sigma^2} + \sum_{i=1}^n \ln I_0\left(\frac{Am_i}{\sigma^2}\right). \quad (5)$$

Since the noise is estimated from the available piecewise constant regions in the image, this estimation neither depends on the image background nor on the SNR of the image.

2.2. NL-Means Filter. We focus on the problem of denoising: an observed image Y is assumed to be a noisy version of an unobserved image f corrupted by white Gaussian noise. Let $\Omega \subset Z^2$ be the indexing set of the pixels. For any pixel $x \in \Omega$,

$$Y(x) = f(x) + \varepsilon(x), \quad (6)$$

where ε is a centered Gaussian random variable with known variance σ^2 and the noise components $\varepsilon(x)$ are independent. For each pixel the output of the procedure is a weighted average of the whole image. The weights used are selected using a ‘‘metric’’ which determines whether two pixels are similar or not. The core idea of the NL-means is to create a metric governed by patches surrounding each pixel, regardless of their position, that is, nonlocal in the image space. For a fixed (odd) width p , a patch P_x is a subimage of width p , centered around the pixel x , and the NL-means estimator of $f(x)$ is:

$$\hat{f}(x) = \frac{\sum_{x' \in \Omega} w(x, x') Y(x')}{\sum_{x' \in \Omega} w(x, x')}, \quad (7)$$

where $w(x, x') = \exp(-\|P_x - P_{x'}\|_{2,a}^2/2h^2)$, which measures the proximity between patches. $h > 0$ is the bandwidth, which has a smoothing effect and plays the same role as the bandwidth for kernel methods in statistics. The larger the bandwidth is, the smoother the image becomes. $\|\cdot\|_{2,a}$ is a weighted Euclidean norm in $R^{|P|}$ ($|P| = p^2$) using the Gaussian kernel, a controlling the concentration of the norm around the central pixel. The denominator is a normalizing factor ensuring the weights sum to one. The patch size P is generally chosen equal to 5, 7, or 9. From the patch estimator, it is possible to recover a pixel estimator by reprojection.

In the following, the proposed filter is realized in three steps: (a) finding the image patches similar to a given patch; (b) applying the Rician estimation on the 3D block; (c) collaborative adaptive filtering.

2.3. Adaptation to Rician Noise Denoising Model. In case of Rician noise, there is no closed form for the ML estimate of the true signal μ given n such measures x_i . However, the even order moments of the Rician law have very simple expressions. In particular, the second-order moment is $E(X_i^2) = \mu^2 + 2\sigma^2$ where σ^2 is the variance of the Gaussian noise of MRI data. The measured value of x_i^2 (and that of x_i) is thus usually overestimated compared to its true, unknown value, which is termed the Rician bias in the following. Using the same remark as in the Gaussian case, that is, $E(\sum_i w_i X_i^2) = \mu^2 + 2\sigma^2$, it then seems natural to restore x as $\sqrt{\sum_i w_i x_i^2 - 2\sigma^2}$, the weights w_i summing to (1). The voxel value x can be restored as

$$\text{NLM}_R(x) = \sqrt{\left(\sum_{x_i \in V} w_i x_i^2\right) - 2\sigma^2}, \quad (8)$$

where σ^2 is the noise variance. As noted by others in case if i.i.d random variables X_i and with $w_i = 1/n$, the term under the square root has a nonnull probability to be negative,

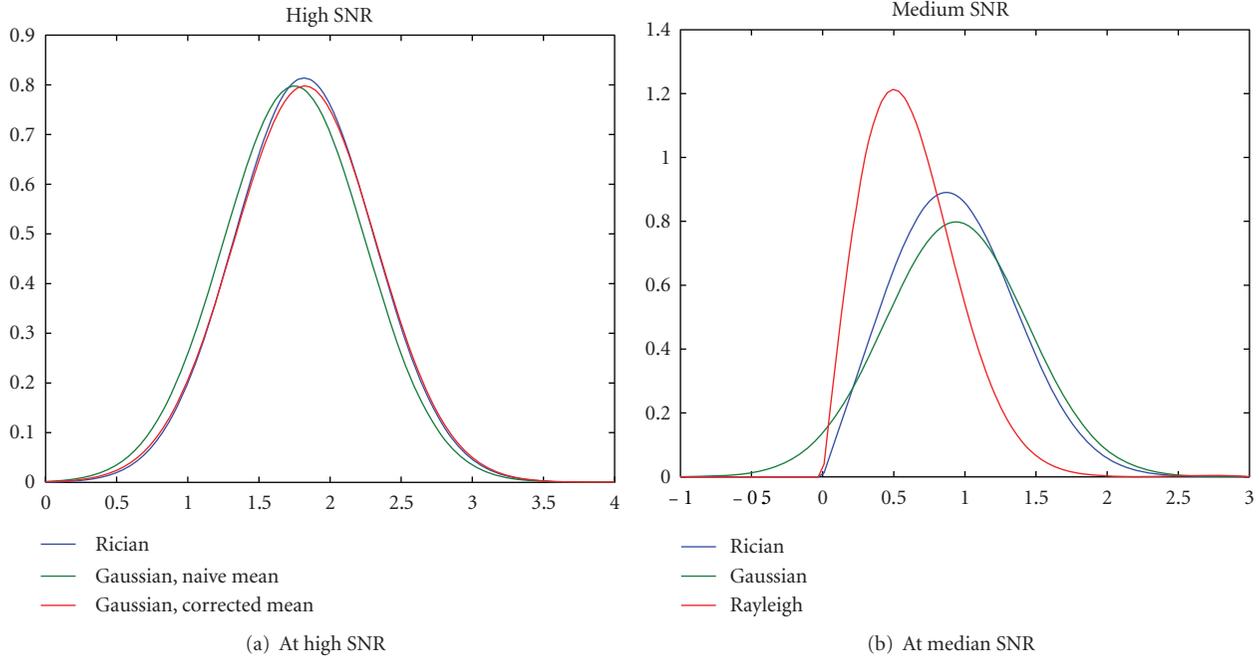


FIGURE 1: At high SNR, Rician data is approximately Gaussian. At low-medium SNR, neither Gaussian nor Rayleigh is a great approximation.

which decreases when n is large. In such cases the restored value is set to zero. In practice, on real data, negative values are mainly found in the background of the images.

On the other hand, we should identify features that capture the underlying geometry of the image patches, without regard to the average intensity of the patches. For this, we make use of the data adaptive steering kernels developed by Takeda et al. [17]. In that work on Steering Kernel Regression (SKR), robust estimates of the local gradients are taken into account in analyzing the similarity between two pixels in a patch. The gradients are then used to describe the shape and size of the kernel. The steering kernel weight at the j th pixel in i th patch, which is a measure of similarity between the two pixels, is then given by

$$w(i, j) = \frac{\sqrt{\det(C_j)}}{2\pi h^2} \exp\left\{-\frac{(x_i - x_j)^T C_j (x_i - x_j)}{2h^2}\right\}, \quad (9)$$

where h is a global smoothing parameter also known as the bandwidth of the kernel. The matrix C_j denotes the gradient covariance formed from the estimated vertical and horizontal gradient of the j th pixel that lies in the i th patch. The 3×3 data-dependent steering matrix C_j can be defined as $C_j = h(H_i)^{-1/2}$, where h is a global smoothing parameter and H_i is a 3×3 covariance matrix based on the sample variations in a local neighborhood around sample x_i . The weight $w(i, j)$ is calculated for each location in the i th patch to form the weight matrix (or kernel). It is interesting to see that the weight matrix thus formed is indicative of the underlying image geometry. This fact is illustrated in Figure 2. Note that in each point of the weight matrix a different C_j is used

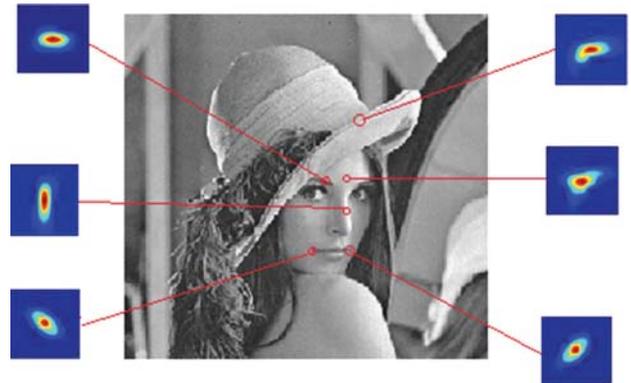


FIGURE 2: Steering kernels at different locations of the Lena image. The patch size is chosen to be 11×11 .

to compute the weight, and hence, the kernels do not have simple elliptical contours.

However, when dealing with nonstationary noise the use of a global noise variance across the image will lead to suboptimal results. To deal with this situation, local noise estimation should be introduced.

Such estimation can be obtained by observing that the expectation of the squared Euclidean distance of two noisy patches as pointed out by Buades et al. is [12]

$$\begin{aligned} d(N_i, N_j) &= E\|u(N_i) - u(N_j)\|_2^2 \\ &= \|u_0(N_i) - u_0(N_j)\|_2^2 + 2\sigma^2, \end{aligned} \quad (10)$$

where u_0 is the noise-free image. Therefore, $d(N_i, N_j) = 2\sigma^2$ if $N_i = N_j$. If we assume that each 2D patch in the volume has at least one patch equal to itself then the noise variance can be estimated as

$$\sigma^2 = \frac{\min(d(N_i, N_j))}{2} \quad \forall j \neq i. \quad (11)$$

However, we found experimentally that this assumption is not normally met in real clinical conditions. In order to relax such an assumption we estimated the local variance as

$$\sigma^2 = \frac{\min(d(R_i, R_j))}{2} \quad \forall j \neq i, R = u - \psi(u), \quad (12)$$

where the distance is calculated from a volume R computed as the subtraction of the original noisy volume u and the lowpass filtered volume $\psi(u)$. We have found experimentally that the minimum distance in this case is approximately equal to σ^2 due to the removal of low-frequency information and the application of the minimum operator.

This Rician adapted filter removes bias intensity using the properties of the second-order moment of a Rice law. In fact, the second-order moment of a random variable X following a Rice distribution can be written as

$$E(X^2) = \mu^2 + 2\sigma^2. \quad (13)$$

Consider a gray-scale image $y = (y(x))_{x \in \Omega}$ defined over a bounded domain $\Omega \subset \mathbb{R}^2$, and $y(x) \in \mathbb{R}_+$ is the noisy observed intensity at pixel $x \in \Omega$. The weighting associated to the patch P is computed from the steering kernel:

$$W_P(i, j) = \frac{\sqrt{\det(C_j)}}{\sigma^2} \exp -\frac{1}{2} \left(\frac{\|y(x) - y(x_i)\|}{\sigma^2} - \sqrt{2n-1} \right)^2, \quad (14)$$

where $\|\cdot\|$ denotes the Euclidean distance. $y(x) := (y(x_k), y_k \in B(x)) \in \mathbb{R}^n$ is a vectorized image patch. $B(x)$ is a $\sqrt{n} \times \sqrt{n}$ neighborhood centered at pixel x ($n = 7$). $\Delta(x)$ is a square neighborhood of $N = |\Delta(x)|$ pixels. $y(x_i)$ is a vectorized image patch such that $x_i \in \Delta(x)$. σ^2 is the noise variance assumed to be known or estimated. The final estimate is given by

$$\text{INL}_{\sigma, n} y(x) = \frac{\sum_P W_P(i, j) y(x_i)}{\sum_P W_P(i, j)}. \quad (15)$$

The algorithm is divided in two identical separate steps, the image is scanned pixel per pixel. Let us denote by P the current reference patch which size is $n \times n$ (with $k = 5$) and x_r the current central pixel of P . The loop on the image is done on x_r .

This approach has two important benefits. On the one hand, it allows finding more similar patches with the same pattern but with different mean level compensating intensity inhomogeneities typically present on MRI data, and on

the other hand, overestimation of the noise variance will be minimized in cases with unique patches in the search volume. Thus, the adaptive filter proposed will set the parameter h^2 equal to the minimum distance estimation as described in (11).

3. Experiments and Results

To evaluate and compare the proposed method with state-of-the-art methods, we did experiments on both synthetic and real medical images. To conduct the experiments on synthetic data, we use the standard MR images phantom of the brain obtained from the BrainWeb database [18]. The proposed algorithm was compared with the following recently proposed methods.

- (1) *NL-means: Nonlocal Means Image Denoising Method* [12]. The size of the patch and research window depend on the value of σ . The search window size used for experiments was $9 \times 9 \times 9$, neighborhood size was $3 \times 3 \times 3$, and value of the decay parameter h and σ were 0.4σ , 20 .
- (2) *NL-PCA: Nonlocal Principal Component Analysis Method* [19]. Local neighborhood size used for the experiments was $3 \times 3 \times 3$. Other parameters are fixed to $\alpha = 2.1$; $K^{\text{hard}} = K^{\text{wien}} = 3$; $n^{\text{hard}} = n^{\text{wien}} = 15$.
- (3) *DCT: Local Discrete Cosine Transform Method* [20]. The method decomposes the image into local patches, and denoises the patches with thresholding estimate in the DCT domain. The local patches of size used for the experiments was $\sqrt{N} = 16 \times 16$.
- (4) *Proposed Method*. The search window n for the experiments was $5 \times 5 \times 5$.

For quantitative analysis of the denoising methods, we used the peak signal-to-noise ratio (PSNR), the structural similarity index matrix (SSIM).

Figure 3 displays the results of the image denoised with NL-means, NL-PCA, DCT, and proposed method. This experiment was conducted on the 2D slice of the synthetic images of the brain in the 3D environment after corrupting the image by uniform Rician noise with $\sigma = 20$. The proposed filter was executed using a neighborhood size for denoising as $13 \times 13 \times 13$ and a neighborhood size for the local computation of range as $5 \times 5 \times 5$. It can be observed from Figure 3 that the image denoised with the proposed method is closer to the original image than the images denoised with other approaches. The graph in Figure 4 shows the quantitative analysis of the proposed method with other recently proposed methods based on the similarity measures PSNR, MSSIM, respectively. This experiment was also conducted on the BrainWeb MR image with σ of the noise ranging from 10 to 30. All the methods with which the proposed method was compared are based on the Rician noise model. In the quantitative analysis, the background was excluded; that is, only the area of the image inside the skull was considered. It can be seen from the graph that the performance of the proposed method is best for each similarity measure.

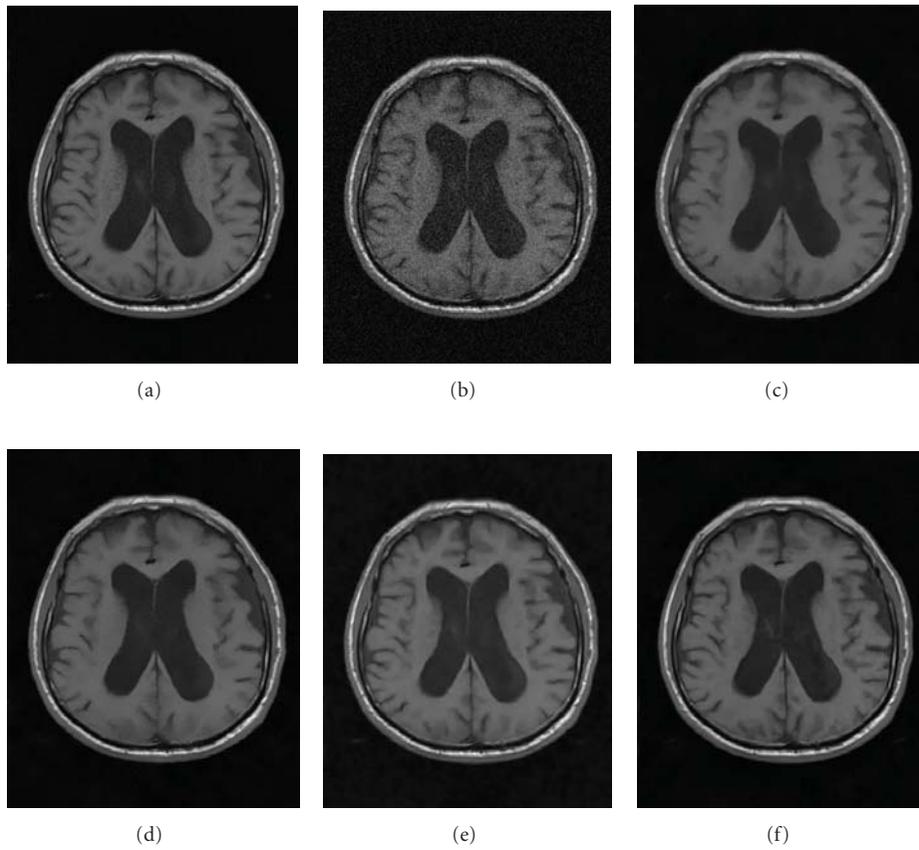


FIGURE 3: Denoising MRI with several methods: (a) original image; (b) original image corrupted by Rician noise of $\sigma = 20$; (c) denoised with NL-means method; (d) denoised with DCT method; (e) denoised with NL-PCA method; (f) denoised with proposed method.

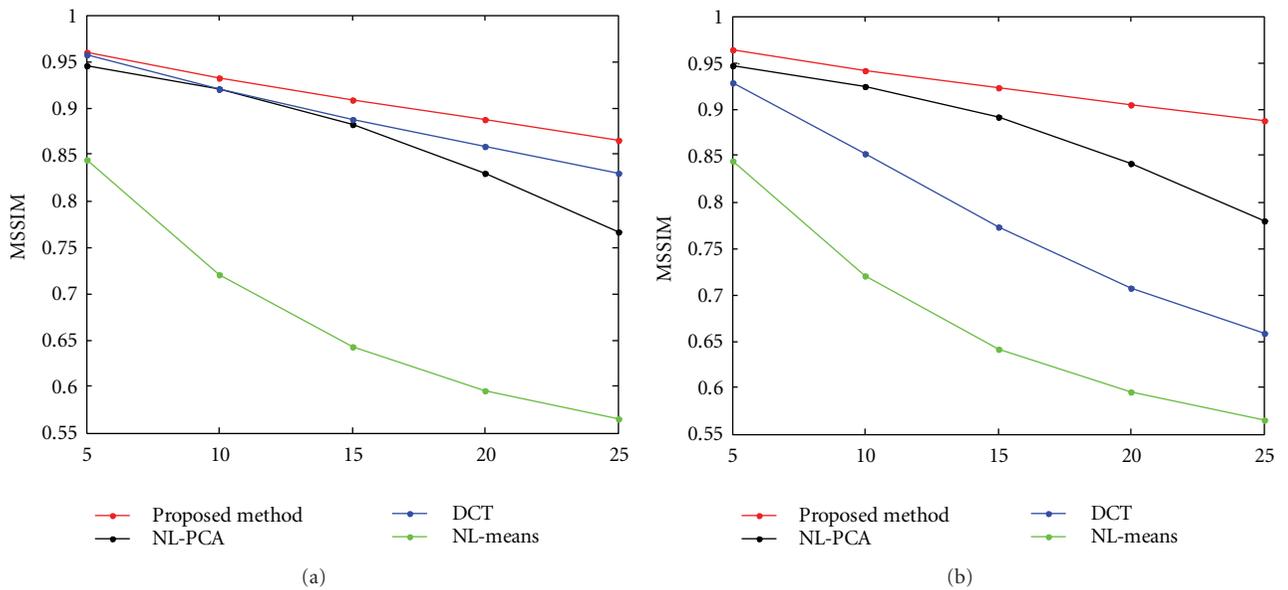


FIGURE 4: Comparative analysis of the proposed method with other methods based on PSNR, MSSIM for different values of the noise standard deviation.

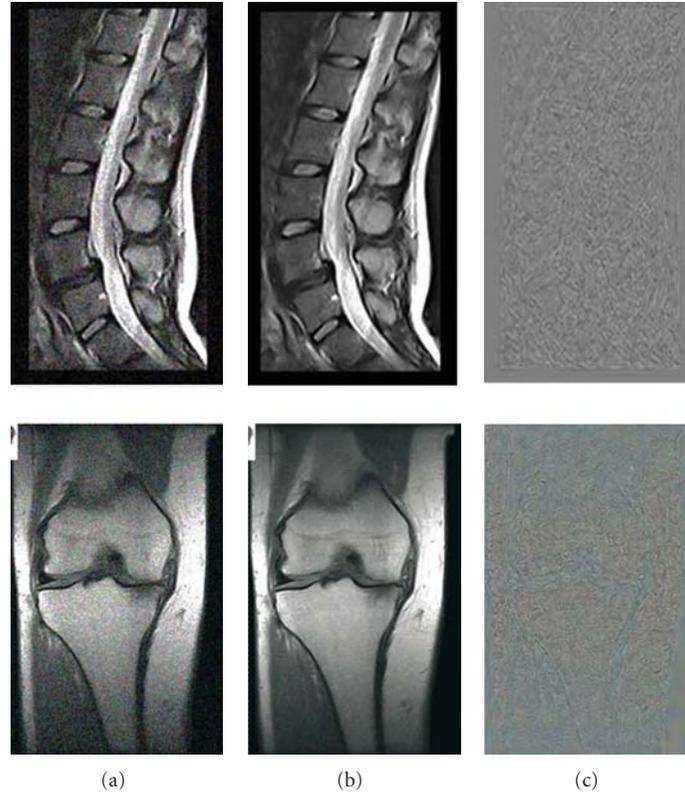


FIGURE 5: The denoising results obtained with the proposed filter. (a) The original noisy images ($\sigma = 30$); (b) the denoised images (PSNR is 38.9 and 37.7, resp.); (c) the differences of the images.

Figure 5 shows the extremely noisy data and we use the proposed method to remove the noise. The absolute value of the residuals of the filtering process clearly show the capabilities of the proposed approach on the extremely noisy data.

4. Conclusion

A new method to denoise the medical images by applying NL-means method is proposed in this paper. To demonstrate the efficiency of the proposed method, experiments were conducted on both simulated and real medical images. Comparative analysis with other recently proposed methods based on the similarity measures, PSNR, MSSIM, proves that the proposed method is superior to them in terms of image quality.

Acknowledgments

This work was supported by the NSFC (Grant nos. 61103130, 61070120, 60982145); National Program on Key basic research Project (973 Programs) (Grant nos. 2010CB731804-1, 2011CB706901-4); Beijing Natural Science Foundation (Grant no. 4112021); Foundation of Beijing Educational Committee (no. KM201111417015); the opening project of Shanghai key laboratory of integrate administration technologies for information security (no. AGK2010005).

References

- [1] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1522–1531, 2000.
- [2] A. Pižurica, W. Philips, I. Lemahieu, and M. Achery, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 545–557, 2002.
- [3] L. Zhang, P. Bao, and X. Wu, "Hybrid inter- and intra-wavelet scale image restoration," *Pattern Recognition*, vol. 36, no. 8, pp. 1737–1746, 2003.
- [4] L. Zhang, P. Bao, and X. Wu, "Multiscale LMMSE-based image denoising with optimal wavelet selection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 4, pp. 469–481, 2005.
- [5] A. Pižurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 654–665, 2006.
- [6] J. L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.
- [7] G. Y. Chen and B. Kégl, "Image denoising with complex ridgelets," *Pattern Recognition*, vol. 40, no. 2, pp. 578–585, 2007.
- [8] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

- [9] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [10] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Transactions on Image Processing*, vol. 16, no. 5, pp. 1395–1411, 2007.
- [11] D. Barash, "A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 844–847, 2002.
- [12] 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.
- [13] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2866–2878, 2006.
- [14] 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, 15 pages, 2012.
- [15] 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.
- [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.
- [17] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, 2007.
- [18] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, Article ID 4359947, pp. 425–441, 2008.
- [19] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, vol. 43, no. 4, pp. 1531–1549, 2010.
- [20] G. Yu and G. Sapiro, "DCT image denoising: a simple and effective image denoising algorithm," *Image Processing On Line*, 2011.

Research Article

Solid Dynamic Models for Analysis of Stress and Strain in Human Hearts

Qiu Guan,^{1,2} Miaomiao Lu,² Xiaoyan Wang,¹ and Cao Jiang²

¹ College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

² College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

Correspondence should be addressed to Qiu Guan, gq@zjut.edu.cn

Received 22 September 2011; Accepted 25 October 2011

Academic Editor: Carlo Cattani

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

This paper proposes a solid model based on four-dimensional trivariate B-spline for strain and stress analysis of ventricular myocardium. With a series of processing steps in the four-dimensional medical images, the feature points of ventricular inner and outer wall are obtained. A B-spline surface is then used to build the dynamic deformation model of the myocardial walls. With such a surface model, a hexahedron control mesh can be constructed by sweeping the cloud data, and the ventricular solid model is built by fitting the trivariate B-spline parameters. Based on these models, a method of isogeometric analysis can be applied to calculate the stress and strain continuously distributed in the ventricle. The model is represented smoothly in the cylindrical coordinate system and is easy to measure myocardium dynamics for finding abnormal motion. Experiments are carried out for comparing the stress and strain distribution. It is found that the solid model can determine ventricular dynamics which can well reflect the deformation distribution in the heart and imply early clues of cardiac diseases.

1. Introduction

Cardiovascular diseases are currently the leading cause of death in the world, and the rate of death is increasing each year in many countries. Therefore, more and more physiologists and researchers make efforts to understand how the heart works and how to diagnosis and treat the heart diseases. Dynamics and kinetics of the left cardiac ventricle are the primary representation of the cardiac motion, as a series of systolic and diastolic motions of the left cardiac ventricle make the heart pump blood to circulate the whole body. Furthermore, the stress and strain express the characteristics of elasticity and motion of myocardial walls [1].

With assist of computer medical imaging technology, such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography, single photon emission computed tomography, ultrasound, and X-ray, these imaging techniques can give different representation of the heart. In this paper, the endocardium and epicardium shapes of a ventricle are constructed with a

segmentation technique in MRI images. For constructing model, finite element (FE) models are the most commonly used in biomechanics. Especially, hexahedral and tetrahedral FE models are popular in the representation of cardiac model, for example, Figure 1 [2, 3]. Comparing with the FE model, left ventricular CAD (computer-aided design) model is more representative of the true ventricle in the shape, such as B-spline surface model. Its continuity and smoothness are better than the FE model. However, it is difficult to apply FE analysis directly to the CAD model [4].

The appearance of isogeometric analysis makes it possible to carry out mechanical analysis directly by a CAD model. As a new computational technique, isogeometric analysis improves on and generalizes the standard FE method and has been proven to be a powerful method exceeding the FE method [5]. Inspired by isogeometric analysis, this paper constructs a solid model based on trivariate B-spline and calculates the strain and stress of a ventricle by this model. Finally, it shows the distribution of the strain and stress on the solid model.

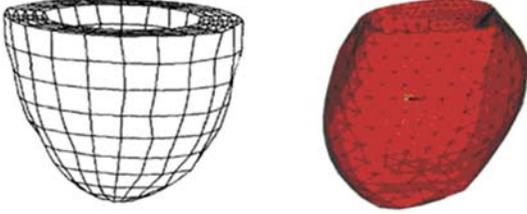


FIGURE 1: Finite element models of left ventricle in the literature [2, 3].

2. Model Representation

2.1. *B-Spline Curves and Surfaces.* One can get a B-spline curve of order k . By [6]

$$C(t) = \sum_{i=0}^n p_i N_{i,k}(t), \quad t_k \leq t < t_{k+1}, \quad (1)$$

where p_0, p_1, \dots, p_n are the control points of $C(t)$, the linear interpolation of control points is called control polygon. $N_{i,k}(u)$ is B-spline basis function of degree k , which is determined by a group of nondecreasing normalized sequence T : $t_0 \leq t_1 \leq \dots \leq t_{n+k+1}$. The sequence can be determined by Riesenfeld method and Hartley-Judd method [7].

The definition of $N_{i,k}(u)$ can be expressed as

$$N_{i,0}(t) = \begin{cases} 1, & \text{if } t_i \leq t < t_{i+1}, \\ 0, & \text{else} \end{cases}$$

$$N_{i,k}(t) = \frac{t - t_i}{t_{i+k} - t_i} N_{i,k-1}(t) + \frac{t_{i+k+1} - t}{t_{i+k+1} - t_{i+1}} N_{i+1,k-1}(t), \quad (2)$$

$$\text{define } \frac{0}{0} = 0.$$

With a tensor product, given a $(n+1) \times (m+1)$ control mesh $p_{i,j}$ and knot vector T : $t_0 \leq t_1 \leq \dots \leq t_{n+k+1}$, a B-spline surface of degree k_1 in the t direction and degree k_2 in the w direction is defined as

$$S(t, w) = \sum_{i=0}^n \sum_{j=0}^m p_{i,j} N_{i,k_1}(t) N_{j,k_2}(w), \quad (3)$$

where S is a point on the defined surface, and t and w are usually representing longitude and latitude; respectively, $N_{i,k_1}(t)$ and $N_{j,k_2}(w)$ are the basis functions in the t and w directions with a degree of k_1 and k_2 .

2.2. *Volumetric B-Splines.* A volumetric B-spline can also be defined by

$$V(t, w, u) = \sum_{i=0}^n \sum_{j=0}^m \sum_{k=0}^l p_{i,j,k} N_{i,k_1}(t) N_{j,k_2}(w) N_{k,k_3}(u). \quad (4)$$

Comparing with the model in Cartesian coordinates, a cylindrical coordinate model is better approximate to the shape of the heart [8]. Volumetric B-spline is the extension of

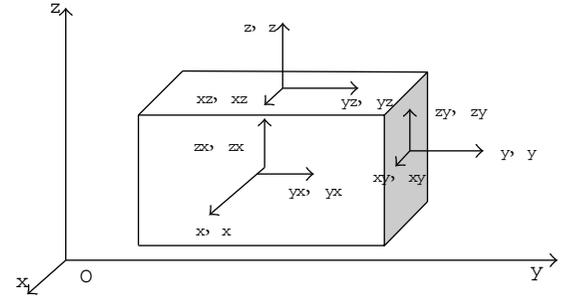


FIGURE 2: Three-dimensional stress and strain components.

B-spline surface, which can be given by the following two steps in cylindrical coordinates.

(1) Coordinate transform. A $(n+1) \times (m+1) \times (l+1)$ control mesh $p_{i,j,l}$ with the points $(x_{ijk}, y_{ijk}, z_{ijk})$ in Cartesian coordinate is transformed into cylindrical coordinate by

$$r_{ijk} = \sqrt{x_{ijk}^2 + y_{ijk}^2},$$

$$\theta_{ijk} = \begin{cases} \pi/2 & \text{if } x_{ijk} = 0, y_{ijk} > 0, \\ 3\pi/2 & \text{if } x_{ijk} = 0, y_{ijk} < 0, \\ a \tan\left(\frac{y_{ijk}}{x_{ijk}}\right) & \text{if } x_{ijk} > 0, y_{ijk} \geq 0, \\ \pi + a \tan\left(\frac{y_{ijk}}{x_{ijk}}\right) & \text{if } x_{ijk} < 0, \\ 2\pi + a \tan\left(\frac{y_{ijk}}{x_{ijk}}\right) & \text{if } x_{ijk} > 0, y_{ijk} < 0, \end{cases} \quad (5)$$

$$z_{ijk} = z_{ijk}.$$

(2) The representation of volumetric B-spline in cylindrical coordinate is

$$r(t, w, u) = \sum_{i=0}^n \sum_{j=0}^m \sum_k^l r_{i,j,k} N_{i,k_1}(t) N_{j,k_2}(w) N_{k,k_3}(u),$$

$$\theta(t, w, u) = \sum_{i=0}^n \sum_{j=0}^m \sum_k^l \theta_{i,j,k} N_{i,k_1}(t) N_{j,k_2}(w) N_{k,k_3}(u), \quad (6)$$

$$z(t, w, u) = \sum_{i=0}^n \sum_{j=0}^m \sum_k^l z_{i,j,k} N_{i,k_1}(t) N_{j,k_2}(w) N_{k,k_3}(u).$$

3. Determination of Stress and Strain

3.1. *Elasticity.* Stress and strain actually reflect elasticity of ventricular myocardial walls [9]. Calculations of the stress and strain can be inspired from the idea of elasticity theory. In 3D space, set u, v, w as the displacements in x, y, z direction. Strain vector and stress vector are $\sigma = [\sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{yz}, \tau_{xz}]^T$ and $\varepsilon = [\varepsilon_x, \varepsilon_y, \varepsilon_z, \gamma_{xy}, \gamma_{yz}, \gamma_{xz}]^T$, respectively, as shown in Figure 2.

The relation between strain and displacement is

$$\varepsilon = \begin{cases} \varepsilon_x = \frac{\partial u}{\partial x}, \\ \varepsilon_y = \frac{\partial v}{\partial y}, \\ \varepsilon_z = \frac{\partial w}{\partial z}, \\ \gamma_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}, \\ \gamma_{yz} = \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}, \\ \gamma_{xz} = \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}. \end{cases} \quad (7)$$

Physical relationship or the relation between strain and stress is

$$\sigma = D\varepsilon, \quad (8)$$

where D is the elastic matrix defined as

$$D = \frac{E(1-\mu)}{(1+\mu)(1-2\mu)} \times \begin{bmatrix} 1 & \frac{\mu}{1-\mu} & \frac{\mu}{1-\mu} & 0 & 0 & 0 \\ \frac{\mu}{1-\mu} & 1 & \frac{\mu}{1-\mu} & 0 & 0 & 0 \\ \frac{\mu}{1-\mu} & \frac{\mu}{1-\mu} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & H & 0 & 0 \\ 0 & 0 & 0 & 0 & H & 0 \\ 0 & 0 & 0 & 0 & 0 & H \end{bmatrix}, \quad (9)$$

where E is elasticity modulus, μ is Poisson's ratio, and $H = (1-2\mu)/2(1-\mu)$. With (9), strain and stress can be determined.

3.2. Isogeometric Analysis. We regard B-spline basis functions as the displacement function and also the basis function of the ventricular shape, to calculate the stress and strain. Solid models at adjacent time points can be reconstructed by (4) with corresponding control points. The continuous displacements of the solid model correspond the displacements of control points, that is,

$$\text{Disp}(t_c, t_l, t_r) = \sum_{i=0}^n \sum_{j=0}^m \sum_{k=0}^l d_{i,j,l} N_{i,k_1}(t_c) N_{j,k_2}(t_l) N_{k,k_3}(t_r), \quad (10)$$

where $d_{i,j,l}$ is the displacements of control points of $\text{Disp}(t_c, t_l, t_r)$. n, m, l define the control net, t_c, t_l , and t_r are the knot vectors in the directions of circumference, long axis and radius, respectively. For X, Y, Z directions, the displacements

DX, DY, DZ are defined as

$$\begin{aligned} DX(t_c, t_l, t_r) &= \sum_{i=0}^n \sum_{j=0}^m \sum_{k=0}^l d_{i,j,l}^x N_{i,k_1}(t_c) N_{j,k_2}(t_l) N_{k,k_3}(t_r), \\ DY(t_c, t_l, t_r) &= \sum_{i=0}^n \sum_{j=0}^m \sum_{k=0}^l d_{i,j,l}^y N_{i,k_1}(t_c) N_{j,k_2}(t_l) N_{k,k_3}(t_r), \\ DZ(t_c, t_l, t_r) &= \sum_{i=0}^n \sum_{j=0}^m \sum_{k=0}^l d_{i,j,l}^z N_{i,k_1}(t_c) N_{j,k_2}(t_l) N_{k,k_3}(t_r), \end{aligned} \quad (11)$$

where $d_{xi,j,l}, d_{yi,j,l}, d_{zi,j,l}$ are the displacements of control points in X, Y, Z directions.

By elasticity theory, the strain of one point in the model can be calculated

$$\begin{aligned} \varepsilon_x &= \frac{\partial u}{\partial x} = \frac{\partial DX}{\partial x} = \frac{\partial N}{\partial x} dx, \\ \varepsilon_y &= \frac{\partial v}{\partial y} = \frac{\partial DY}{\partial y} = \frac{\partial N}{\partial y} dy, \\ \varepsilon_z &= \frac{\partial w}{\partial z} = \frac{\partial DZ}{\partial z} = \frac{\partial N}{\partial z} dz, \\ \gamma_{xy} &= \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = \frac{\partial DX}{\partial y} + \frac{\partial DY}{\partial x} = \frac{\partial N}{\partial y} dx + \frac{\partial N}{\partial x} dy, \\ \gamma_{yz} &= \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} = \frac{\partial DY}{\partial z} + \frac{\partial DZ}{\partial y} = \frac{\partial N}{\partial z} dy + \frac{\partial N}{\partial y} dz, \\ \gamma_{xz} &= \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} = \frac{\partial DX}{\partial z} + \frac{\partial DZ}{\partial x} = \frac{\partial N}{\partial z} dx + \frac{\partial N}{\partial x} dz, \end{aligned} \quad (12)$$

where $N = N_{i,k_1}(t_c) N_{j,k_2}(t_l) N_{k,k_3}(t_r)$ is the B-spline basis function, and dx, dy , and dz are displacement values of the corresponding points.

Then the derivatives of B-spline basis functions in X, Y, Z directions can be derived in each parameter direction

$$\begin{aligned} \frac{\partial N}{\partial t_c} &= (k_1 - 1) \left[\frac{N_{i,k_1-1}(t_c)}{t_{c,i+k_1-1} - t_{c,i}} - \frac{N_{i+1,k_1-1}(t_c)}{t_{c,i+k_1} - t_{c,i+1}} \right] \\ &\quad \times N_{j,k_2}(t_l) N_{k,k_3}(t_r), \\ \frac{\partial N}{\partial t_l} &= (k_2 - 1) \left[\frac{N_{j,k_2-1}(t_l)}{t_{l,j+k_2-1} - t_{l,j}} - \frac{N_{j+1,k_2-1}(t_l)}{t_{l,j+k_2} - t_{l,j+1}} \right] \\ &\quad \times N_{i,k_1}(t_c) N_{k,k_3}(t_r), \\ \frac{\partial N}{\partial t_r} &= (k_3 - 1) \left[\frac{N_{k,k_3-1}(t_r)}{t_{r,k+k_3-1} - t_{r,i}} - \frac{N_{k+1,k_3-1}(t_r)}{t_{r,k+k_3} - t_{r,k+1}} \right] \\ &\quad \times N_{j,k_2}(t_l) N_{i,k_1}(t_r). \end{aligned} \quad (13)$$

To get the derivatives of B-spline basis functions in X, Y, Z directions, the partial derivative transform of isoparametric

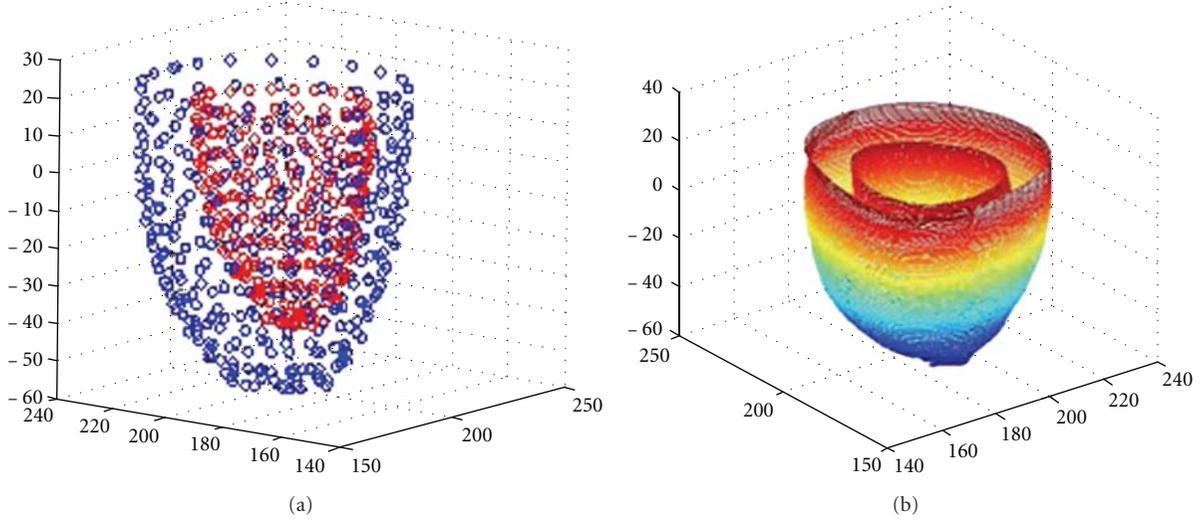


FIGURE 3: (a) Ventricular point distribution model (b) B-spline surface model.

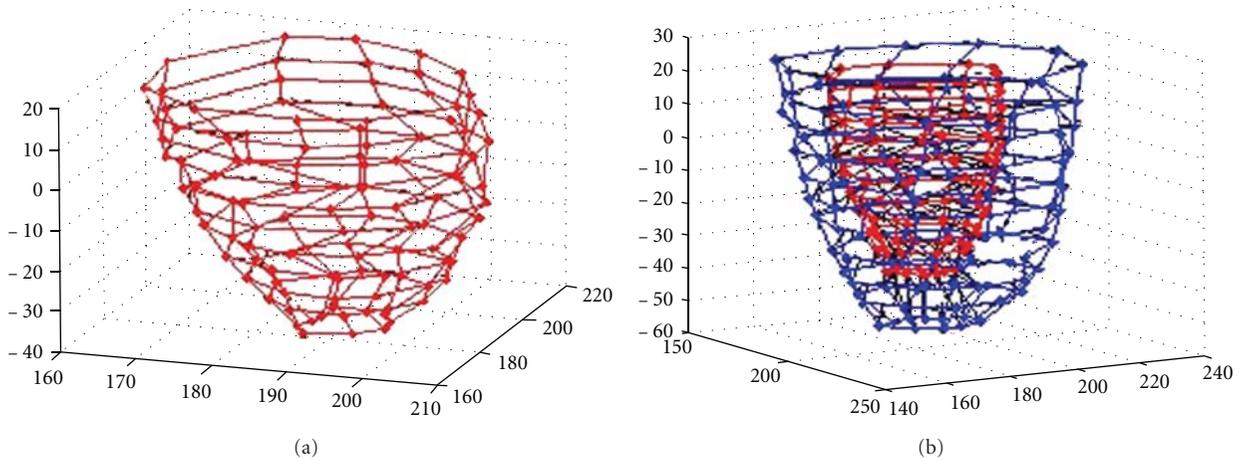


FIGURE 4: (a) Left ventricular inside wall quadrilateral mesh, (b) hexahedral control mesh.

principles [10] is used

$$\begin{bmatrix} \frac{\partial N}{\partial x} \\ \frac{\partial N}{\partial y} \\ \frac{\partial N}{\partial z} \end{bmatrix} = J^{-1} \begin{bmatrix} \frac{\partial N}{\partial t_c} \\ \frac{\partial N}{\partial t_l} \\ \frac{\partial N}{\partial t_r} \end{bmatrix}, \quad (14)$$

where J is Jacobian matrix as follows:

$$J = \begin{bmatrix} \frac{\partial x}{\partial t_c} & \frac{\partial y}{\partial t_c} & \frac{\partial z}{\partial t_c} \\ \frac{\partial x}{\partial t_l} & \frac{\partial y}{\partial t_l} & \frac{\partial z}{\partial t_l} \\ \frac{\partial x}{\partial t_r} & \frac{\partial y}{\partial t_r} & \frac{\partial z}{\partial t_r} \end{bmatrix}. \quad (15)$$

Here, take $\frac{\partial x}{\partial t_c}$ as an example. $p_{i,j,l}^x$ is the control point coordinate in the X direction.

$$\frac{\partial x}{\partial t_c} = \sum_{i=0}^m \sum_{j=0}^l N_{j,k_2}(t_l) N_{k,k_3}(t_r) \sum_{i=0}^n N_{i,k_1-1}(t_c) \frac{p_{i,j,l}^x - p_{i-1,j,l}^x}{t_{c,i+k_1-1} - t_{c,i}}. \quad (16)$$

Therefore, the strain can be calculated in this way. By (8), the stress condition can also be calculated.

4. Experiments and Results

4.1. Construction of Ventricular Models. Both inside and outside data points of ventricular myocardial walls can be obtained from 3D medical images [11, 12]. Figure 3(a) shows the points obtained by model-based segmentation [13]. Figure 3(b) is the corresponding B-spline surface model.

A sweeping method [14] can be used to obtain the control hexahedral, for example, the one in Figure 4, where

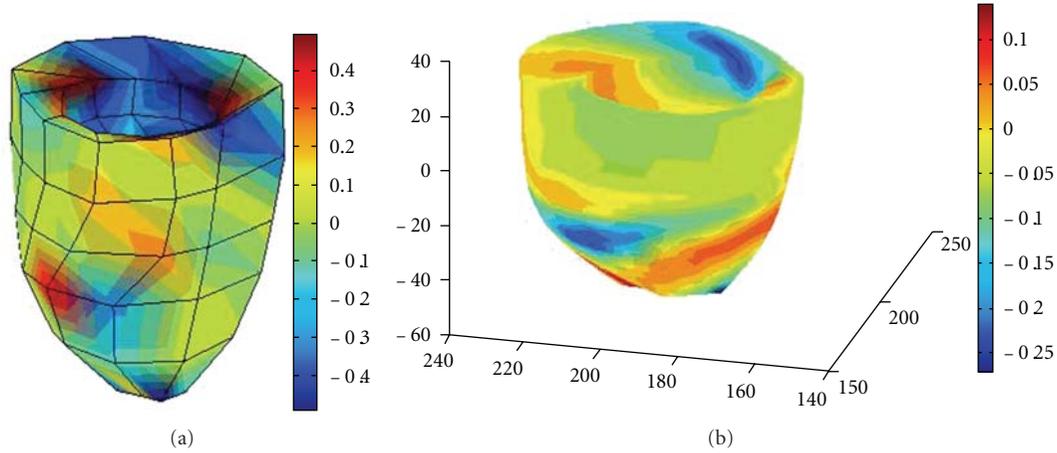


FIGURE 5: Principal strain distribution in X direction: (a) the results by finite element model, (b) result by B-spline solid model.

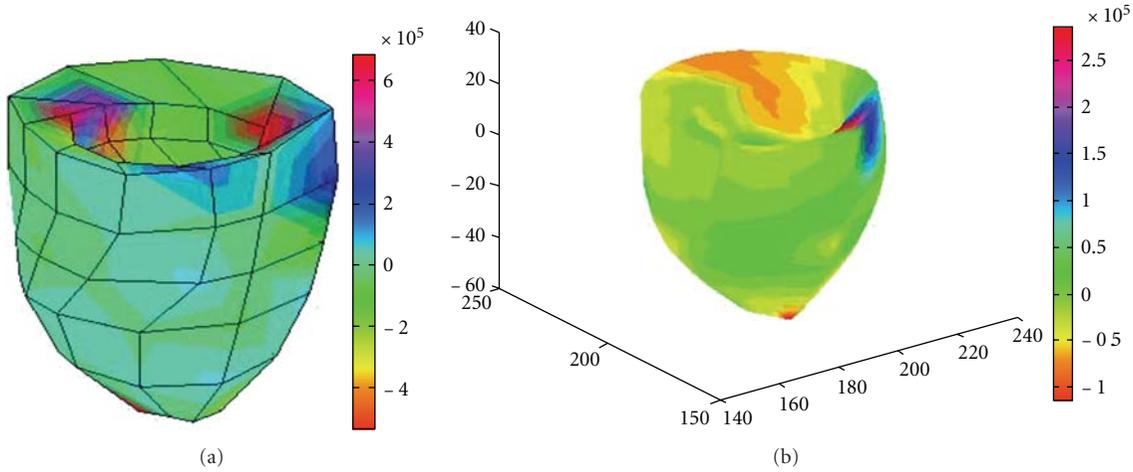


FIGURE 6: Principal stress distribution in X direction: (a) the results by finite element model, (b) result by B-spline solid model.

Figure 4(a) is the quadrilateral mesh of inside wall, and Figure 4(b) is the hexahedral mesh. While control hexahedral mesh is obtained, the corresponding ventricular B-spline solid model can be reconstructed by (4) and (5).

4.2. Stress and Strain. Stress and strain of a ventricle are calculated based on the steps described in the above sections. Here, for calculation of the stress results, we set myocardial elastic modulus 11 Kpa and Poisson’s ratio of 0.49. Figure 5 shows the principal strain distribution in X direction, and Figure 6 shows the principal stress distribution. In these figures, the left one shows the result by the finite element model, and the right shows that by the B-spline solid model. The color represents the change in stress or strain distribution, with the specific reference of the color bar. From the results, it can be seen that the stress and strain of the left ventricular model show overall nonuniform distribution [15, 16], which is consistent with the results by other researchers [17].

5. Conclusion

With the situation that traditional finite element methods are difficult for direct use in mechanics analysis which has continuous distribution in space, this paper proposes a solid B-spline model to construct a continuous ventricular mechanical model and applies isogeometric analysis. Stress and strain calculative formulas are derived. The proposed model features continuous, smooth, and inseparable. According to a set of ventricular hexahedral solid B-spline models sampled at different times in a cardiac cycle, strain and stress are determined for medical analysis.

Acknowledgments

Thanks are due to Z. Xu, T. T. Jiang, Y. Q. Xu, Y. H. Du, et al. for their helpful work and laboratory implementation in the project during their study in the university. This work was supported in part by the National Natural Science Foundation of China (NSFC nos. 61103140 and 61105073) and Zhejiang Provincial S&T Department (2010R10006 and 201-0C33095).

References

- [1] S. Chen, J. Zhang, H. Zhang et al., "Myocardial motion analysis for determination of tei-index of human heart," *Sensors*, vol. 10, no. 12, pp. 11428–11439, 2010.
- [2] G. H. Wu, F. Liu, L. Xia, and W. X. Lu, "Construction of three dimensional composite finite element mechanical model of human left ventricle," *Chinese Journal of Biomedical Engineering*, vol. 21, no. 5, pp. 404–410, 2002.
- [3] F. Liu, J. You, L. Guo et al., "Strain analysis of left ventricle by integrating finite element model and non-rigid motion estimation," *Journal of Computer-Aided Design and Computer Graphics*, vol. 19, no. 8, pp. 1028–1033, 2007.
- [4] J. A. Cottrell, T. J. R. Hughes, and Y. Bazilevs, *Isogeometric Analysis: Toward Intergration of CAD and FEA [M]*, Wiley, 2009.
- [5] C. de Falco, A. Reali, and R. V'azquez, "GeoPDEs: a research tool for IsoGeometric Analysis of PDEs," <http://geopdes.sourceforge.net>.
- [6] T. T. Jiang, S. Y. Chen, and Y. Xu, "3-D representation and volumetric measurement of human heart from a cylindrical B-spline surface model," in *Proceedings of the International Conference on BioMedical Engineering and Informatics*, pp. 765–769, Sanya, China, May 2008.
- [7] W. P. Segars, D. S. Lalush, E. C. Frey, D. Manocha, M. A. King, and B. M. W. Tsui, "Improved dynamic cardiac phantom based on 4D NURBS and tagged MRI," *IEEE Transactions on Nuclear Science*, vol. 56, no. 5, Article ID 5280500, pp. 2728–2738, 2009.
- [8] T.-T. Jiang, S. Chen, and Y. Xu, "3-D representation and volumetric measurement of human heart from a cylindrical B-spline surface model," *BioMedical Engineering and Information*, vol. 1, pp. 765–769, 2008.
- [9] Q. Guan, S. Chen, Y. Du, and W. Wang, "Strain and stress analysis of cardiac ventricles from 4D images," *Journal of Computer-Aided Design and Computer Graphics*, vol. 22, no. 12, pp. 2221–2236, 2010.
- [10] D. J. Benson, Y. Bazilevs, E. de Luycker et al., "A generalized finite element formulation for arbitrary basis functions: from isogeometric analysis to XFEM," *International Journal for Numerical Methods in Engineering*, vol. 83, no. 6, pp. 765–785, 2010.
- [11] S. Y. Chen and Y. F. Li, "Determination of stripe edge blurring for depth sensing," *IEEE Sensors Journal*, vol. 11, no. 2, Article ID 5585653, pp. 389–390, 2011.
- [12] S. Y. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.
- [13] H. C. van Assen and M. G. Danilouchkine, "et. Segmentation method for cardiac region in CT images based on active shape model," in *Proceedings of the International Conference on Control, Automation and Systems*, vol. 10, pp. 2074–2077, 2010.
- [14] J. Zhai, X. Qiao, and Q. Ding, "Research and implementation of hexahedral mesh generation algorithm based on sweeping," *Journal of Nanjing University of Aeronautics and Astronautics*, vol. 39, no. 1, pp. 71–74, 2007.
- [15] S. Y. Chen, "Cardiac deformation mechanics from 4D images," *Electronics Letters*, vol. 43, no. 11, pp. 609–611, 2007.
- [16] T. T. Jiang, S. Y. Chen, and Y. Xu, "3-D representation and volumetric measurement of human heart from a cylindrical B-Spline surface model," in *Proceedings of the 1st International Conference on BioMedical Engineering and Informatics (BMEI '08)*, pp. 765–769, May 2008.
- [17] Q. Guan, Y. Du, S. Y. Chen, and F. Wang, "Calculation of Tei index and LV motion amplitude based on cardiac physical phase," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '09)*, vol. 12, pp. 1027–1031, December 2009.

Research Article

Characteristics of Evoked Potential Multiple EEG Recordings in Patients with Chronic Pain by Means of Parallel Factor Analysis

Juan Wang,¹ Xiaoli Li,^{1,2} Chengbiao Lu,¹ Logan J. Voss,³
John P. M. Barnard,³ and Jamie W. Sleight³

¹Institute of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China

²National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 10088, China

³Department of Anesthesia, Waikato Hospital, Hamilton 3204, New Zealand

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

Received 29 September 2011; Accepted 25 October 2011

Academic Editor: Sheng-yong Chen

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

This paper presents an alternative method, called as parallel factor analysis (PARAFAC) with a continuous wavelet transform, to analyze of brain activity in patients with chronic pain in the time-frequency-channel domain and quantifies differences between chronic pain patients and controls in these domains. The event related multiple EEG recordings of the chronic pain patients and non-pain controls with somatosensory stimuli (pain, random pain, touch, random touch) are analyzed. Multiple linear regression (MLR) is applied to describe the effects of aging on the frequency response differences between patients and controls. The results show that the somatosensory cortical responses occurred around 250 ms in both groups. In the frequency domain, the neural response frequency in the pain group (around 4 Hz) was less than that in the control group (around 5.5 Hz) under the somatosensory stimuli. In the channel domain, cortical activation was predominant in the frontal region for the chronic pain group and in the central region for controls. The indices of active ratios were statistical significant between the two groups in the frontal and central regions. These findings demonstrate that the PARAFAC is an interesting method to understanding the pathophysiological characteristics of chronic pain.

1. Introduction

Chronic pain is a complex disease characterized by pain persisting after damage or pathology has healed. Effective treatment of chronic pain is hampered by an incomplete understanding of the pathophysiological changes that occur in the nervous system of chronic pain sufferers. The electroencephalogram (EEG) records the electrical activity from the scalp produced by the firing of neurons within the cerebral cortex [1] and has been widely used to analyze neural activity in chronic pain subjects [2, 3].

To investigate the physiological basis of chronic pain, event-related potentials (ERPs) have been used to explore pain-related modulation of the latency, location, amplitude, and frequency of evoked EEG responses to sensory stimulation [4–6]. Previous studies have shown that frequency and time domain characteristics of EEG recordings from certain brain regions are altered with chronic pain [7–9].

Traditionally, the ERP components are analyzed at specific cortical locations, that is, the vertex and frontotemporal region (e.g., [10, 11]). With this approach, however, the information provided is limited to the particular region under investigation and neglects the importance of wider cortical regions in information processing [12].

Multiple EEG recordings can be used to characterise the electrical activity across the whole cortex. Traditional methods of PCA and ICA analysis process the multiple EEG signals in two-way domains such as time channel and frequency channel [13]. To effectively characterise multiple EEG signals, development of an analysis method that captures time-frequency-channel information is required.

In this study, we have investigated differences in multiple ERPs between chronic pain patients and pain-free individuals. To characterise the EEG signals in the time-frequency-channel domain, a parallel factor analysis (PARAFAC) method with wavelet transforms was developed to

decompose the multiple EEG recordings. The PARAFAC method has been successfully employed to detect abnormal EEG activity in neurological diseases such as epilepsy and Alzheimer's disease [14]. Herein, this novel method is employed to analyze the EEG in chronic pain subjects. The results show differences in cortical evoked activity between chronic pain and pain-free individuals and indicate that the PARAFAC is an effective method for extracting the characteristics of multiple EEG recordings in the time-frequency-channel domain.

2. Materials and Methods

2.1. Subjects. Subjects were 13 chronic pain patients recruited through the Waikato Hospital Pain Clinic and 13 pain-free volunteers. A range of conditions were represented in the patient group, including chronic lower back pain, neck pain, abdominal pain, and throat pain. None of the subjects in either group had a history of other neurological disease or head injury. Ethical approval was obtained from the Waikato Ethics Committee, and all subjects signed written informed consent.

2.2. Sensory Stimulations. The stimuli consisted of brief (10 ms), repetitive (at least 120) electrical shocks delivered to the dominant index finger. The electrodes were positioned on the dorsal aspect of the distal interphalangeal joint and a fine wire with a soldered tip, applied to the pulp of the fingertip. The stimulus intensity was recorded as the percent maximum voltage and rated by each subject on a 1–10 analogue scale. Two intensities of electric shock were tailored to each subject, one that was easily felt but not painful (from here on referred to as the “touch” stimulus) and one that was rated as “moderately painful” (the “pain” stimulus). The pain stimulus was felt as a sharp pricking sensation, predominantly under the wire electrode on the finger pulp. The shocks were given in three sequences at a constant frequency of 1 every 1.5 s as follows: (1) 120 sequential touch shocks (“touch”); (2) 120 sequential painful shocks (“pain”); (3) a random sequence of 300 touch and painful shocks at a 4 : 1 ratio. For the randomized protocol, the touch and pain stimuli were analyzed separately (“random touch” and “random pain”, resp.). Stimulus intensity and delivery were controlled by MatLab software (Matlab 6.0 Mathworks, Natick, MA, USA) running on a laptop computer that interfaced directly with the stimulus generator.

2.3. EEG Recording and Experimental Protocol. The subjects were comfortably seated, and the stimulating and EEG recording electrodes were attached. The latter consisted of a 28-channel bipolar montage configured in accordance with the international 10 : 20 system. The electrodes were Ag/AgCl sintered ring electrodes (Falk Minow, Herrsching, Germany) (1 cm outer diameter) that fastened securely to plastic loops imbedded in a prefabricated scalp cap (Easycap, Falk Minow, Herrsching, Germany). One of two cap sizes was chosen to give the correct positioning of the electrodes on the head relative to the nasion and inion, in accordance with the

international 10 : 20 system. The centres of the electrodes were filled with an electrolyte gel, and attention was given to ensure the gel made contact with the scalp. Two reference electrodes were positioned behind each ear. The EEG electrodes were connected to two 16-channel biosignal amplifiers (Guger Technologies, Herbersteinstrasse, Austria) and digitised (Gdaqsys, Guger Technologies, Herbersteinstrasse, Austria) to computer at 100 Hz for continuous display and later offline analysis. The amplifiers were powered using mains-charged battery packs. One of the spare channels on the amplifier was used as an event marker from the electrical stimulus generator. Application of the electrodes took approximately 1 hour. The quality of the EEG was assessed by visual inspection and corrective measures taken to improve the quality of “noisy” channels. This usually involved checking the contact of the electrolyte gel between the electrode and the scalp. Time restraints, particularly with the requirement for patients to be seated for up to two hours to complete the study, meant it was not practicable to check and monitor individual channel impedances.

During delivery of the stimulation sequences, the subjects were instructed to keep their eyes closed, refrain from talking, and relax as much as possible. The subjects were not specifically instructed to either attend to or ignore the stimuli. The subjects could stop the stimulation at any time by pressing a button. The three sequences took approximately 30 minutes to complete.

2.4. Data Analysis

2.4.1. Preprocessing. The EEG was preprocessed by a band-pass filter and further analyzed using EEGLAB [15] software in MatLab. The raw EEG was 1 to 50 Hz band-pass filtered. Each trace was visually inspected, and data predominated by electrical noise was discarded.

2.4.2. Wavelet Transforms. Wavelet transform was used to transform a single-channel EEG signal into a time-frequency map. In this study, the continuous wavelet transform (CWT) was applied, and the Morlet wavelet was employed [16], and it is

$$\psi_0(t) = \pi^{-1/4} e^{iwt} e^{-1/2t^2}, \quad (1)$$

where w is the wavelet central angle frequency, often $w \geq 6$, which is an optimal value to adjust the time-frequency resolution [17]. In this study, $w = 6$ was applied. Then, a family of wavelets can be generated: $\psi_s(t) = (1/\sqrt{2})\psi_0(t/s)$, $s \in (0, +\infty)$, and s is called a scale. The CWT at scale s and time t of a signal $x(t)$ is defined as

$$W(s, \tau) = \frac{1}{\sqrt{s}} \int x(t) \psi_s * \left(\frac{t - \tau}{s} \right) dt, \quad (2)$$

where ψ_s is a parent wavelet function, and $*$ denotes complex conjugation. By adjusting the scale s and the translation τ , a series of different frequency resolutions in the signal can be projected on the two-dimension space (scale s and translation τ). The factor \sqrt{s} normalizes energy across the different scales.

2.4.3. *PARAFAC*. After wavelet transformation of all EEG channels, a three-way tensor $X(t, f, c)$ (time-frequency-channel), giving the energy at time t , frequency f , and channel c , was obtained. To decompose the three-way tensor into time, frequency, and channel modes, the PARAFAC method was applied, and a linear combination of the three-way tensor was obtained by means of the alternating least squares (ALS) algorithm [14, 18]. The PARAFAC model is defined as

$$X_{T \times F \times C} = \sum_{n=1}^N a_n \circ b_n \circ c_n + E, \quad (3)$$

where N is the number of signal factors of $X_{T \times F \times C}$, and a_n , b_n , and c_n indicate the n th column of the loading matrices $A \in R^{T \times N}$, $B \in R^{F \times N}$, and $C \in R^{C \times N}$, respectively. A , B , and C represent the time, frequency, and channel modes and provide information on the interactions between modes. $E \in R^{T \times F \times C}$ is the residual information in the decomposition. The operator \circ represents the outer product of two vectors. Illustration of a 2-factor PARAFAC model on a three-way dataset is shown in Figure 1(A).

In the PARAFAC method, determination of the number of factors is a key issue. There are several methods to determine the number of factors, including the visual appearance of loadings, the residual analysis, the core consistency, and the number of iterations of the algorithm [19]. In this study, the core consistency method was employed. The core consistency represents the resemblance between the Tucker3 core and the PARAFAC core [14]. A Tucker3 model is similar to the PARAFAC model, and both of them are an extension of bilinear factor analysis to tensors [20]. The principle of the core consistency is as follows: (1) the PARAFAC model is available when the core consistency value is greater than 90%; (2) the PARAFAC model is not available when the core consistency value is less than 50%; (3) the PARAFAC model is probably available when the core consistency value is between 50 and 90%. The core consistency is defined as follows:

$$\text{Core-Consistency} = \left(1 - \frac{\sum_{t=1}^R \sum_{j=1}^R \sum_{k=1}^R (g_{ijk} - t_{ijk})^2}{R} \right) \times 100, \quad (4)$$

where g_{ijk} and t_{ijk} are the Tucker3 core and the PARAFAC core, respectively; R is the factor number. In the PARAFAC core, $t_{ijk} = 1$ if $i = j = k$, otherwise $t_{ijk} = 0$, and in the Tucker3 core, g_{ijk} can be nonzero for all i , j , and k . This method for determining the factor number has been successfully applied to multiple neural data [13, 21].

2.4.4. *Statistics*. Data were analysed in the time, frequency, and channel domains for all subjects. For the channel domain, the average energy for each stimulus was calculated for all channels and compared between the two groups, with outlier detection based on the generalized extreme studentized deviate (GESD) [22]. The average energy distribution for each group was obtained by averaging across all subjects.

Cortical locations were grouped into 5 zones (frontal (FPz, FP1, FP2, Fz, F3, F4, F7, F8), central (Fz, Cz, FC3, FC4, C3, C4, CP3, CP4), occipital (Pz, CP3, CP4, P3, P4, P7, P8, O1, O2), left temporal (F7, F3, Fz, FT7, FC3, T7, C3, TP7, CP3, P7), and right temporal (F8, F4, Fz, FT8, FC4, T8, C4, TP8, CP4, P8)), as shown in Figure 1(C). The active ratios (average energy of one zone/the sum of energy across all zones) were assessed using the t -test. In time and frequency domains, the differences between pain and control groups were evaluated by t -test.

The mean age of the chronic pain group (49 ± 11 years, $n = 13$) was significantly greater than the pain-free group (39 ± 11 years, $n = 13$) ($P < 0.05$, t -test). Multiple linear regression (MLR) was used to investigate the effect of age on the frequency domain parameter. The pain-free subjects and chronic pain subjects were represented by 0 and 1, respectively, and the effect of age and subject grouping differentiated by MLR as shown in Figure 4. In the MLR analysis, frequency was the dependent variable, while the age and subject grouping were independent variables. The effect of MLR was evaluated by the F -test.

3. Results

An example showing the epoch decomposition process using the PARAFAC method is illustrated in Figure 1(B). Wavelet decomposition was employed on the raw EEG recordings, generating time-frequency information corresponding to each channel. The frequency range for the wavelet decomposition was 1–50 Hz with an interval of 0.5 Hz. The factors were then extracted by the PARAFAC model from each epoch for every subject. Every factor was decomposed into three-way information at the time-frequency-channel domain. In the case shown (Figure 1(B) a, b, and c), the three-way information for one factor revealed a neural response frequency in the frontal region of 4 Hz and response time of 220 ms after the stimulus. Three-way information for each subject was obtained by averaging the results across all epochs. Comparisons between groups of the spatial topography, frequency, and time responses are shown in Figure 2.

3.1. *Comparison in Frequency Domain*. As shown in Figure 2 (middle), the frequency response in the chronic pain group (around 4 Hz) was lower than that in control group (around 5.5 Hz) and was statistically significant for the pain stimulus (chronic pain group: 3.456 ± 1.716 Hz; control group: 5.608 ± 2.315 Hz; $P < 0.05$, t -test).

3.2. *Comparison in Time Domain*. As shown in Figure 2 (right), the response time was around 250 ms in both groups and showed no significant difference for any of the stimuli.

3.3. *Comparison in Location*. As shown in Figure 2 (left), the topographical analysis indicated that the active zone was mainly in the frontal region in pain group and in the central region in the control group. These differences were quantified by comparing the active ratios in the 5 cortical regions (see Figure 1(C)). The results are shown in Figure 3(a);

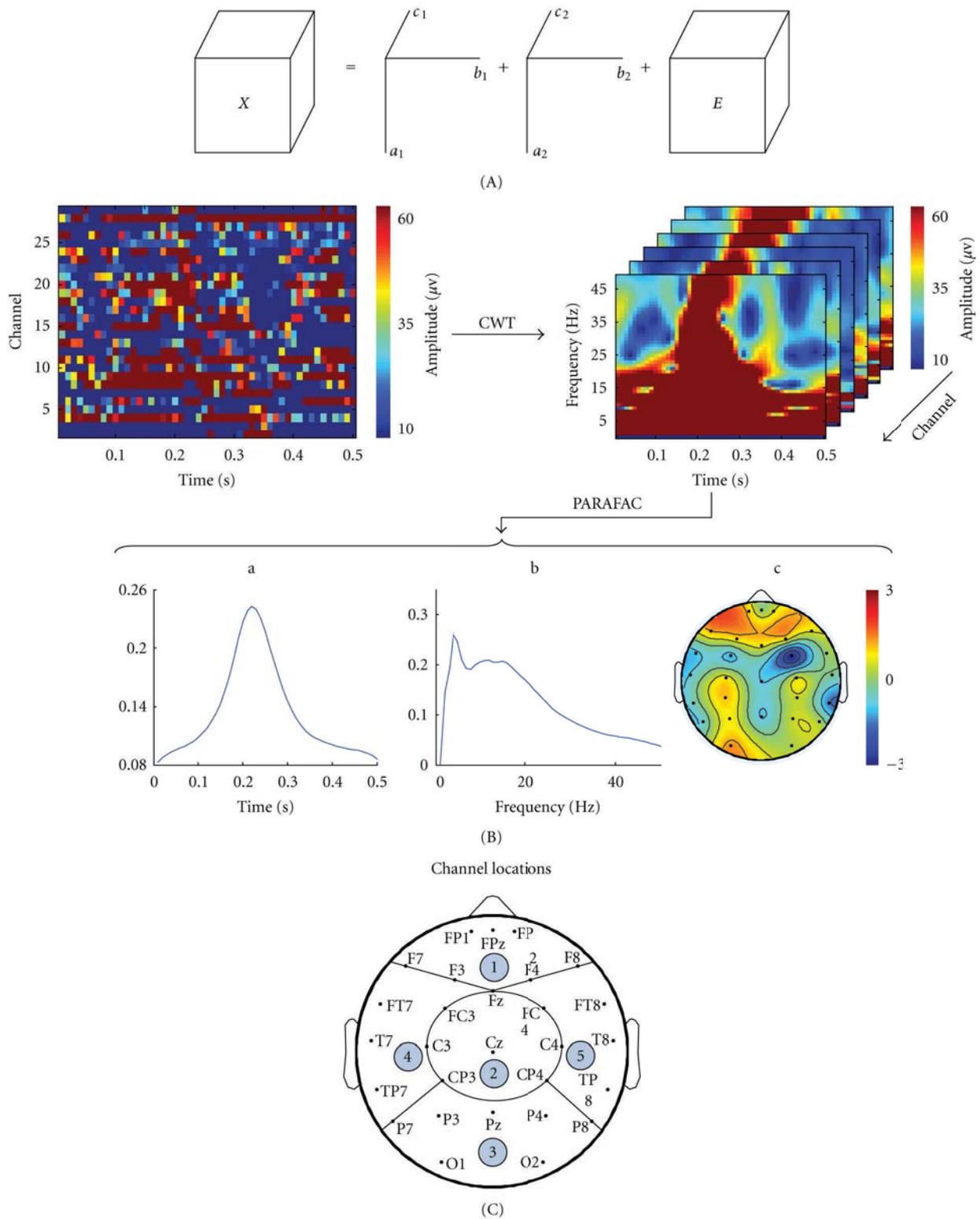


FIGURE 1: The PARAFAC model and factors extracted by the PARAFAC model from one case. (A) PARAFAC modeling of a three-way tensor. Each component ($R = 2$) is the outer product of a , b , and c of rank-1, and E is a residual tensor. (B) Flowchart describing the procedure of the epoch decomposition process by the PARAFAC method. (C) Whole brain cortex is clustered into 5 zones (1 = frontal, 2 = central, 3 = occipital, 4 = left temporal, and 5 = right temporal).

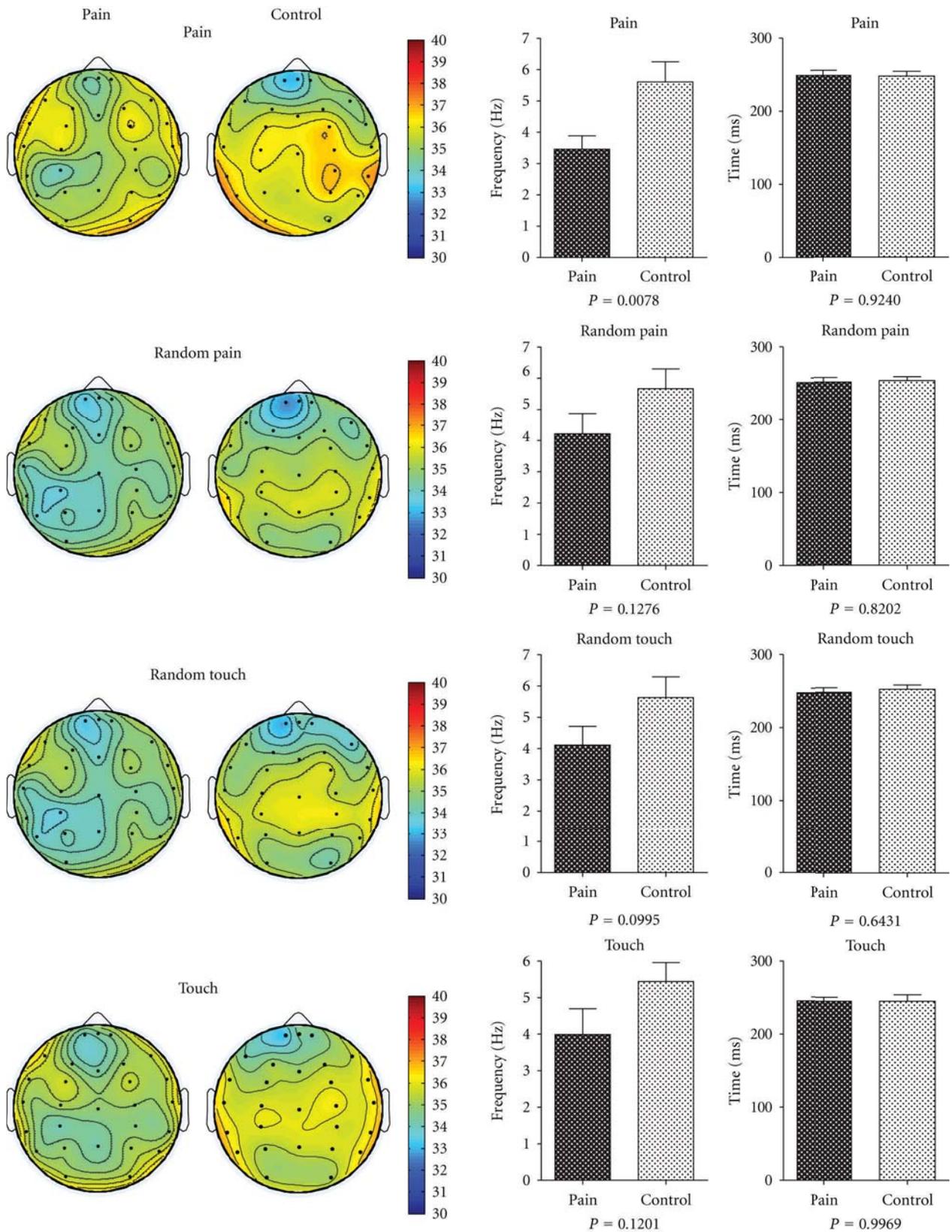


FIGURE 2: Results at the level of group corresponding to the four different stimuli. Left: the average energy distribution (pain, random pain, random touch, and touch, resp.). Middle: the statistical results in the frequency domain. Right: the statistical results in the time domain.

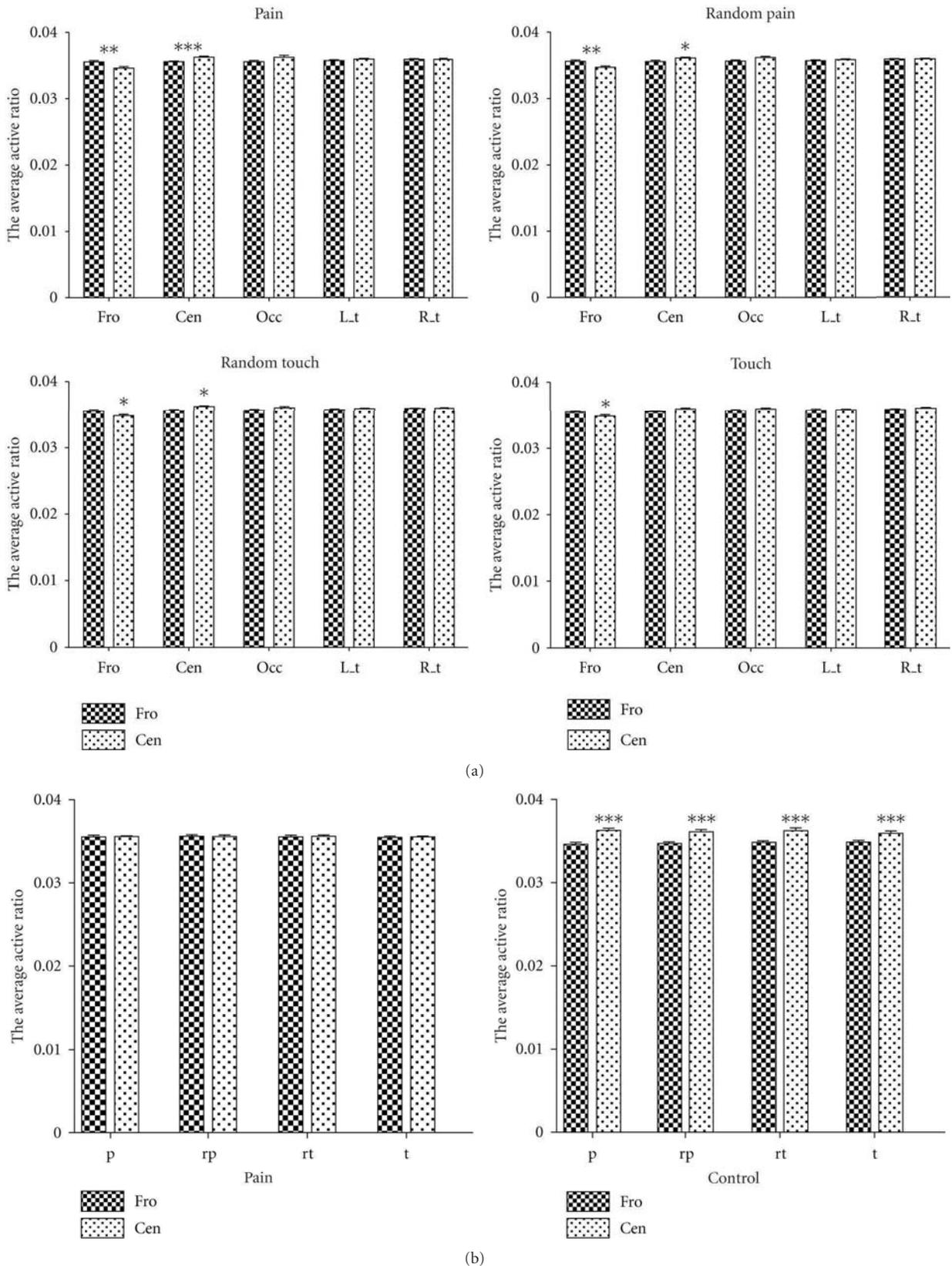


FIGURE 3: (a) Statistical results of the active ratio between the two groups corresponding to 5 zones under the four different stimuli. (b) Within-group comparison of the active ratio between the frontal and the central zones in the chronic pain group (left) and the corresponding result in the control group (right).

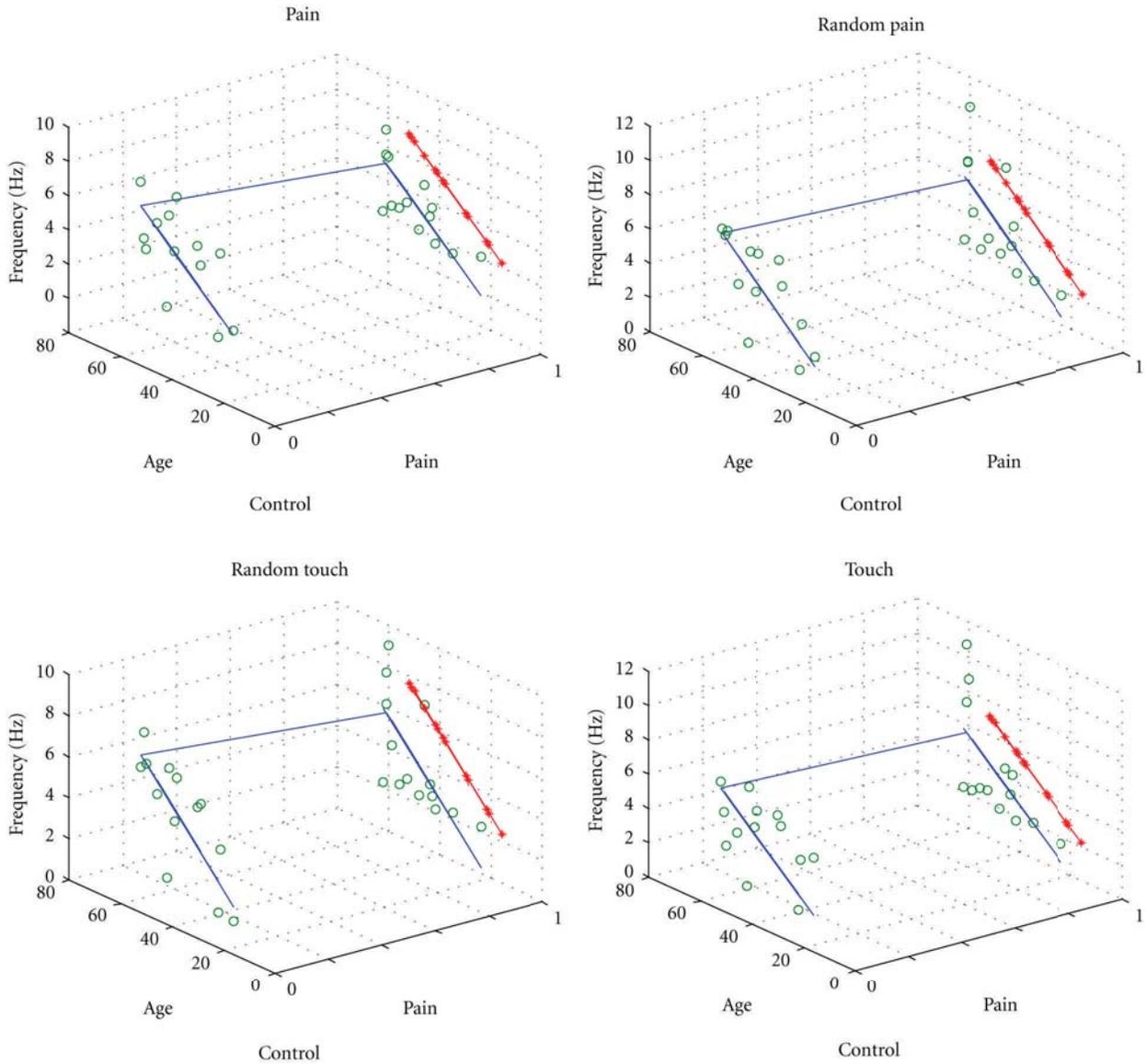


FIGURE 4: Influence of age and pain status on the frequency responses of subjects by MLR during the four different stimuli.

in the frontal region, the active ratio of neural response in the chronic pain group was significantly higher than that in the control group for all 4 tests stimuli (chronic pain group: 0.03551 ± 0.00085 , control group: 0.03459 ± 0.00080 , pain: $P < 0.01$; chronic pain group: 0.03557 ± 0.00088 , control group: 0.03469 ± 0.00076 , random pain: $P < 0.01$; chronic pain group: 0.03551 ± 0.00077 , control group: 0.03482 ± 0.00074 , random touch: $P < 0.05$; chronic pain group: 0.03545 ± 0.00061 , control group: 0.03485 ± 0.00073 , touch: $P < 0.05$, *t*-test). In comparison, in the central region, the active ratio of neural response in the chronic pain group was lower than that in the control group and was statistically significant for all but the touch stimulus (chronic pain group: 0.03555 ± 0.00043 , control group: 0.03624 ± 0.00051 , pain: $P < 0.001$; chronic pain group: 0.03557 ± 0.00077 , con-

trol group: 0.03608 ± 0.00050 , random pain: $P < 0.05$; chronic pain group: 0.0356 ± 0.0006 , control group: 0.03621 ± 0.00043 , random touch: $P < 0.01$; chronic pain group: 0.0355 ± 0.00042 , control group: 0.03591 ± 0.00065 , touch: $P = 0.0526$, *t*-test). In other regions, there were no significant differences between the two groups.

Furthermore, we were interested in the differences in active ratio between the frontal and central regions within each group. As shown in Figure 3(b), in the pain group, the active ratios were not significantly different, whereas in the control group, the active ratios in the central region were significantly higher than in the frontal region for all stimuli (frontal: 0.03459 ± 0.00080 , central: 0.03624 ± 0.00051 , pain: $P < 0.0001$; frontal: 0.03469 ± 0.00076 , central: 0.03608 ± 0.00050 , random pain: $P < 0.0001$; frontal:

0.03482 ± 0.00074 , central: 0.03621 ± 0.00043 , random touch: $P < 0.0001$; frontal: 0.03485 ± 0.00073 , central: 0.03591 ± 0.00065 , touch: $P < 0.001$, t -test).

3.4. Effects of the Age on the Time Frequency and Active Ratio. The mean age of the subjects in the pain group was significantly greater than in the control group. This study considered the effects of the age on the time, frequency, and active ratio parameters. Firstly, linear regression analysis was carried out on age versus response time, frequency, and active ratio; the only significant correlation was observed between age and response frequency. MLR was used to determine whether the difference in the frequency domain between two groups could be attributed to an age effect and is shown in Figure 4. The line marked by “*” was the vertical mapping of the line of control group to the same plane with the line of the chronic pain group. From the comparisons of these lines, it is clear that the frequencies of the chronic pain group were significantly lower than those of the control group at the same age for the pain ($R^2 = 0.4124$, $F = 24.131$, $P = 0.0017$), random pain ($R^2 = 0.3398$, $F = 6.177$, $P = 0.0069$), random touch ($R^2 = 0.4064$, $F = 8.2145$, $P = 0.0019$), and touch ($R^2 = 0.2156$, $F = 3.2985$, $P = 0.0542$) stimuli. The differences in the frequency domain between the two groups therefore cannot be attributed entirely to the difference in age between groups.

4. Discussion

4.1. PARAFAC as an Effective Method for EEG Decomposition. In this study, the PARAFAC method was applied to somatosensory evoked potential recordings to analyze EEG time-frequency-channel domain [23] characteristics in chronic pain subjects. The advantage of this method is that it can extract more information in comparison with the two-way models (PCA and ICA) but also takes into account the frequency content of the signals in specific time periods across different channels [20]. The PARAFAC method has been successfully used to characterise the structure of epileptic seizure [14, 24, 25]. Analyzing the EEG signals in the time-frequency-channel domain, we found that the response latency was about 250 ms for all stimuli in both groups, while the frequency responses of the chronic pain group were lower than those of the control group. Furthermore, topographical analysis showed that the chronic pain group exhibited predominantly frontal cortical activity, compared to central activation in the control group. These findings are in accordance with previous results showing a response latency of about 250 ms [26], an increased size and reduction in population spike frequency [27], and restriction of some ERP components to frontal-central regions in patients with fibromyalgia syndrome [28]. Our findings indicate that the PARAFAC method is an effective technique for extracting the characteristics of multiple EEG recordings in the time-frequency-channel domain.

4.2. Activity Regions Involved in Chronic Pain. Previous studies have shown that the scale distribution of laser-evoked potentials (LEPs) around the chronic pain ERP

components extends into vertex and frontocentral leads in fibromyalgia syndrome (FS) patients, indicating more widespread nociceptive activation outside the cortical hand area [28]. One question addressed in this study is whether the active ratio of frontal or central regions versus the whole cortex is different between the two groups. In this study, the active ratio was used to assess the activity within different regions of the cortex. The statistical results showed that the active ratios of the chronic pain group were significantly higher than those of the control group in the frontal region. The active ratios calculated from the other cortical regions showed no differences between the two groups. These results imply that evoked EEG activity in the frontal and central cortical regions may help discriminate between chronic pain and pain-free subjects. These findings are in agreement with the previous studies showing that some ERP components are more restricted to frontocentral regions in patients with fibromyalgia syndrome [28].

In keeping with the above findings, the active ratio in the central region was significantly higher than that in the frontal region in the control group, and vice versa for the chronic pain group. These results further indicate that the frontal cortical regions are involved in somatosensory processing in the chronic pain condition compared to central regions for pain-free subjects, in accordance with [29–32]. These findings suggest that passive functions (emotion, attention, etc.) are presented more frontally [33], and pain beliefs influence patients behavioral and psychological functioning because of their persistent pain experience [34].

4.3. Relationship between Frequency and Chronic Pain. In this study, we found that the cortical neural responses to somatosensory stimuli occurred around 250 ms in both of the groups and that the response frequency in the chronic pain group was lower than in the control group. In particular, the dominant activity was in the delta frequency range (around 4 Hz) for the chronic pain patients, compared to the theta frequency range (around 5.5 Hz) in the controls. Taken together, these results indicate that frontal cortical activation and a lower response frequency are characteristics of evoked EEG responses in chronic pain subjects.

In imaging studies, the functional connectivity between cortical structures receiving input arising from nociceptors has documented that experimental pain is processed in multiple pain-related areas, often characterized as a “pain network” [35, 36]. This “pain network” is not fixed but changes as a function of the pain-related task [37]. Moreover, in the study of cortical pathophysiology, pain-related neural networks are larger in patients with sympathetically mediated chronic pain (SMP) compared to acute pain states [32]. Therefore, the pain-related network of chronic pain patients is more extensive than that of the controls. This may be partly due to the effects of the patients’ persistent pain experience, resulting in processing experimental pain accompanied with subjective experience. The larger pain-related network is accompanied with a lower frequency rhythm, which is consistent with data from isolated hippocampal slice experiments showing a reduction in the frequency of

population spikes with increasing size of the participating neuronal population [27].

Theta activity is associated with alertness, attention, and the efficient processing of cognitive and perceptual tasks [38], while delta band activity is associated with pathological conditions associated with impairment of brain networks such as amnesic mild cognitive impairment (MCI) and Alzheimer's disease (AD) [39]. Chronic pain is also found to be frequently associated with psychiatric disorders [40]. Thus, the delta rhythm of chronic pain patients is suggestive of impairment to pain-related brain network processing.

4.4. Influence of Age on the Response Frequency. In this study, the mean age of the chronic pain group was significantly greater than that of the control group. MRL was applied to analyze the influences of age and pain status on the frequency responses of subjects. Our results showed that the frequency responses of the chronic pain group were significantly lower than those of the control group across all ages. Thus, the differences in the frequency domain between the two groups cannot be attributed entirely to the difference in age between groups.

In summary, the PARAFAC method with a continuous wavelet was used to extract time-frequency-channel domain information from somatosensory-evoked EEG recordings from chronic pain and pain-free subjects. We found that the response latency was about 250 ms, the chronic pain group had lower response frequency, and the central and frontal regions were the crucial regions of cortical activation. Further analysis indicated that the frontal regions were more involved in the chronic pain condition than the control condition. Application of MLR to the analysis of the relationship between frequency and age showed that the lower frequency response in the chronic pain group was not attributable to the difference in subject age. The conclusion from these findings is that the PARAFAC method is an effective tool for characterising multiple EEG recordings in the time-frequency-channel domain.

Acknowledgments

This research was partly supported by the National Natural Science Foundation of China (nos. 61025019, 90820016), the Program for New Century Excellent Talents in University (NECT-07-0735), and the Natural Science Foundation of Hebei, China.

References

- [1] E. Niedermeyer and F. L. da Silva, *Electroencephalography: basic Principles, Clinical Applications, and Related Fields*, Lippincott Williams & Wilkins, Philadelphia, Pa, USA, 5th edition, 2004.
- [2] A. B. Ettinger and C. E. Argoff, "Use of antiepileptic drugs for nonepileptic conditions: psychiatric disorders and chronic pain," *Neurotherapeutics*, vol. 4, no. 1, pp. 75–83, 2007.
- [3] M. Zhuo, "Cortical excitation and chronic pain," *Trends in Neurosciences*, vol. 31, no. 4, pp. 199–207, 2008.
- [4] V. de Pascalis, M. R. Magurano, A. Bellusci, and A. C. Chen, "Somatosensory event-related potential and autonomic activity to varying pain reduction cognitive strategies in hypnosis," *Clinical Neurophysiology*, vol. 112, no. 8, pp. 1475–1485, 2001.
- [5] J. Sarnthein and D. Jeanmonod, "High thalamocortical theta coherence in patients with neurogenic pain," *NeuroImage*, vol. 39, no. 4, pp. 1910–1917, 2008.
- [6] S. W. G. Derbyshire, A. K. P. Jones, F. Creed et al., "Cerebral responses to noxious thermal stimulation in chronic low back pain patients and normal controls," *NeuroImage*, vol. 16, no. 1, pp. 158–168, 2002.
- [7] P. Y. Geha and A. V. Apkarian, "Brain imaging findings in neuropathic pain," *Current Pain and Headache Reports*, vol. 9, no. 3, pp. 184–188, 2005.
- [8] M. de Tommaso, G. Libro, M. Guido, V. Scirucchio, L. Losito, and F. Puca, "Heat pain thresholds and cerebral event-related potentials following painful CO₂ laser stimulation in chronic tension-type headache," *Pain*, vol. 104, no. 1-2, pp. 111–119, 2003.
- [9] J. Stern, D. Jeanmonod, and J. Sarnthein, "Persistent EEG over-activation in the cortical pain matrix of neurogenic pain patients," *NeuroImage*, vol. 31, no. 2, pp. 721–731, 2006.
- [10] M. Valeriani, L. Rambaud, and F. Mauguière, "Scalp topography and dipolar source modelling of potentials evoked by CO₂ laser stimulation of the hand," *Electroencephalography and Clinical Neurophysiology*, vol. 100, no. 4, pp. 343–353, 1996.
- [11] B. Bromm and J. Lorenz, "Neurophysiological evaluation of pain," *Electroencephalography and Clinical Neurophysiology*, vol. 107, no. 4, pp. 227–253, 1998.
- [12] J. Medina and H. B. Coslett, "From maps to form to space: touch and the body schema," *Neuropsychologia*, vol. 48, no. 3, pp. 645–654, 2010.
- [13] F. Miwakeichi, E. Martínez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.
- [14] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics*, vol. 23, no. 13, pp. i10–i18, 2007.
- [15] 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.
- [16] X. Li, X. Yao, J. Fox, and J. G. Jefferys, "Interaction dynamics of neuronal oscillations analysed using wavelet transforms," *Journal of Neuroscience Methods*, vol. 160, no. 1, pp. 178–185, 2007.
- [17] M. Farge, E. Goirand, Y. Meyer, F. Pascal, and M. V. Wickerauser, "Improved predictability of two-dimensional turbulent flows using wavelet packet compression," *Fluid Dynamics Research*, vol. 10, no. 4–6, pp. 229–250, 1992.
- [18] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis with Applications in the Chemical Sciences*, John Wiley & Sons, England, UK, 2004.
- [19] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *Journal of Chemometrics*, vol. 17, no. 5, pp. 274–286, 2003.
- [20] E. Acar and B. Yener, "Unsupervised multiway data analysis: a literature survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 1, pp. 6–20, 2009.

- [21] F. Estienne, N. Matthijs, D. L. Massart, P. Ricoux, and D. Leibovici, "Multi-way modelling of high-dimensionality electroencephalographic data," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 1, pp. 59–71, 2001.
- [22] X. Li, C. P. Bowers, and T. Schnier, "Classification of energy consumption in buildings with outlier detection," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 11, pp. 3639–3644, 2010.
- [23] R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, no. 1, pp. 1–84, 1970.
- [24] M. de Vos, A. Vergult, L. de Lathauwer et al., "Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone," *NeuroImage*, vol. 37, no. 3, pp. 844–854, 2007.
- [25] E. Acar, C. A. Bingo, H. Bingo, and B. Yener, "Computational analysis of epileptic focus localization," in *Proceedings of the 4th IASTED International Conference on Biomedical Engineering*, pp. 317–322, February 2006.
- [26] C. Pazzaglia and M. Valeriani, "Brain-evoked potentials as a tool for diagnosing neuropathic pain," *Expert Review of Neurotherapeutics*, vol. 9, no. 5, pp. 759–771, 2009.
- [27] J. E. Fox, M. Bikson, and J. G. Jefferys, "The effect of neuronal population size on the development of epileptiform discharges in the low calcium model of epilepsy," *Neuroscience Letters*, vol. 411, no. 2, pp. 158–161, 2007.
- [28] J. Lorenz, K. Grasedyck, and B. Bromm, "Middle and long latency somatosensory evoked potentials after painful laser stimulation in patients with fibromyalgia syndrome," *Electroencephalography and Clinical Neurophysiology*, vol. 100, no. 2, pp. 165–168, 1996.
- [29] K. L. Casey, J. Lorenz, and S. Minoshima, "Insights into the pathophysiology of neuropathic pain through functional brain imaging," *Experimental Neurology*, vol. 184, no. 1, pp. S80–S88, 2003.
- [30] V. Neugebauer, V. Galhardo, S. Maione, and S. C. Mackey, "Forebrain pain mechanisms," *Brain Research Reviews*, vol. 60, no. 1, pp. 226–242, 2009.
- [31] X. Moisset and D. Bouhassira, "Brain imaging of neuropathic pain," *NeuroImage*, vol. 37, supplement 1, pp. S80–S88, 2007.
- [32] A. V. Apkarian, P. S. Thomas, B. R. Krauss, and N. M. Szeverenyi, "Prefrontal cortical hyperactivity in patients with sympathetically mediated chronic pain," *Neuroscience Letters*, vol. 311, no. 3, pp. 193–197, 2001.
- [33] N. Picard and P. L. Strick, "Motor areas of the medial wall: a review of their location and functional activation," *Cerebral Cortex*, vol. 6, no. 3, pp. 342–353, 1996.
- [34] M. P. Jensen, J. M. Romano, J. A. Turner, A. B. Good, and L. H. Wald, "Patient beliefs predict patient functioning: further support for a cognitive-behavioural model of chronic pain," *Pain*, vol. 81, no. 1-2, pp. 95–104, 1999.
- [35] P. Rainville, B. Carrier, R. K. Hofbauer, M. C. Bushnell, and G. H. Duncan, "Dissociation of sensory and affective dimensions of pain using hypnotic modulation," *Pain*, vol. 82, no. 2, pp. 159–171, 1999.
- [36] J. L. Davis, "Use of sibutramine hydrochloride monohydrate in the treatment of the painful peripheral neuropathy of diabetes," *Diabetes Care*, vol. 23, no. 10, pp. 1594–1595, 2000.
- [37] S. Ohara, N. E. Crone, N. Weiss, and F. A. Lenz, "Analysis of synchrony demonstrates "pain networks" defined by rapidly switching, task-specific, functional connectivity between pain-related cortical structures," *Pain*, vol. 123, no. 3, pp. 244–253, 2006.
- [38] R. M. Stern, W. J. Ray, and K. S. Quigley, *Psychophysiological Recording*, Oxford University Press, New York, NY, USA, 2nd edition, 2001.
- [39] C. Babiloni, G. Frisoni, F. Vecchio et al., "Global functional coupling of resting EEG rhythms is abnormal in mild cognitive impairment and Alzheimer's disease: a multicenter EEG study," *Journal of Psychophysiology*, vol. 23, no. 4, pp. 224–234, 2010.
- [40] J. Dersh, P. B. Polatin, and R. J. Gatchel, "Chronic pain and psychopathology: research findings and theoretical considerations," *Psychosomatic Medicine*, vol. 64, no. 5, pp. 773–786, 2002.

Research Article

A Batch Rival Penalized Expectation-Maximization Algorithm for Gaussian Mixture Clustering with Automatic Model Selection

Jiechang Wen,¹ Dan Zhang,² Yiu-ming Cheung,² Hailin Liu,¹ and Xinge You³

¹ Faculty of Applied Mathematics, Guangdong University of Technology, Guangzhou 510520, China

² Department of Computer Science, Hong Kong Baptist University, Kowloon, Hong Kong

³ Department of Electronics and Information Engineering, Huazhong University of Science & Technology, Wuhan, China

Correspondence should be addressed to Yiu-ming Cheung, ymc@comp.hkbu.edu.hk

Received 31 August 2011; Accepted 9 October 2011

Academic Editor: Sheng-yong Chen

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

Within the learning framework of maximum weighted likelihood (MWL) proposed by Cheung, 2004 and 2005, this paper will develop a batch Rival Penalized Expectation-Maximization (RPEM) algorithm for density mixture clustering provided that all observations are available before the learning process. Compared to the adaptive RPEM algorithm in Cheung, 2004 and 2005, this batch RPEM need not assign the learning rate analogous to the Expectation-Maximization (EM) algorithm (Dempster et al., 1977), but still preserves the capability of automatic model selection. Further, the convergence speed of this batch RPEM is faster than the EM and the adaptive RPEM in general. The experiments show the superior performance of the proposed algorithm on the synthetic data and color image segmentation.

1. Introduction

As a typical statistical technique, clustering analysis has been widely applied to a variety of scientific areas such as data mining [1], vector quantization [2, 3], image processing [4–7], and so forth. In particular, clustering analysis provides a useful tool to solve the several computer vision problems, for example, multithresholding of gray level images, analysis of the Hough space, and range image segmentation, formulated in the feature space paradigm [8]. In general, one kind of clustering analysis can be formulated as a density mixture modeling, in which each mixture component represents the density distribution of a data cluster. Subsequently, the task of clustering analysis is to identify the dense regions of the input (also called *observation* interchangeably) densities in a mixture. Such a clustering is therefore called a density mixture clustering.

In general, the Expectation-Maximum (EM) algorithm [9, 10] has provided a general solution for the parameter estimation of a density mixture model. Nevertheless, it needs to preassign an appropriate number of density components, that is, the number of clusters. Roughly, the mixture may overfit the data if too many components are utilized, whereas

a mixture with too few components may not be flexible enough to approximate the true underlying model. Subsequently, the EM almost always leads to a poor estimate result if the number of components is misspecified. Unfortunately, from the practical viewpoint, it is hard or even impossible to know the exact cluster number in advance. In the literature, one promising way is to develop a clustering algorithm that is able to perform a correct clustering without preassigning the exact number of clusters. Such algorithms include the RPCL algorithm [11] and its improved version, namely, RPCCL [12]. More recently, Cheung [13, 14] has proposed a general learning framework, namely, Maximum Weighted Likelihood (MWL), through which an adaptive Rival Penalized EM (RPEM) algorithm has been proposed for density mixture clustering. The RPEM learns the density parameters by making mixture component compete each other at each time step. Not only are the associated parameters of the winning density component updated to adapt to an input, but also all rivals' parameters are penalized with the strength proportional to the corresponding posterior density probabilities. Therefore, this intrinsic rival penalization mechanism enables the RPEM to automatically select an appropriate number of densities by gradually fading out the

redundant densities from a density mixture. Furthermore, a simplified version of RPEM has included RPCL and RPCCL as its special cases with some new extensions.

In the papers [13, 14], the RPEM algorithm learns the parameters via a stochastic gradient ascending method; that is, we update the parameters immediately and adaptively once the current observation is available. In general, the adaptiveness of the RPEM makes it more applicable to the environment changed over time. Nevertheless, the convergence speed of the RPEM relies on the value of learning rate. Often, by a rule of thumb, we arbitrarily set the learning rate at a small positive constant. If the value of learning rate is assigned too small, the algorithm will converge at a very slow speed. On the contrary, if it is too large, the algorithm may even oscillate. In general, it is a nontrivial task to assign an appropriate value to the learning rate, although we can pay extra efforts to make the learning rate dynamically change over time, for example, see [15].

In this paper, we further study the MWL learning framework and develop a batch RPEM algorithm accordingly provided that all observations are available before the learning process. Compared to the adaptive RPEM, this batch one need not assign the learning rate analogous to the EM, but still preserves the capability of automatic model selection. Further, the convergence speed of this batch RPEM is faster than the EM and the adaptive RPEM in general. The experiments have shown the superior performance of the proposed algorithm on the synthetic data and color image segmentation.

The remainder of this paper is organized as follows. Section 2 reviews the MWL learning framework. In Section 3, we present the batch RPEM algorithm in detail, in which the weights involve a coefficient ε . We will therefore further explore the assignment of ε in Section 4. Section 5 shows the detailed experiment results. Finally, we draw a conclusion in Section 6.

2. Overview of Maximum Weighted Likelihood (MWL) Learning Framework

Suppose that an input $\mathbf{x} \in \mathcal{R}^d$ comes from the following density mixture model:

$$P(\mathbf{x} | \Theta) = \sum_{j=1}^k \alpha_j p(\mathbf{x} | \theta_j), \quad \sum_{j=1}^k \alpha_j = 1, \quad (1)$$

$$\alpha_j > 0, \quad \forall 1 \leq j \leq k,$$

where Θ is the parameter set of $\{\alpha_j, \theta_j\}_{j=1}^k$. Furthermore, k is the number of components, α_j is the mixture proportion of the j th component, and $p(\mathbf{x} | \theta_j)$ is a multivariate probability density function (pdf) of \mathbf{x} parameterized by θ_j . As long as we know the value of Θ , an input \mathbf{x} can be classified into a certain cluster via its posterior probability:

$$h(j | \mathbf{x}, \Theta) = \frac{\alpha_j p(\mathbf{x} | \theta_j)}{P(\mathbf{x} | \Theta)} \quad (2)$$

using the winner-take-all rule, that is, assigning an input \mathbf{x} to Cluster c if $c = \arg \max_j h(j | \mathbf{x}, \Theta)$ or taking its soft

version which assigns \mathbf{x} to Cluster j with the probability $h(j | \mathbf{x}, \Theta)$. Therefore, how to estimate the parameter set Θ , particularly without knowing the correct value of k in advance, is a key issue in density mixture clustering.

In the MWL learning framework [13, 14], the parameter set Θ is learned via maximizing the following Weighted Likelihood (WL) cost function:

$$l(\Theta) = \omega(\Theta; \mathbf{x}) + \nu(\Theta; \mathbf{x}) \quad (3)$$

with

$$\omega(\Theta; \mathbf{x}) = \int \sum_{j=1}^k g(j | \mathbf{x}, \Theta) \ln[\alpha_j p(\mathbf{x} | \theta_j)] dF(\mathbf{x}), \quad (4)$$

$$\nu(\Theta; \mathbf{x}) = - \int \sum_{j=1}^k g(j | \mathbf{x}, \Theta) \ln h(j | \mathbf{x}, \Theta) dF(\mathbf{x}),$$

where $g(j | \mathbf{x}, \Theta)$'s are the designable weights satisfying the two conditions:

- (1) $\sum_{j=1}^k g(j | \mathbf{x}, \Theta) = 1$,
- (2) for all j , $g(j | \mathbf{x}, \Theta) = 0$ if $h(j | \mathbf{x}, \Theta) = 0$.

Suppose that a set of N i.i.d. observations, denoted as $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, comes from the density mixture model in (1). The empirical WL function of (3), written as $Q(\Theta; \mathbf{X})$, can be given as

$$Q(\Theta; \mathbf{X}) = \omega(\Theta; \mathbf{X}) + \nu(\Theta; \mathbf{X}) \quad (5)$$

with

$$\omega(\Theta; \mathbf{X}) = \frac{1}{N} \sum_{t=1}^N \sum_{j=1}^k g(j | \mathbf{x}_t, \Theta) \ln[\alpha_j p(\mathbf{x}_t | \theta_j)], \quad (6)$$

$$\nu(\Theta; \mathbf{X}) = - \frac{1}{N} \sum_{t=1}^N \sum_{j=1}^k g(j | \mathbf{x}_t, \Theta) \ln h(j | \mathbf{x}_t, \Theta).$$

Moreover, the weights $g(j | \mathbf{x}_t, \Theta)$'s have been generally designed as [13, 14]

$$g(j | \mathbf{x}_t, \Theta) = (1 + \varepsilon_t) I(j | \mathbf{x}_t, \Theta) - \varepsilon_t h(j | \mathbf{x}_t, \Theta), \quad (7)$$

where ε_t is a coefficient varying with the time step t in general. Please note that $g(j | \mathbf{x}_t, \Theta)$'s in (7) can be negative as well as positive. For simplicity, we hereinafter set ε_t as a constant, denoted as ε . Moreover, $I(j | \mathbf{x}_t, \Theta)$ is an indicator function with

$$I(j | \mathbf{x}_t, \Theta) = \begin{cases} 1, & \text{if } j = c = \arg \max_{1 \leq j \leq k} h(j | \mathbf{x}_t, \Theta), \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Subsequently, under a specific weight design, the papers [13, 14] have presented the adaptive RPEM to learn Θ via maximizing the WL function of (5) using a stochastic gradient ascent method, which is able to fade out the redundant densities gradually from a density mixture. Consequently, it can automatically select an appropriate number of density components in density mixture clustering.

Interested readers may refer to the paper [14] for more details. We summarize the main steps of the adaptive RPEM in Algorithm 1. Although the experiments have shown the superior performance of the adaptive RPEM on automatic model selection, its convergence speed, however, relies on the value of learning rate. Under the circumstances, we will present a batch version without the learning rate in the next section.

3. Batch RPEM Algorithm

To estimate the parameter set within the MWL framework, we have to maximize the empirical WL function $Q(\Theta; \mathbf{X})$ in (5). In general, we update the parameters via maximizing the first term of (5), that is, $\omega(\Theta; \mathbf{X})$, by fixing the second term $\nu(\Theta; \mathbf{X})$. Subsequently, we need to solve the following nonlinear optimization problem:

$$\tilde{\Theta} = \arg \max_{\Theta} \{\omega(\Theta; \mathbf{X})\} \quad (9)$$

subject to the constraints as shown in (1). We adopt the Lagrange method analogous to the EM by introducing a Lagrange multiplier λ into the Lagrange function. Subsequently, we have

$$\mathcal{L}(\Theta, \lambda) = \omega(\Theta; \mathbf{X}) + \lambda \left(\sum_{j=1}^k \alpha_j - 1 \right). \quad (10)$$

In this paper, we concentrate on the Gaussian mixture model only, that is, each component $p(\mathbf{x} | \theta_j)$ in (1) is a Gaussian density. We then have

$$p(j | \mathbf{x}_t, \theta_j) = (2\pi)^{-d/2} |\mathbf{C}_j|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{x}_t - \boldsymbol{\mu}_j)^T \mathbf{C}_j^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_j) \right], \quad (11)$$

where $\theta_j = (\boldsymbol{\mu}_j, \mathbf{C}_j)$, $\boldsymbol{\mu}_j$ and \mathbf{C}_j are the mean (also called *seed points* interchangeably) and the covariance of the j th density, respectively.

Through optimizing (10), we then finally obtain the batch RPEM algorithm as shown in Algorithm 2. If a covariance matrix $\mathbf{C}_j^{(n+1)}$ is singular, then it indicates that the corresponding j th density component is degenerated and can be simply discarded without being learned any more in the subsequent iterations. In this case, we have to normalize those remaining $\alpha_r^{(n+1)}$'s ($r \neq j$) so that their sum is always kept to be 1.

In the above batch RPEM, its capability of automatic model selection is controlled by the weight functions $g(j | \mathbf{x}_t, \theta)$'s, which further rely on the parameter ε as shown in (7). Subsequently, a new question is arisen: how to assign an appropriate value of ε ? The next section will answer this question.

4. How to Assign Parameter ε ?

To deal with how to assign an appropriate value of ε , we rewrite (7) as the following form:

$$g(j | \mathbf{x}_t, \Theta) = \begin{cases} (1 + \varepsilon)I(c | \mathbf{x}_t, \Theta) - \varepsilon h(c | \mathbf{x}_t, \Theta), & \text{if } j = c, \\ -\varepsilon h(j | \mathbf{x}_t, \Theta), & \text{otherwise} \end{cases} \\ = \begin{cases} h(c | \mathbf{x}_t, \Theta) + (1 + \varepsilon)(1 - h(c | \mathbf{x}_t, \Theta)), & \text{if } j = c, \\ h(j | \mathbf{x}_t, \Theta) - (1 + \varepsilon)h(j | \mathbf{x}_t, \Theta), & \text{otherwise.} \end{cases} \quad (12)$$

Intuitively, the term $(1 + \varepsilon)(1 - h(c | \mathbf{x}_t, \Theta))$ can be regarded as the award received by the winning density component (i.e., the c th density with $I(c | \mathbf{x}_t, \Theta) = 1$), and meanwhile the term $-(1 + \varepsilon)h(j | \mathbf{x}_t, \Theta)$ is the penalty of the rival components (i.e., those densities with $I(j | \mathbf{x}_t, \Theta) = 0$). In general, it is expected that the award is positive and the penalty is negative. That is, ε should be greater than -1 . Otherwise, as $\varepsilon < -1$, we will meet an awkward situation; that is, the amount of award is negative and the penalty one becomes positive. This implies that we will penalize the winner and award the rivals, which evidently violates our expectations. Furthermore, as $\varepsilon = -1$, both of the award and penalty amount become zero. In this special case, the batch RPEM is actually degenerated into the EM without the property of automatic model selection. As a result, ε is required to be greater than -1 . In addition, ε in the batch RPEM should be a negative value. Otherwise, the weights of the rival components $g(j | \mathbf{x}_t, \Theta) = -\varepsilon h(j | \mathbf{x}_t, \Theta)$'s become negative, resulting in some α_j 's to be negative finally. Hence, an appropriate selection of ε in the batch RPEM would be a negative value and greater than -1 . That is, ε should be fallen into the range of $(-1, 0)$.

Furthermore, our empirical studies have found that a smaller ε will lead the batch RPEM algorithm to a more robust performance, especially when the value of k is large and the data are overlapped considerably. In other words, the algorithm has a poor capability of automatic model selection if ε is close to zero. To illustrate this scenario, we have utilized two synthetic data sets that are generated from the two bivariate three-Gaussian mixtures individually as shown in Figures 1(a) and 1(b), where each data set consists of 1,000 observations with the true mixture proportions: $\alpha_1^* = 0.4$, $\alpha_2^* = 0.3$, and $\alpha_3^* = 0.3$. Also, the true $\boldsymbol{\mu}_j^*$'s and \mathbf{C}_j^* 's of data set 1 in Figure 1(a) are

$$\boldsymbol{\mu}_1^* = \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}, \quad \boldsymbol{\mu}_2^* = \begin{pmatrix} 1.0 \\ 5.0 \end{pmatrix}, \quad \boldsymbol{\mu}_3^* = \begin{pmatrix} 5.0 \\ 5.0 \end{pmatrix}, \\ \mathbf{C}_1^* = \begin{pmatrix} 0.3 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}, \quad \mathbf{C}_2^* = \begin{pmatrix} 0.2 & -0.1 \\ -0.1 & 0.3 \end{pmatrix}, \quad (13) \\ \mathbf{C}_3^* = \begin{pmatrix} 0.30 & -0.20 \\ -0.20 & 0.25 \end{pmatrix},$$

Initialization: Given a specific k ($k \geq k^*$, k^* is the true number of clusters), initialize the parameter Θ .

Step 1: Given the current input \mathbf{x}_t and the parameter estimate, written as $\Theta^{(n)}$, compute $h(j | \mathbf{x}_t, \Theta^{(n)})$'s and $g(j | \mathbf{x}_t, \Theta^{(n)})$'s via (2) and (7), respectively.

Step 2: Given $h(j | \mathbf{x}_t, \Theta^{(n)})$'s and $g(j | \mathbf{x}_t, \Theta^{(n)})$'s, we update Θ by $\Theta^{(n+1)} = \Theta^{(n)} + \eta(\omega_t(\Theta; \mathbf{x}_t)/\Theta)|_{\Theta^{(n)}}$, with $\omega_t(\Theta; \mathbf{x}_t) = \sum_{j=1}^k g(j | \mathbf{x}_t, \Theta) \ln[\alpha_j p(\mathbf{x}_t | \theta_j)]$, where η is a small positive learning rate.

Step 3: Let $n = n + 1$, and go to *Step 1* for the next iteration until Θ is converged.

ALGORITHM 1: Adaptive RPEM algorithm.

Initialization: Given a specific k ($k \geq k^*$, k^* is the true number of clusters), initialize the parameter Θ .

Step 1: Given $\Theta^{(n)}$, we compute $h(j | \mathbf{x}_t, \Theta^{(n)})$'s and $g(j | \mathbf{x}_t, \Theta^{(n)})$'s for all \mathbf{x}_t 's via (2) and (7), respectively.

Step 2: Fixing $h(j | \mathbf{x}_t, \Theta^{(n)})$'s and $g(j | \mathbf{x}_t, \Theta^{(n)})$'s, we update Θ by $\alpha_j^{(n+1)} = \phi_j^{(n)} / \sum_{j=1}^k \phi_j^{(n)}$, $\mu_j^{(n+1)} = (1/\phi_j^{(n)}) \sum_{t=1}^N \mathbf{x}_t g(j | \mathbf{x}_t, \Theta^{(n)})$, $\mathbf{C}_j^{(n+1)} = (1/\phi_j^{(n)}) \sum_{t=1}^N g(j | \mathbf{x}_t, \Theta^{(n)}) (\mathbf{x}_t - \mu_j^{(n)}) (\mathbf{x}_t - \mu_j^{(n)})^T$, where $\phi_j^{(n)} = \sum_{t=1}^N g(j | \mathbf{x}_t, \Theta^{(n)})$.

Step 3: Let $n = n + 1$, and go to *Step 1* for the next iteration until Θ is converged.

ALGORITHM 2: Batch RPEM algorithm.

while the true parameters of data set 2 in Figure 1(b) are

$$\begin{aligned} \mu_1^* &= \begin{pmatrix} 1.0 \\ 1.0 \end{pmatrix}, & \mu_2^* &= \begin{pmatrix} 1.0 \\ 2.5 \end{pmatrix}, & \mu_3^* &= \begin{pmatrix} 2.5 \\ 2.5 \end{pmatrix}, \\ \mathbf{C}_1^* &= \begin{pmatrix} 0.3 & 0.1 \\ 0.1 & 0.4 \end{pmatrix}, & \mathbf{C}_2^* &= \begin{pmatrix} 0.3 & 0.0 \\ 0.0 & 0.3 \end{pmatrix}, & (14) \\ \mathbf{C}_3^* &= \begin{pmatrix} 0.30 & -0.05 \\ -0.05 & 0.25 \end{pmatrix}. \end{aligned}$$

It can be seen that the clusters in data set 1 are well separated, whereas the clusters in data set 2 are overlapped considerably.

For each data set, we conducted the three experiments by setting $k = 3$, $k = 8$, and $k = 20$, respectively. Also, all α_j 's and \mathbf{C}_j 's were initialized at $1/k$ and the identity matrix, respectively. During the learning process, we discarded those densities whose covariance matrices \mathbf{C}_j 's were singular. Table 1 shows the performance of the batch RPEM over the parameter ε . We found that, as $k = 3$ and $k = 8$, all ε 's we have tried from -0.9 to -0.1 lead to the good performance of the algorithm when using the data set 1. For example, as $k = 8$ and $\varepsilon = -0.8$, we randomly initialized the eight seed points in the input space as shown in Figure 2(a). After all the parameters were converged, 2 out of 8 density components had been discarded and the mixture proportions of the remaining components were converged to $\alpha_1 = 0.2960$, $\alpha_2 = 0.0036$, $\alpha_3 = 0.2900$, $\alpha_4 = 0.0058$,

$\alpha_5 = 0.0136$, and $\alpha_6 = 0.3910$. It can be seen that the three principal mixing proportions, α_1 , α_3 , and α_6 , have well estimated the true ones, while the other proportions were inclined to zero. The corresponding three μ_j 's and \mathbf{C}_j 's were

$$\begin{aligned} \mu_1 &= \begin{pmatrix} 5.06 \\ 4.96 \end{pmatrix}, & \mu_3 &= \begin{pmatrix} 0.98 \\ 4.98 \end{pmatrix}, & \mu_6 &= \begin{pmatrix} 1.00 \\ 0.96 \end{pmatrix}, \\ \mathbf{C}_1 &= \begin{pmatrix} 0.29 & -0.17 \\ -0.17 & 0.22 \end{pmatrix}, & \mathbf{C}_3 &= \begin{pmatrix} 0.18 & -0.08 \\ -0.08 & 0.25 \end{pmatrix}, \\ \mathbf{C}_6 &= \begin{pmatrix} 0.29 & 0.19 \\ 0.19 & 0.39 \end{pmatrix}. \end{aligned} \quad (15)$$

As shown in Figure 2(b), the three μ_j 's have successfully stabilized at the corresponding cluster centers, meanwhile the other three redundant seed points have been pushed away and stably located at the boundary of the clusters. That is, the redundant densities have been fade out through the learning, thus the batch RPEM can select the model automatically as well as the adaptive version.

Nevertheless, when k is set at a large value, for example, say $k = 20$, it is found that the proposed algorithm could not fade out the redundant density components from a mixture if ε is close to 0. Instead, we should set ε at a value close to -1 . For example, as $k^* = 3$, $k = 20$, and $\varepsilon = -0.9$, we ran the proposed algorithm. It was found that

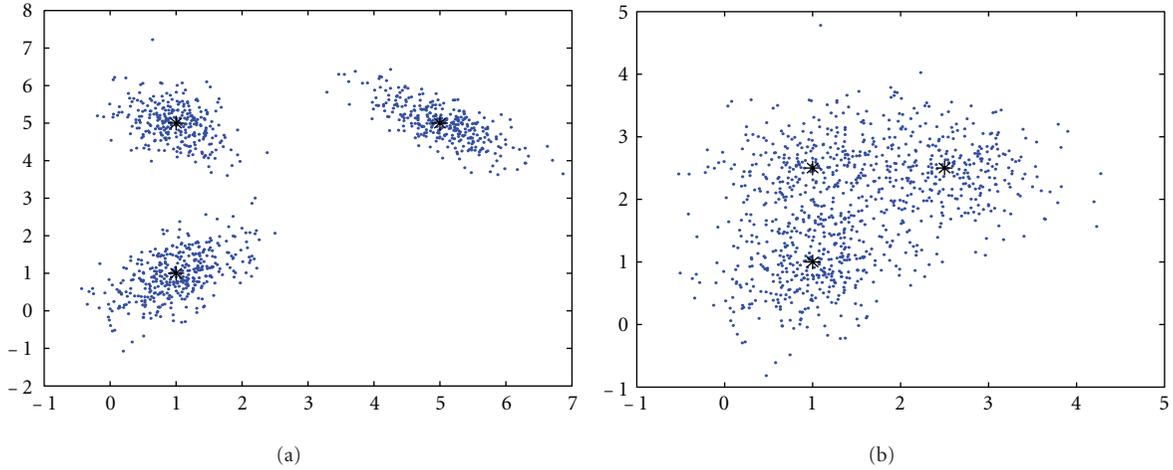


FIGURE 1: (a) Synthetic data set 1 with the well-separated clusters, and (b) synthetic data set 2 with the clusters overlapped considerably.

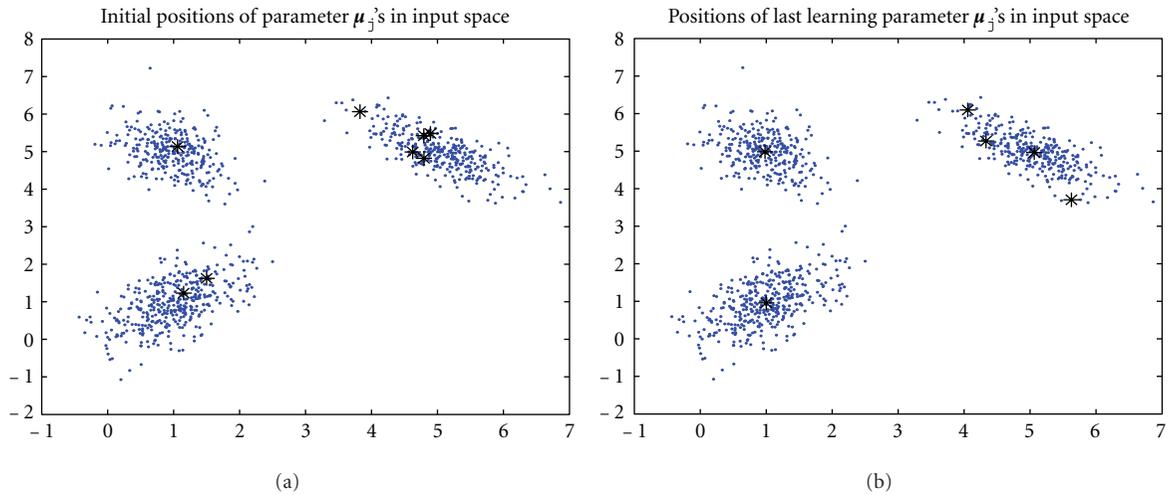


FIGURE 2: The performance of the batch RPEM as $k^* = 3$, $k = 8$, and $\varepsilon = -0.8$: (a) initial positions of seed points; (b) converged positions of seed points.

13 of 20 seed points were maintained by discarding those whose covariance matrices C_j 's were singular. The mixture proportions of the remaining components were converged to $\alpha_1 = 0.0421$, $\alpha_2 = 0.0169$, $\alpha_3 = 0.0051$, $\alpha_4 = 0.2349$, $\alpha_5 = 0.0036$, $\alpha_6 = 0.0149$, $\alpha_7 = 0.3444$, $\alpha_8 = 0.0049$, $\alpha_9 = 0.0210$, $\alpha_{10} = 0.0029$, $\alpha_{11} = 0.0057$, $\alpha_{12} = 0.2944$, and $\alpha_{13} = 0.0091$. The three principal mixing proportions, α_4 , α_7 , and α_{12} , have also well estimated the true ones while the other proportions were tended to zero. Furthermore, the corresponding μ_j 's were $\mu_1 = [1.0872, 4.9986]^T$, $\mu_2 = [0.9897, 0.9640]^T$, and $\mu_3 = [5.0754, 4.9552]^T$. As shown in Figure 3(a), the learned μ_j 's are correctly allocated at the center of the three clusters and the other redundant seed points were driven away to the boundaries of clusters. Hence, the batch algorithm performed a good model selection by assigning $\varepsilon = -0.9$. In contrast, if we assign ε to some value close to zero, the algorithm will lead to a poor model selection. We take $\varepsilon = -0.1$ for instance. The mixture proportions of the remaining 19 out of 20 components were

converged to $\alpha_1 = 0.0461$, $\alpha_2 = 0.0121$, $\alpha_3 = 0.0439$, $\alpha_4 = 0.1404$, $\alpha_5 = 0.0070$, $\alpha_6 = 0.0258$, $\alpha_7 = 0.0178$, $\alpha_8 = 0.0348$, $\alpha_9 = 0.0659$, $\alpha_{10} = 0.0513$, $\alpha_{11} = 0.0493$, $\alpha_{12} = 0.0352$, $\alpha_{13} = 0.0362$, $\alpha_{14} = 0.0528$, $\alpha_{15} = 0.0587$, $\alpha_{16} = 0.0171$, $\alpha_{17} = 0.1916$, $\alpha_{18} = 0.0882$, and $\alpha_{19} = 0.0260$. It can be seen that none of α_j 's tends to zero. As shown in Figure 3(b), all the converged positions have a bias from the cluster centers. In other words, the algorithm has a poor performance as ε get close to zero. Hence, if k is large, it would be better to choose a relative smaller value of ε between -1 and 0 .

In addition, we also investigated the assignment of ε on data set 2, where the data are considerably overlapped. We take $k^* = 3$, $k = 20$, and $\varepsilon = -0.9$ for instance. The converged positions of the seed points are shown in Figure 4(a), where the learned positions converged to the cluster centers while driving the redundant seed points to the boundaries of the clusters. That is, the proposed batch algorithm can work quite well as $\varepsilon = -0.9$. Also, we let $k^* = 3$, $k = 20$, and $\varepsilon = -0.2$ to run the algorithm again

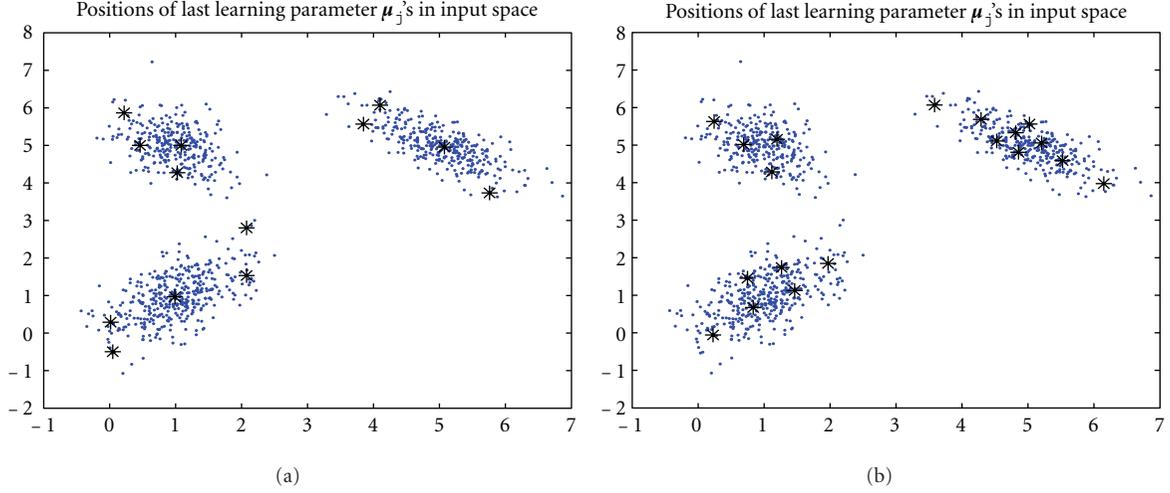


FIGURE 3: The converged positions of the seed points as $k^* = 3$ and $k = 20$: (a) $\varepsilon = -0.9$, (b) $\varepsilon = -0.1$.

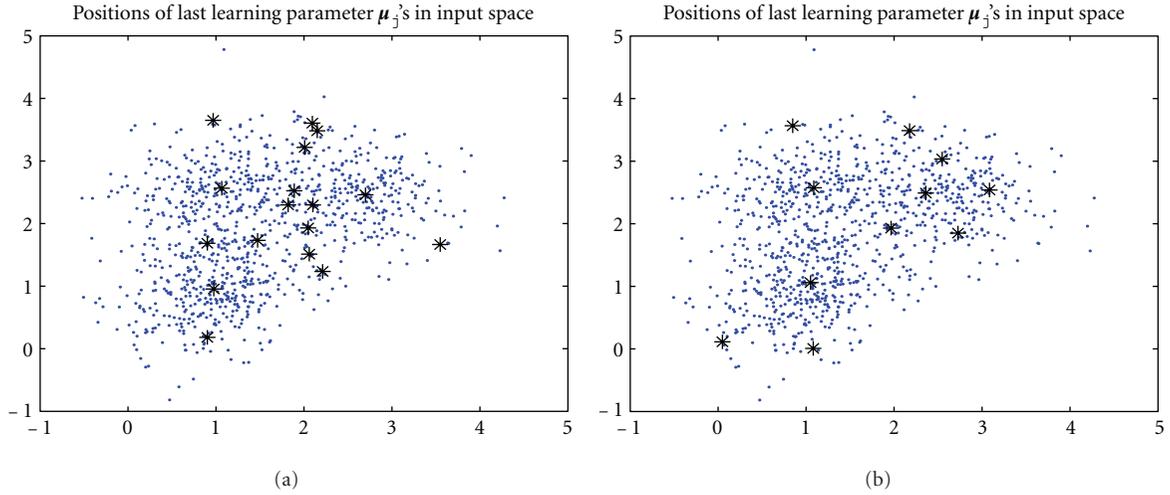


FIGURE 4: The converged positions of the seed points learned via the batch RPEM as $k^* = 3$ and $k = 20$: (a) $\varepsilon = -0.9$; (b) $\varepsilon = -0.2$.

for comparison. As shown in Figure 4(b), the converged positions of the seed points have a bias from the cluster centers. This implies that the values of ε that are close to zero cannot work well in this case. More examples can be found in Table 1. It can be seen that the feasible region of ε is shrunk over the overlap level of the data. For example, the appropriate values of ε are in the range of $[-0.9, -0.6]$ only when using the date set 2 with $k^* = 3$ and $k = 3$ or 8, respectively. In contrast, ε is feasible in the full range of $[-0.9, -0.1]$ where we have tried so far as data set 1 is used. Hence, by a rule of thumb, we should choose an appropriate value of ε close to -1 . Nevertheless, we have also noted that it is not a good choice if ε is too close to -1 . In fact, the proposed algorithm will gradually degenerate to the EM as ε tends to -1 ; that is, the capability of the proposed algorithm on model selection will be reduced as ε tends to -1 . Hence, to sum up, empirical studies have found that $[-0.9, -0.8]$ is an appropriate feasible region of ε . In the next section, we therefore arbitrarily set ε at -0.8 .

5. Experimental Results

To evaluate the performance of the batch RPEM algorithm, we have conducted the following three experiments.

5.1. Experiment 1: Batch RPEM on Synthetic Data with $K = K^$.* This experiment was to evaluate the convergence speed of the batch RPEM. We utilized 1,000 data points from a mixture of three bivariate Gaussian densities with the true parameters as follows:

$$\begin{aligned} \alpha_1^* &= 0.3, & \alpha_2^* &= 0.4, & \alpha_3^* &= 0.3, \\ \boldsymbol{\mu}_1^* &= [1.0, 1.0]^T, & \boldsymbol{\mu}_2^* &= [1.0, 2.5]^T, & \boldsymbol{\mu}_3^* &= [2.5, 2.5]^T, \\ \mathbf{C}_1^* &= \begin{pmatrix} 0.20 & 0.05 \\ 0.05 & 0.30 \end{pmatrix}, & \mathbf{C}_2^* &= \begin{pmatrix} 0.2 & 0.0 \\ 0.0 & 0.2 \end{pmatrix}, \\ & & \mathbf{C}_3^* &= \begin{pmatrix} 0.2 & -0.1 \\ -0.1 & 0.2 \end{pmatrix}. \end{aligned} \tag{16}$$

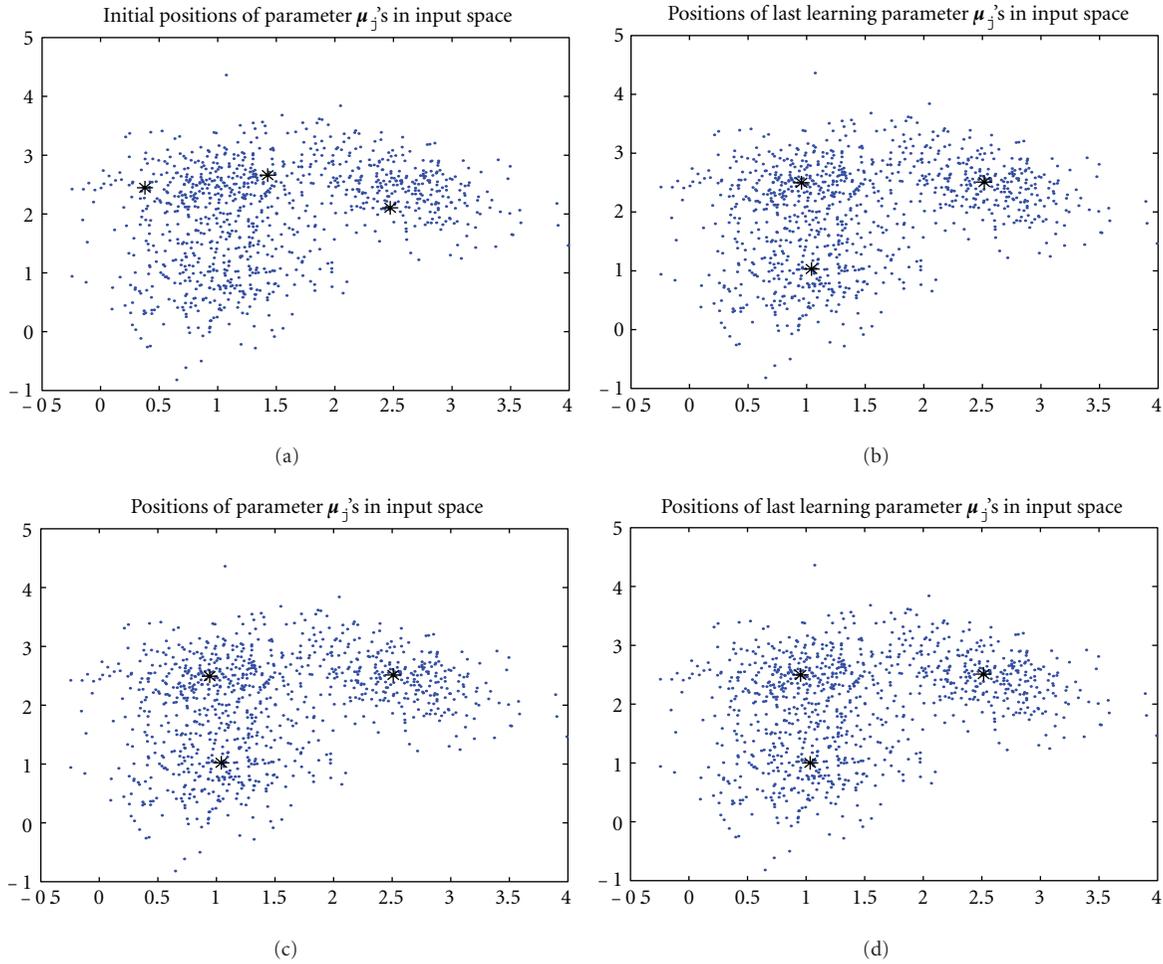


FIGURE 5: (a) The initial positions of the three seed points and their converged positions learned by (b) EM, (c) adaptive RPEM, and (d) batch RPEM, respectively.

TABLE 1: Performance of the Batch RPEM over the Parameter ε , where “G” stands for a good model selection capability of the algorithm, while “P” represents a poor model selection capability.

| ε | $k^* = 3, k = 3$ | | $k^* = 3, k = 8$ | | $k^* = 3, k = 20$ | |
|---------------|------------------|------------|------------------|------------|-------------------|------------|
| | Data set 1 | Data set 2 | Data set 1 | Data set 2 | Data set 1 | Data set 2 |
| -0.9 | G | G | G | G | G | G |
| -0.8 | G | G | G | G | G | G |
| -0.7 | G | G | G | G | G | G |
| -0.6 | G | G | G | G | G | G |
| -0.5 | G | P | G | P | G | G |
| -0.4 | G | P | G | P | P | P |
| -0.3 | G | P | G | P | P | P |
| -0.2 | G | P | G | P | P | P |
| -0.1 | G | P | G | P | P | P |

We let $k = 3$, which is equal to the true mixture number $k^* = 3$. The three seed points were randomly allocated in the observation space as shown in Figure 5(a), where the data are considerably overlapped. Moreover, all α_j 's and C_j 's

were initialized at $1/k$ and the identity matrix, respectively. Figure 5(d) shows the positions of the three converged seed points, which are all stably located at the corresponding cluster centers. For comparison, we also implemented the EM under the same experimental environment. Figure 5(b) shows that the EM had successfully located the three seed points as well as the batch RPEM.

Nevertheless, as shown in Figures 6(c) and 7, the batch RPEM converges at 20 epochs, while the EM needs 60 epochs as shown in Figure 6(a). That is, the convergence speed of the batch RPEM is significantly faster than the EM. This indicates that the intrinsic rival-penalization scheme of the batch RPEM, analogous to the RPCL [11], RPCCL [12], and the adaptive RPEM [14], is able to drive away the rival seed points so that they can be more quickly towards the other cluster centers. As a result, the batch RPEM converges much faster than the EM. Moreover, we also compared it with the adaptive RPEM, in which we set the learning rate $\eta = 0.001$. Figure 5(c) shows the convergent results of the seed points. It can be seen that the adaptive RPEM works quite well in this case, but it needs around 70 epochs as shown in Figure 6(b). Actually, the adaptive RPEM can be further speed up if an

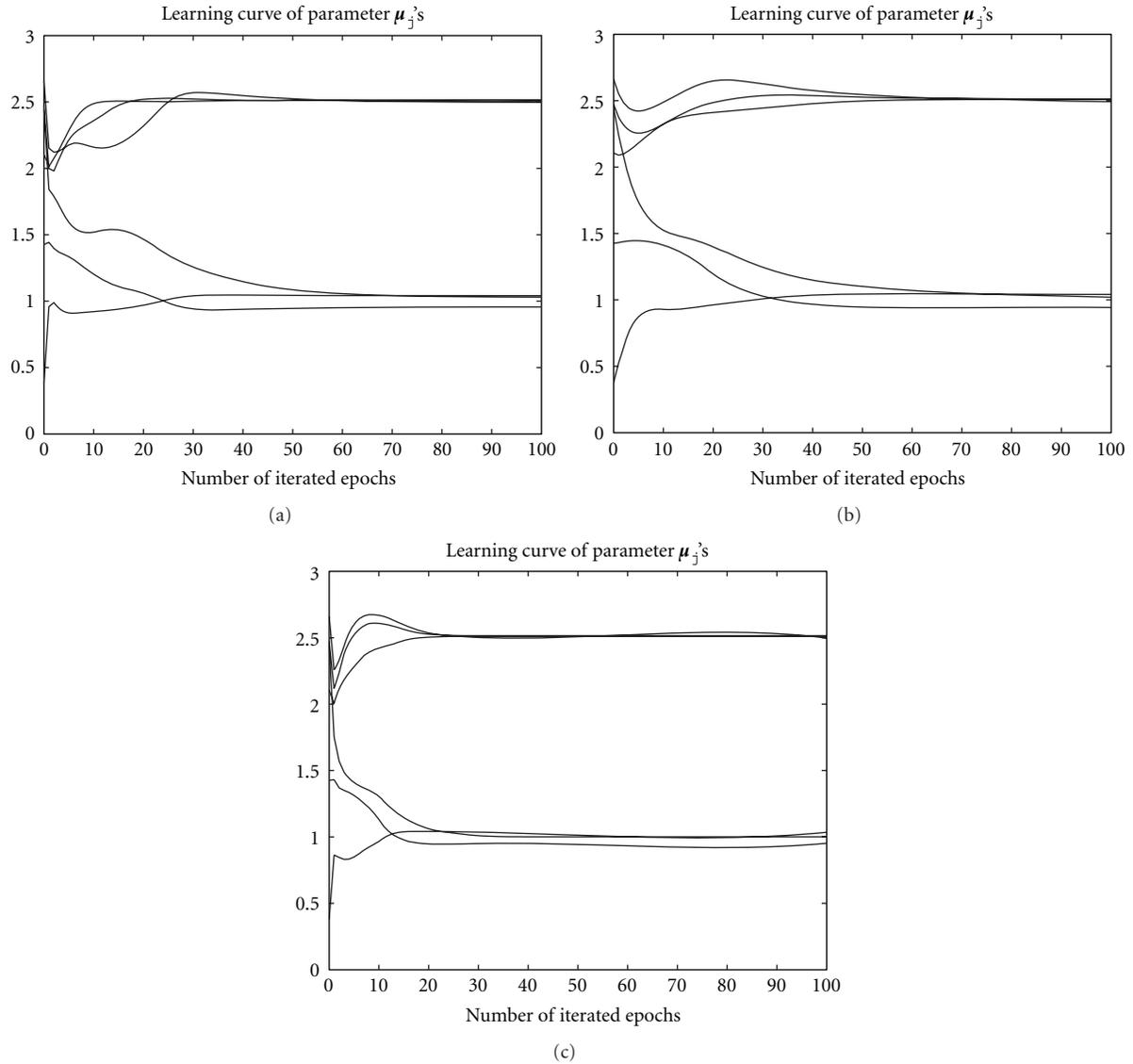


FIGURE 6: Learning curves of μ_j 's by (a) EM, (b) adaptive RPEM, and (c) batch RPEM, respectively.

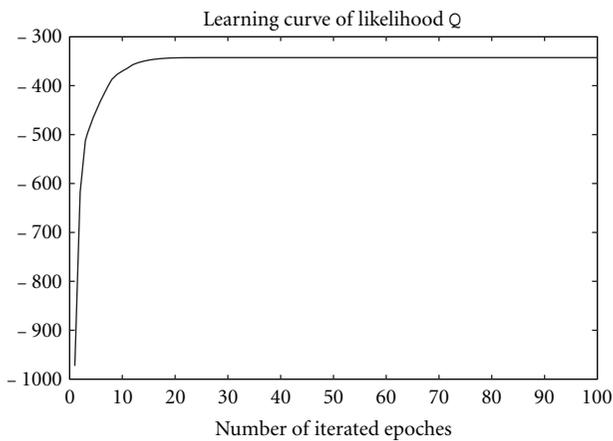


FIGURE 7: The value of the cost function Q over the epochs.

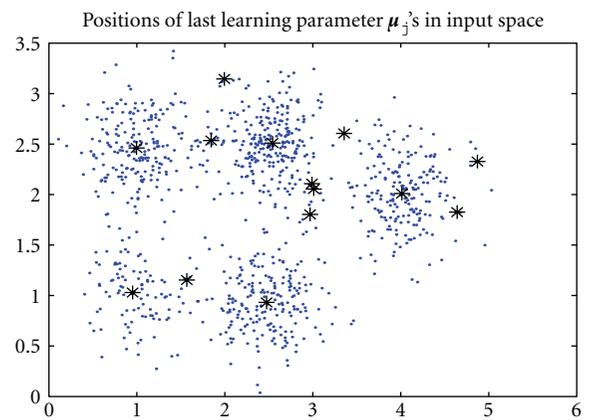


FIGURE 8: The converged positions of the seed points learned by the batch RPEM.

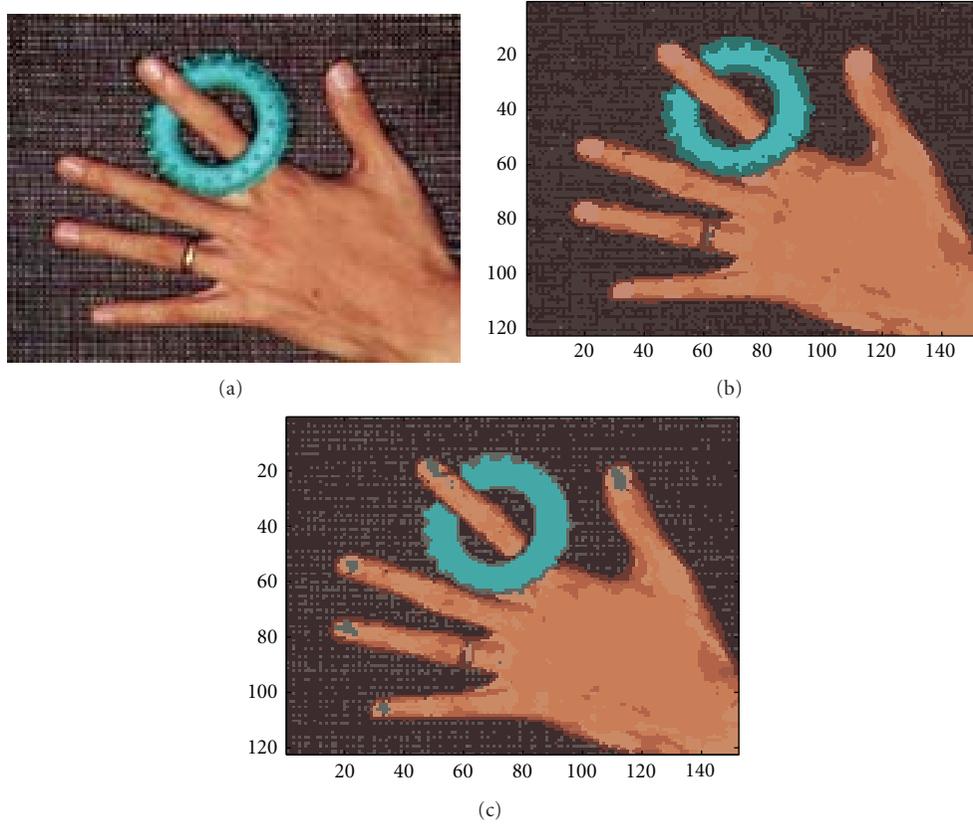


FIGURE 9: Segmentation of the hand image: (a) original image, (b) the result given by the EM, and (c) the result given by the batch RPEM.



FIGURE 10: The original house image.

appropriate learning rate is adopted, which, however, is not a trivial task.

5.2. Experiment 2: Batch RPEM on Synthetic Data with $K > K^*$. This experiment will investigate the performance of batch RPEM performance as $k > k^*$. We generated 1,000 observations from a mixture of five bivariate Gaussian density distributions with the mixing proportions:

$$\alpha_1^* = 0.1, \quad \alpha_2^* = 0.2, \quad \alpha_3^* = 0.3, \quad \alpha_4^* = 0.2, \quad \alpha_5^* = 0.2 \quad (17)$$

and the true cluster centers:

$$\begin{aligned} \boldsymbol{\mu}_1^* &= [1.0, 1.0]^T, & \boldsymbol{\mu}_2^* &= [1.0, 2.5]^T, & \boldsymbol{\mu}_3^* &= [2.5, 2.5]^T, \\ \boldsymbol{\mu}_4^* &= [2.5, 1.0]^T, & \boldsymbol{\mu}_5^* &= [4.0, 2.0]^T. \end{aligned} \quad (18)$$

15 seed points were initialized in the input space arbitrarily. During the learning, the three density components were discarded because their corresponding covariances became singular. As a result, the remaining 12 converged proportions were $\alpha_1 = 0.0065$, $\alpha_2 = 0.0113$, $\alpha_3 = 0.1929$, $\alpha_4 = 0.0030$, $\alpha_5 = 0.0068$, $\alpha_6 = 0.2013$, $\alpha_7 = 0.2084$, $\alpha_8 = 0.0074$, $\alpha_9 = 0.0083$, $\alpha_{10} = 0.0986$, $\alpha_{11} = 0.2531$, and $\alpha_{12} = 0.0022$. It can be seen that the five principal values $\alpha_3, \alpha_6, \alpha_7, \alpha_{10}$, and α_{11} were estimated well, while the others were learned towards zero. A snapshot of the corresponding $\boldsymbol{\mu}_j$'s were $\boldsymbol{\mu}_3 = [4.0348, 2.0075]^T$, $\boldsymbol{\mu}_6 = [0.9990, 2.4571]^T$, $\boldsymbol{\mu}_7 = [2.4725, 0.9220]^T$, $\boldsymbol{\mu}_{10} = [0.9553, 1.0277]^T$, and $\boldsymbol{\mu}_{11} = [2.5189, 2.5199]^T$. As shown in Figure 8, these five seed points have successfully allocated in the cluster centers, meanwhile the batch RPEM drove the redundant seed points to the boundaries of the clusters.

5.3. Experiment 3: Batch RPEM on Color Image Segmentation. This experiment further investigated the batch RPEM algorithm on color image segmentation in comparison to the EM algorithm. We implemented the image segmentation in

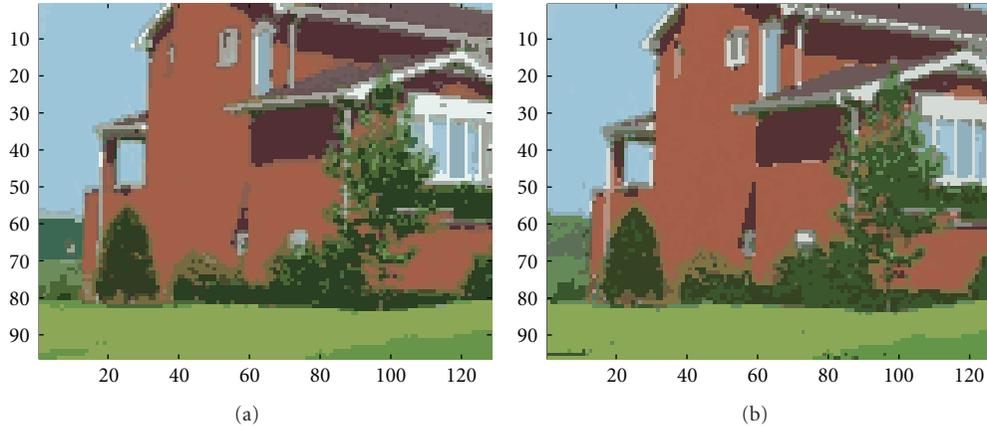


FIGURE 11: Segmentation of the house image by (a) EM; (b) batch RPEM.

the red-green-blue (RGB) color space model that represents each pixel in an image by a three-color vector. We conducted color image segmentation on a 122×152 hand image and a 96×128 house image as shown in Figures 9(a) and 10, respectively. For the former, we initially assigned 10 seed points randomly. After the convergence of the algorithms' performance, a snapshot of their segmentation results is shown in Figures 9(b) and 9(c). It can be seen that the blue tiny swim ring-shaped region after segmentation process by the batch RPEM is smoother than the EM. Further, the tiny nail regions have been partitioned by the batch RPEM but the EM is not. In other words, the batch RPEM algorithm performs better than the EM algorithm.

For the house image, we initially assigned the seed points to be 80. A snapshot of the converged segmentation results of the EM and the batch RPEM is shown in Figure 11. It can be seen that the texture on the red wall and the green lawn has no longer maintained after the segmentation process both by the EM and the RPEM. However, the small white regions of windows on red wall were disappeared by the EM as well as the triangle shadow area on the wall. In contrast, the batch RPEM algorithm partitioned these regions well as shown in Figure 11(b). Actually, the batch RPEM has drove out the redundant seed points far away and maintained some principal components correctly, which therefore leads to a better performance in color image segmentation.

6. Conclusion

In this paper, we have developed a batch RPEM algorithm based on MWL learning framework for Gaussian mixture clustering. Compared to the adaptive RPEM, this new one need not select the value of learning rate. As a result, it can learn faster in general and still preserve the capability of automatic model selection analogous to the adaptive one. We have evaluated the proposed batch RPEM algorithm on both synthetic data and color image segmentation. The numerical results have shown the efficacy of the proposed algorithm.

Acknowledgments

This work was jointly supported by the grants from the Research Grant Council of the Hong Kong SAR with the

Project Code: HKBU 210309, the Natural Science Foundation of China (60974077), the Natural Science Foundation of Guangdong Province (s2011010005075), Guangzhou Technology Projects (11c42110781), the Grants 60403011 and 60973154 from the NSFC, and NCET-07-0338 from the Ministry of Education, China. This work was also partially supported by the Fundamental Research Funds for the Central Universities, HUST:2010ZD025, and Hubei Provincial Science Foundation under Grant 2010CDA006, China.

References

- [1] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, MIT Press, 1996.
- [2] B. Fritzke, "The LBG-U method for vector quantization— an improvement over LBG inspired from neural networks," *Neural Processing Letters*, vol. 5, no. 1, pp. 35–45, 1997.
- [3] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, 1980.
- [4] S. Y. Chen, H. Y. Tong, Z. J. Wang, S. Liu, M. Li, and B. W. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, Article ID 416963, 12 pages, 2011.
- [5] S. Y. Chen, H. Y. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
- [6] Y. W. Lim and S. U. Lee, "On the color image segmentation algorithm based on the thresholding and the fuzzy C-means techniques," *Pattern Recognition*, vol. 23, no. 9, pp. 935–952, 1990.
- [7] T. Uchiyama and M. A. Arib, "Color image segmentation using competitive learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 12, pp. 1197–1206, 1994.
- [8] J. M. Jolion, P. Meer, and S. Bataouche, "Robust clustering with applications in computer vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 791–802, 1991.
- [9] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, no. 1, pp. 1–38, 1977.
- [10] Bilmes A. Jeff, *A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gausssian Mixture and Hidden Markov Models*, International Computer Science

- Institute, and Computer Science Division, Department of Electrical Engineering and Computer Science, Berkeley, Calif, USA, 1998, TR-97-021.
- [11] L. Xu, A. Krzyzak, and E. Oja, "Rival penalized competitive learning for clustering analysis, RBF Net, and curve detection," *IEEE Transactions on Neural Networks*, vol. 4, no. 4, pp. 636–649, 1993.
 - [12] Y. M. Cheung, "Rival penalized controlled competitive learning for data clustering with unknown cluster number," in *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP '02)*, 2002.
 - [13] Y. M. Cheung, "A rival penalized EM algorithm towards maximizing weighted likelihood for density mixture clustering with automatic model selection," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 4, pp. 633–636, Cambridge, UK, 2004.
 - [14] Y. M. Cheung, "Maximum weighted likelihood via rival penalized EM for density mixture clustering with automatic model selection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 750–761, 2005.
 - [15] X. M. Zhao, Y. M. Cheung, L. Chen, and K. Aihara, "A new technique for adjusting the learning rate of RPEM algorithm automatically," in *Proceedings of the 12th International Symposium on Artificial Life and Robotics*, pp. 597–600, Oita, Japan, January 2007.

Research Article

The Finite Element Method Applied to a Problem of Blood Flow in Vessels

Gabriela Nuț,¹ Ioana Chiorean,¹ and Maria Crișan²

¹ Faculty of Mathematics and Computer Science, Babeș-Bolyai University 1, Mihail Kogălniceanu Street, 400084 Cluj Napoca, Romania

² Clinic of Dermatology, Department of Histology, Iuliu Hatieganu University of Medicine and Pharmacy 8 Victor Babeș Street, 400012 Cluj-Napoca, Romania

Correspondence should be addressed to Maria Crișan, mcrisan7@yahoo.com

Received 29 August 2011; Accepted 10 October 2011

Academic Editor: Carlo Cattani

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

We use the finite element method to solve a convection-diffusion equation when convection is dominating, a problem which describes the behavior of the concentration of a solute in a blood vessel. A new technique for computing the discrete problem is used.

1. Introduction

Due to the fact that various chemicals, such as oxygen, carbon dioxide, or lipids, are transported by blood to and from the other tissues, including the skin, the study of the blood flow in the human vascular system is of great interest in medicine and medical engineering.

The skin is the largest organ of the human body, having very important psychosocial implications. It represents the only system completely displayed at the body surface, offering essential information regarding the homeostasis of the internal organs and thus the general senescence process. Senescence involves a complex of factors, of which vascularisation plays an important role. The perfusion degree of the tissues depends on the microscopical structure of the vascular walls, as well as the metabolical and biochemical changes associated with age. It is well known that the collagen glycation processes that occur in the vascular walls, and are involved in the senescence process, are associated with vascularisation deficits, that are specific to the ageing process generally, and to the associated pathology specifically (atherosclerosis, Alzheimer, metabolical diseases, rheumatoid arthritis). These modifications can be observed at the skin level by means of images techniques (see [1, 2]).

These are only few reasons for which the study of the possible solutes in blood vessels is so important. In this sense, many studies were made and different mathematical models were given, depending on various factors, such as the health

state of the patient (the existence of pathologies like atherosclerosis, etc.). For instance, in [3], an implementation of adaptive anisotropic meshes for this class of problems by developing an a posteriori error analysis for a simpler situation, namely, a single steady advection-diffusion-reaction equation with a given convective field, is given.

2. The Model Problem

In order to study the transport problem of a solute in a vessel, we consider the following partial derivative equation:

$$\begin{aligned} m\Delta c + \mathbf{n}\nabla c &= f, & (x, y) \in \Omega, \\ c &= g_1, & (x, y) \in \Gamma_{\text{in}}, \\ c &= g_2, & (x, y) \in \Gamma_{\text{out}}, \\ c &= \xi g_3, & (x, y) \in \Gamma. \end{aligned} \quad \mathbf{n} = (n_1, n_2). \quad (1)$$

Here, $c = c(x, y)$ represents the concentration of the solute in the blood. The first term in (1) describes the diffusion, the later contains two more terms: one for the diffusion along the x -direction: $n_1(\partial c/\partial x)$, and another for the diffusion along the y -direction: $n_2(\partial c/\partial y)$. m represents the diffusivity of the solute, \mathbf{n} a given velocity field, α is a reaction coefficient, f —a possible forcing term for the solution concentration due, for example, to chemical reactions.

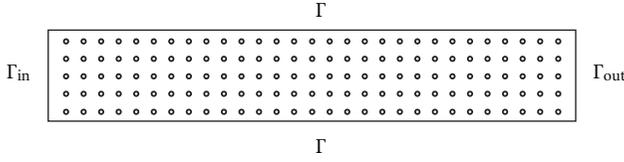


FIGURE 1

The ratio between coefficients m and \mathbf{n} determines the predominance of the convection or the diffusion in the physical process. If:

$$\frac{|\mathbf{n}|}{m} \gg 1, \quad (2)$$

then the convection is dominant. computing the numerical solution of a convection–diffusion problem of the form (1) becomes increasingly difficult (converge slow or not at all) as the ratio (2) increases (i.e., convection is dominant in the process), and this is the case in the model problem (1) (the ratio here is 10^4 , see Section 5).

The domain Ω chosen for this problem represents a part of an artery (considered to be rigid) and is shown in Figure 1. The borders of Ω are as follows.

- (i) Γ_{in} , which corresponds to the point in where the solution is injected in the blood, thus the concentration of the solution is known here: $u = g_1, (x, y) \in \Gamma_{\text{in}}$.
- (ii) Γ_{out} , which corresponds to the point where the concentration of the solution can be measured: $u = g_2, (x, y) \in \Gamma_{\text{out}}$.
- (iii) Γ : in these points the solution is in contact with the vessel walls, ξ represents the wall permeability, and g_3 is a function that decreases from g_1 to g_2 .

3. The Discretization of the Problem

In this paper, we use the finite element method (see [4, 5]). In order to do this, as in [6–8], the domain $[a, b] \times [c, d]$ is divided in rectangular subdomains.

For the discretization of the problem we use the finite element method. The domain is divided in rectangular subdomains, having the step h_x on the Ox -direction and h_y on Oy . The solution of the systems generated through discretization is obtained by Gauss full elimination method. The first level on which the solution is computed is l_0 , then this particular solution is used for obtaining the solution on higher order levels. The grid on the l level is divided by the one from the l_0 level in subdomains. On each of these, the corresponding system of linear equations will be solved, and the solutions from the subdomains are used in order to generate the solution on the l level.

The partial differential equations will be replaced by a linear system of equations through the discretization method.

In order to achieve this, and keeping the notations used in [7], we choose the grid steps $h_{lx} = (b - a)/2^{l+1}$ and $h_{ly} = (d - c)/2^{l+1}$, l being the number of the level. The corresponding number of interior grid points is $n_l = 2^{l+1} - 1$

on each direction. The largest possible step corresponds to the level denoted by $l = 0$ on which the grid has a single point: $((b - a)/2, (d - c)/2)$. The grid on the level l will contain the points (x_i, y_j) , $i, j = 1, 2, \dots, n_l$, and will be denoted by G_l . The value of the exact solution in the point (x_i, y_j) is denoted by $c_{i,j}$.

3.1. Finite Element Discretization. According to [4], in order to apply the finite element discretization, some transformations of the given equation have to be made. So, the equation to be discretized is multiplied by a test function $v \in U(\Omega) = \{v \mid v \in H^1(\Omega), v = g \text{ on } \Gamma\}$, then is integrated on the domain Ω ; $H^1(\Omega) \subset C^0$ is the space of functions with square integrable derivatives on Ω ,

$$-m \iint_{\Omega} \Delta c v \, dx \, dy + \iint_{\Omega} \mathbf{n} \nabla c v \, dx \, dy + \alpha \iint_{\Omega} c v \, dx \, dy = \iint_{\Omega} f v \, dx \, dy. \quad (3)$$

Using Green's formula, the equation above becomes as follows:

$$m \iint_{\Omega} \nabla c \nabla v \, dx \, dy - m \int_{\delta\Omega} \frac{\partial c}{\partial n} v \, ds + \iint_{\Omega} \mathbf{n} \nabla c v \, dx \, dy + \alpha \iint_{\Omega} c v \, dx \, dy = \iint_{\Omega} f v \, dx \, dy, \quad u, v \in U(\Omega). \quad (4)$$

The functions c and v are approximated using some continuous functions, Φ_i ($\Phi_i(x_j, y_j) = \delta_{ij}$, $i, j = 1, \dots, N$, $N = n_l^2$ being the number of interior points of the grid on level l), through the following relations:

$$c \approx \sum_{i=1}^N u_i \Phi_i, \quad v \approx \sum_{j=1}^N v_j \Phi_j, \quad (5)$$

where $c_i = c(x_i, y_i)$, $i = 1, \dots, N$. Replacing these approximations in (4), the system obtained is:

$$\sum_{j=1}^N K_{ij} c_j = F_i, \quad i = 1, \dots, N, \quad (6)$$

where

$$K_{ij} = \iint_{\Omega} \left[m \left(\frac{\partial \Phi_i}{\partial x} \frac{\partial \Phi_j}{\partial x} + \frac{\partial \Phi_i}{\partial y} \frac{\partial \Phi_j}{\partial y} \right) + n_1 \Phi_i \frac{\partial \Phi_j}{\partial x} + n_2 \Phi_i \frac{\partial \Phi_j}{\partial y} + \alpha \Phi_i \Phi_j \right] dx \, dy, \quad (7)$$

$$F_i = \iint_{\Omega} f \Phi_i \, dx \, dy.$$

The restrictions of K and F on a domain $\Omega_A = [a, b] \times [c, d]$ are as follows:

$$\begin{aligned} k_{ij}^A &= m \iint_{\Omega_A} \left(\frac{\partial \Psi_i^A}{\partial x} \frac{\partial \Psi_j^A}{\partial x} + \frac{\partial \Psi_i^A}{\partial y} \frac{\partial \Psi_j^A}{\partial y} \right) dx dy \\ &+ n_1 \iint_{\Omega_A} \Psi_i^A \frac{\partial \Psi_j^A}{\partial x} dx dy + n_2 \iint_{\Omega_A} \Psi_i^A \frac{\partial \Psi_j^A}{\partial y} dx dy \\ &+ \alpha \iint_{\Omega_A} \Psi_i^A \Psi_j^A dx dy, \quad i, j = 1, \dots, 4, \end{aligned} \quad (8)$$

$$f_k^A = \iint_{\Omega_A} f(x, y) \Psi_k^A(x, y) dx dy, \quad k = 1, \dots, 4. \quad (9)$$

If $x = (d - c)/(b - a)$, the values of the integral (8) for the problem (1) are given in the following 4×4 matrix:

$$\begin{aligned} k &= (k_{ij})_{i,j=1:4} \\ &= \frac{m}{6} \begin{bmatrix} 2x + \frac{2}{x} & -2x + \frac{1}{x} & -x - \frac{1}{x} & x - \frac{2}{x} \\ -2x + \frac{1}{x} & 2x + \frac{2}{x} & x - \frac{2}{x} & -x - \frac{1}{x} \\ -x - \frac{1}{x} & x - \frac{2}{x} & 2x + \frac{2}{x} & -2x + \frac{1}{x} \\ x - \frac{2}{x} & -x - \frac{1}{x} & -2x + \frac{1}{x} & 2x + \frac{2}{x} \end{bmatrix} \\ &+ n_1 \frac{d-c}{12} \begin{bmatrix} -2 & 2 & 1 & -1 \\ -2 & 2 & 1 & -1 \\ -1 & 1 & 2 & -2 \\ -1 & 1 & 2 & -2 \end{bmatrix} + n_2 \frac{b-a}{12} \begin{bmatrix} -2 & -1 & 1 & 2 \\ -1 & -2 & 2 & 1 \\ -1 & -2 & 2 & 1 \\ -2 & -1 & 1 & 2 \end{bmatrix} \\ &+ \alpha \frac{(b-a)(d-c)}{36} \begin{bmatrix} 4 & 2 & 1 & 2 \\ 2 & 4 & 2 & 1 \\ 1 & 2 & 4 & 2 \\ 2 & 1 & 2 & 4 \end{bmatrix}. \end{aligned} \quad (10)$$

Thus the differential equation (1) is approximated in a grid point (x_i, y_j) , $i, j = 1, \dots, n_l$ by the following system of linear equations (see [9]):

$$\begin{bmatrix} k_{24}^D & k_{23}^D + k_{14}^C & k_{13}^C \\ k_{34}^A + k_{21}^D & k_{33}^A + k_{44}^B + k_{11}^C + k_{22}^D & k_{43}^B + k_{12}^C \\ k_{31}^A & k_{32}^A + k_{41}^B & k_{42}^B \end{bmatrix} c_{i,j} = f_3^A + f_4^B + f_1^C + f_2^D, \quad (11)$$

where

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix} u_{i,j} = au_{i-1,j+1} + bu_{i,j+1} + cu_{i+1,j+1} + du_{i-1,j} \\ + eu_{i,j} + fu_{i+1,j} + gu_{i-1,j-1} + hu_{i,j-1} \\ + ku_{i+1,j-1}. \quad (12)$$

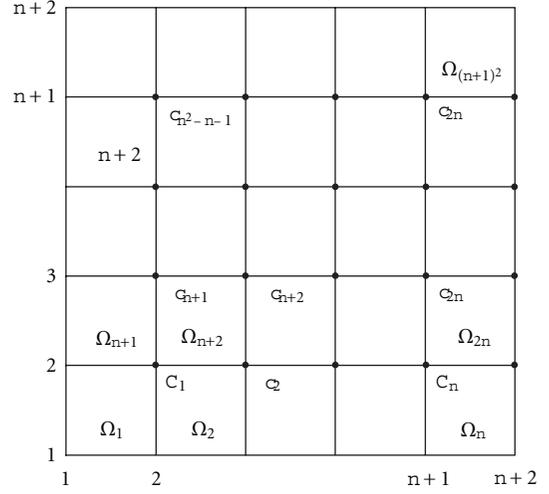


FIGURE 2

4. Solving Method

The systems generated in Section 3 can be written on any level l . Each system contains n_l^2 unknowns. The solution is exactly computed on a level l_0 , for example, on $l_0 = 2$ or $l_0 = 3$ using Gauss elimination method.

The exact solution on the level l_0 , for the problem is approximated by c_i , $i \in \{1, 2, \dots, n_{l_0}^2\}$ (Figure 2), which only contains an error term due to the discretization. In order to solve problem on the level l , the grid already obtained will be further divided. Thus, each domain from the grid, Ω_k , will be splitted into n_i subdomains, where $n_i = 2^{l+1} - 1$ and $l_i = l - l_0 - 1$.

On each subdomain Ω_k , the discretization of the differential equation leads to a system whose matrix has the same form as the one on l_0 level. But on the level l_0 the boundary values were given in the hypothesis. For the systems on the level l to be precisely solved on Ω_k , one has to determine as accurate as possible the n_i values on each of the sides of the domain Ω_k . Two possible ways to accomplish this are given in the following subsections.

4.1. Pondered Arithmetic Mean Prolongation. As in [7], the value of the approximation on level l is denoted by $c^{(l)}$. On the borders of Ω_k , these values are defined through the following relations:

$$\begin{aligned} c_{jN+1, iN+1}^{(l)} &= c_{(i-1)n+j}^{(l_0)}, \\ c_{jN+1, iN+1+k}^{(l)} &= \frac{1}{N} \left(kc_{in+j}^{(l_0)} + (N-k)c_{(i-1)n+j}^{(l_0)} \right), \\ & \quad i = 0, \dots, n, \quad j = 1, \dots, n, \quad N = n_i + 1; \quad (13) \\ c_{jN+1+k, iN+1}^{(l)} &= \frac{1}{N} \left(kc_{(i-1)n+j+1}^{(l_0)} + (N-k)c_{(i-1)n+j}^{(l_0)} \right), \\ & \quad i = 1, \dots, n, \quad j = 0, \dots, n, \quad k = 1, \dots, n_i. \end{aligned}$$

4.2. Stellar Prolongation. In [9], a new type of prolongation which we called "stellar prolongation" because the nodes

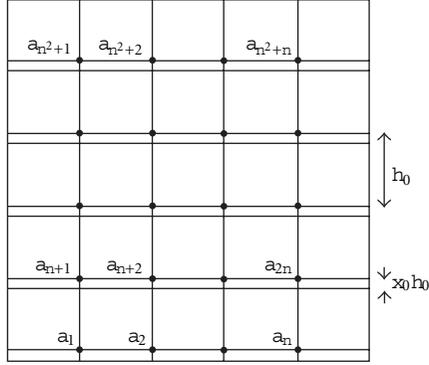


FIGURE 3

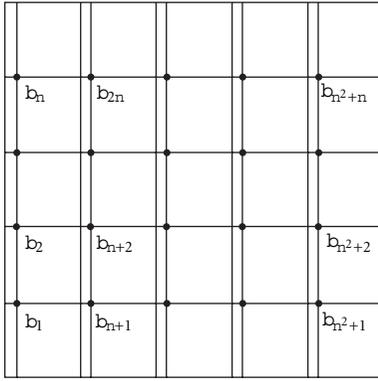


FIGURE 4

involved in computation are in the shape of a star, is presented. We use this technique in what follows, too.

In order to determine more accurately the values of the solution on the borders of Ω_k , instead of pondered arithmetic mean prolongation one can use the solutions of the systems obtained by discretizing the initial equation in the grid points corresponding to the values a_i and b_i , $i = 1, 2, \dots, n^2 + n$ from Figures 3 and 4.

The values a_k , $k = 1, 2, \dots, n(n+1)$ are computed by solving a system with the following matrix:

$$A = \begin{bmatrix} C & D & \Theta & \cdots & \Theta & \Theta \\ S & C & D & \cdots & \Theta & \Theta \\ \Theta & S & C & \cdots & \Theta & \Theta \\ \vdots & & & \ddots & & \\ \Theta & \Theta & \Theta & \cdots & S & C \end{bmatrix}. \quad (14)$$

If the discretization is made by the finite element method, using the notation:

$$\begin{bmatrix} l_1 & l_2 & l_3 \\ l_4 & l_5 & l_6 \\ l_7 & l_8 & l_9 \end{bmatrix} = \begin{bmatrix} k_{24}^D & k_{23}^D + k_{14}^C & k_{13}^C \\ k_{34}^A + k_{21}^D & k_{33}^A + k_{44}^B + k_{11}^C + k_{22}^D & k_{43}^B + k_{12}^C \\ k_{31}^A & k_{32}^A + k_{41}^B & k_{42}^B \end{bmatrix}, \quad (15)$$

where k_{ij} is given by (10), the matrix A has:

$$C = \begin{bmatrix} l_5 & l_6 & 0 & \cdots & 0 \\ l_4 & l_5 & l_6 & \cdots & 0 \\ 0 & l_4 & l_5 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & l_5 \end{bmatrix},$$

$$D = \begin{bmatrix} l_2 & l_3 & 0 & \cdots & 0 \\ l_1 & l_2 & l_3 & \cdots & 0 \\ 0 & l_1 & l_2 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & l_2 \end{bmatrix}, \quad (16)$$

$$S = \begin{bmatrix} l_8 & l_9 & 0 & \cdots & 0 \\ l_7 & l_8 & l_9 & \cdots & 0 \\ 0 & l_7 & l_8 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & l_8 \end{bmatrix}.$$

The components of the constant terms vector are in this case:

$$z_{in+j} = f_3^A + f_4^B + f_1^C + f^D - \begin{bmatrix} l_1 & l_2 & l_3 \\ l_4 & l_5 & l_6 \\ l_7 & l_8 & l_9 \end{bmatrix} c_{fr}(jh, (i+x_0)h),$$

$$i = 0, \dots, n, \quad j = 1, \dots, n. \quad (17)$$

c_{fr} is a function which is zero inside the domain Ω on which the system is solved and is equal to the border values on $\delta\Omega$, and h is the grid step on l_0 level.

The values b_k , $k = 1, 2, \dots, n(n+1)$ are obtained from a system whose matrix is also of the form (14), but in which:

$$C = \begin{bmatrix} l_5 & l_2 & 0 & \cdots & 0 \\ l_8 & l_5 & l_2 & \cdots & 0 \\ 0 & l_8 & l_5 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & l_5 \end{bmatrix},$$

$$D = \begin{bmatrix} l_6 & l_3 & 0 & \cdots & 0 \\ l_9 & l_6 & l_3 & \cdots & 0 \\ 0 & l_9 & l_6 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & l_6 \end{bmatrix}, \quad (18)$$

$$S = \begin{bmatrix} l_4 & l_1 & 0 & \cdots & 0 \\ l_7 & l_4 & l_1 & \cdots & 0 \\ 0 & l_7 & l_4 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & l_4 \end{bmatrix}.$$

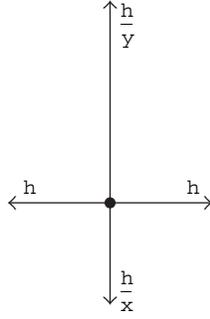


FIGURE 5

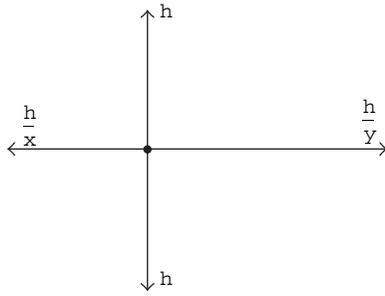


FIGURE 6

The components of the constant terms vector are now

$$\begin{aligned}
 z_{in+j} &= f_3^A + f_4^B + f_1^C + f_2^D \\
 &- \left[\begin{array}{ccc} l_1 & l_2 & l_3 \\ l_4 & l_5 & l_6 \\ l_7 & l_8 & l_9 \end{array} \right] c_{fr}((i+x_0)h, jh), \\
 & \quad i = 0, \dots, n, \quad j = 1, \dots, n.
 \end{aligned}
 \tag{19}$$

In the matrix A : $x = x_0$ and $y = x_0$ for the first line of blocks in (10), on Ω_A and Ω_B , $x = 1$ and $y = 1$ on Ω_C and Ω_D , while $x = 1$, $y = 1$ for the last line of blocks on Ω_A and Ω_B , and $x = 1 - x_0$, $y = 1 - x_0$ on Ω_C and Ω_D . For the remainder of the lines: $x = 1$, $y = 1$ (see Figure 5).

For the matrix B : $x = 1/x_0$ and $y = 1$ for the first line of blocks on Ω_A and Ω_D , $x = 1$ and $y = 1$ on Ω_B and Ω_C . The last line has: $x = 1$, $y = 1$ on Ω_A and Ω_D , and on Ω_B and Ω_C : $x = 1/(1 - x_0)$, $y = 1$. For the other lines: $x = 1$, $y = 1$ (see Figure 6).

The system obtained by discretizing the problem on every subdomain Ω_{iN+j} , $i = 0, \dots, n$, $j = 1, \dots, n$ has n_i^2 equations and unknowns and will be solved using the Gauss full elimination method. The solutions a and b computed above will be used as boundary conditions on this domain (Figure 7).

Reuniting the solutions computed on the grid corresponding to the level l_0 and the ones from every subdomain, one gets the final solution on the work level l .

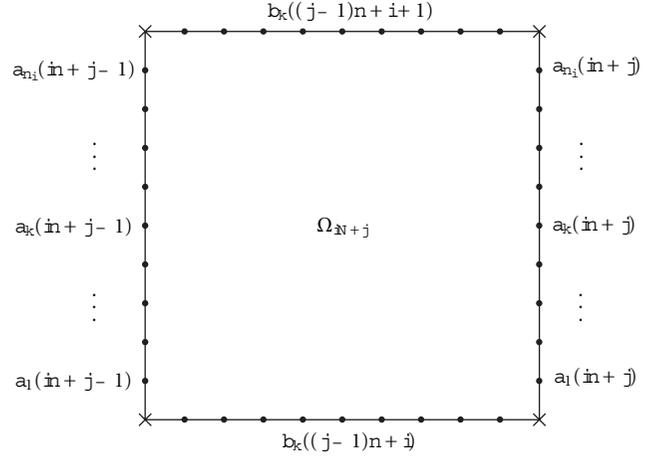
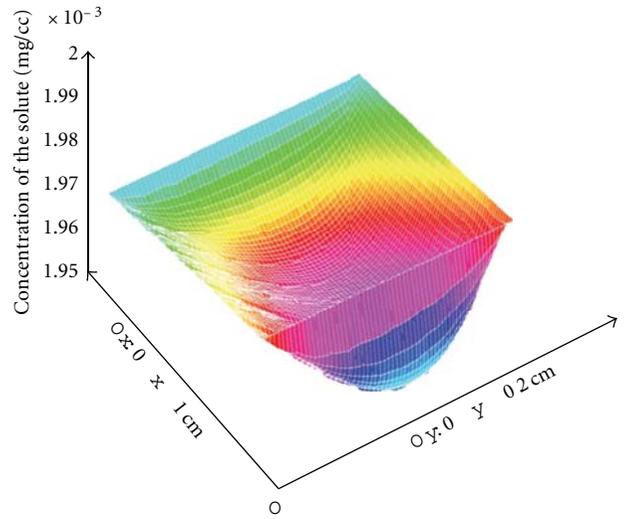


FIGURE 7


 FIGURE 8: The concentration computed on the domain $[0,1] \times [0,0.2]$.

5. Numerical Results and Conclusions

The method presented above allows the computing of the concentration of the solvent in blood at any point of the grid for different grid steps, and can be done for various dimensions of the vessel or values of the coefficients. The importance of such a method comes from the fact that the evaluation of cutaneous circulation can be a predictive parameter for the age-related pathology.

The values of the parameters used here are as follows:

- (i) the diffusion coefficient: $m = 10^{-3} \text{ cm}^2 \text{ s}^{-1}$;
- (ii) the velocity vector: $\mathbf{n} = (10, 10) \text{ cm s}^{-1}$;
- (iii) the wall permeability: $\xi = 1 \text{ cm s}^{-1}$;
- (iv) the concentration on Γ_{in} : $g_1 = 2 \cdot 10^{-3}$;
- (v) the concentration on Γ_{out} : $g_1 = 1.98 \cdot 10^{-3}$;
- (vi) the forcing term for the solute concentration $f = 0$;
- (vii) the vessel dimensions have been chosen to be $1 \text{ cm} \times 0.2 \text{ cm}$.

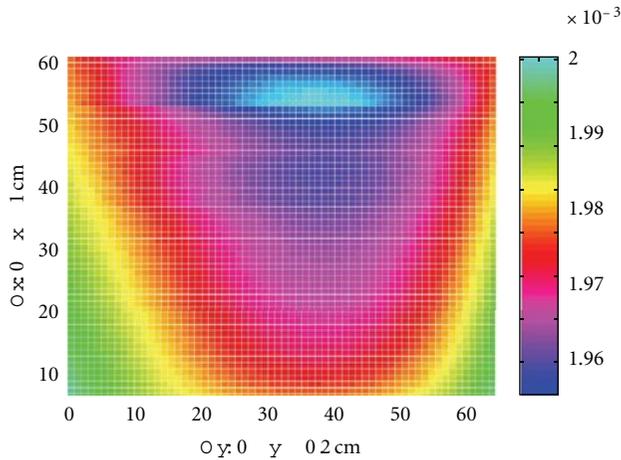


FIGURE 9: The concentration map on the domain $[0, 1] \times [0, 0.2]$.

For the convection-diffusion problem (1), the results are presented in Figures 8 and 9.

Acknowledgment

The paper is funded by research Grant CNCSIS IDEI 2624.

References

- [1] M. Crisan, C. Cattani, R. Badea et al., “Modelling cutaneous senescence process,” in *Proceedings of the International Conference on Computational Science and Its Applications, Part II (ICCSA '10)*, D. Taniar, O. Gervasi, B. Murgante, E. Pardede, and B. O. Apduhan, Eds., vol. 6017 of *Lecture Notes in Computer Science*, pp. 215–224, Springer, Berlin, Germany, March 2010.
- [2] M. Crisan, C. Cattani, R. Badea et al., “Complex histological, genetical, ultrasonography and clinical studies in early noninvasive diagnosis of the photoinduced cutaneous senescence and in the photoinduced skin cancers, using computerized imaging, modern biotechnology and mathematical modelling methods,” *Automatic Computers Applied Mathematics*, vol. 18, no. 2, pp. 231–255, 2009.
- [3] L. Formaggia, S. Perotto, and P. Zunino, “An anisotropic a-posteriori error estimate for a convection-diffusion problem,” *Computing and Visualization in Science*, vol. 4, no. 2, pp. 99–104, 2001.
- [4] E. B. Becker, G. F. Carey, and J. T. Oden, *Finite Elements. An Introduction*, Texas Institute for Computational Mechanics, UT Austin, 1981.
- [5] W. Hackbush, *Elliptic Differential Equations*, Springer, New York, NY, USA, 1992.
- [6] A. Brandt, “Multi-level adaptive solutions to boundary-value problems,” *Mathematics of Computation*, vol. 31, no. 138, pp. 333–390, 1977.
- [7] W. Hackbush, *Multigrid Method and Applications*, Springer, Berlin, Germany, 1985.
- [8] U. Trottenberg, C. W. Oosterlee, and A. Schüller, *Multigrid*, Elsevier Academic Press, London, UK, 2001.
- [9] G. Nuț, “A prolongation technique for solving partial differential equations with a multigrid method,” in *Proceedings of the Modelling and Development of Intelligent Systems Conference*, Sibiu, Romania, 2011.

Review Article

Functional Magnetic Resonance Imaging for Imaging Neural Activity in the Human Brain: The Annual Progress

Shengyong Chen¹ and Xiaoli Li²

¹ College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

² National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

Correspondence should be addressed to Shengyong Chen, sy@ieee.org

Received 29 September 2011; Accepted 25 October 2011

Academic Editor: Carlo Cattani

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

Functional magnetic resonance imaging (fMRI) is recently developed and applied to measure the hemodynamic response related to neural activity. The fMRI can not only noninvasively record brain signals without risks of ionising radiation inherent in other scanning methods, such as CT or PET scans, but also record signal from all regions of the brain, unlike EEG/MEG which are biased towards the cortical surface. This paper introduces the fundamental principles and summarizes the research progress of the last year for imaging neural activity in the human brain. Aims of functional analysis of neural activity from fMRI include biological findings, functional connectivity, vision and hearing research, emotional research, neurosurgical planning, pain management, and many others. Besides formulations and basic processing methods, models and strategies of processing technology are introduced, including general linear model, nonlinear model, generative model, spatial pattern analysis, statistical analysis, correlation analysis, and multimodal combination. This paper provides readers the most recent representative contributions in the area.

1. Introduction

Functional magnetic resonance imaging (functional MRI or fMRI) is based on the increase in blood flow to the local vasculature that accompanies neural activity in the brain. This results in a corresponding local reduction in deoxyhemoglobin because the increase in blood flow occurs without that similar magnitude in oxygen extraction. Deoxyhemoglobin is paramagnetic, and it alters the weighted MRI signal and thus is sometimes referred to as an endogenous contrast-enhancing agent. It also serves as the source of the signal for fMRI. Using an appropriate imaging sequence, human cortical functions can be identified without the use of exogenous contrast-enhancing agents on a clinical strength scanner. It has been confirmed that functional activity of the human brain from the MR signal is in anatomically distinct areas in the visual cortex, the motor cortex, and Broca's area of language-related activities. For example, Stroop test is commonly used as a behavior-testing tool for psychological examinations that are related to attention and cognitive control of the brain [1].

Over 100 years ago, it has been known that changes in blood flow and blood oxygenation (i.e., hemodynamics) are closely linked to neural activities in the brain. When neural cells are active, they increase the consumption of energy from glucose and switch to less energetically effective, but more rapid anaerobic glycolysis. The local response to this energy consumption is to increase blood flow to regions of increased neural activity, which occurs after a 1-2-second delay. The hemodynamic response rises to a peak over 4–6 seconds, before falling back to its baseline. This leads to changes in local cerebral blood volume and local changes in the concentration of oxyhemoglobin, which are detectable through the paramagnetic effects [2].

fMRI is highly interdisciplinary, and many studies are from several different fields, for example, physics (underlying fMRI signals and understanding of the principles), psychology (cognitive psychological, cognitive psychophysiological, and psychophysical experiments for obtaining extra measurements in addition to behavioral or electroencephalographic measurements), neuroanatomy (linking fMRI signals to understanding of the neuroanatomy), statistics (for

correct observations and avoiding false-positive results), and electrophysiology (neuronal behavior at the electrophysiological level) [3].

In early 1990s, it has been recognized the potential importance of blood-oxygen-level dependence (BOLD), which is the MRI contrast of blood deoxyhemoglobin, for functional brain imaging with MRI. The first successful fMRI study was reported in *Science* journal by Belliveau et al. in 1991 [4]. Now fMRI has come to dominate the brain mapping field due to its relatively low invasiveness, absence of radiation exposure, and relatively wide availability [3]. Further, rapidly emerging studies correspond findings between fMRI and conventional electrophysiological techniques to locate specific functions of the brain [5]. Consequently, the number of medical and research centers with fMRI capabilities and investigational programs continues to escalate [2]. Now BOLD-based fMRI becomes a powerful tool for studying brain function not only locally but also on the large scale [6]. The particular imaging methods and procedures vary from every individual institute. Yet there is no completely standardized package of software for clinical use.

Although the current fMRI uses BOLD as the method for determining active areas as the result of various experiences, the signals are relative and not individually quantitative. The recent fMRI technology extends traditional anatomical MR imaging from brain hemodynamics [7] or mental operations to brain functions [8]. fMRI provides the ability to observe both the structures and also which structures participate in specific functions. fMRI provides high-resolution, noninvasive observation of neural activity. This ability to directly observe brain function opens good opportunities to advance our understanding of brain organization. This paper briefly introduces the fundamental principles of fMRI and some recent directions [2].

Integration of electroencephalography (EEG) and fMRI has been pursued in an effort to achieve greater spatiotemporal resolution of imaging dynamic brain activity [9]. Recently, simultaneous EEG-fMRI measurements have been used to investigate the relation between the two signals. Previous attempts at the analysis of simultaneous EEG-fMRI data reported significant correlations between regional BOLD activations and modulation of both event-related potential (ERP) and oscillatory EEG power, mostly in the alpha but also in other frequency bands [10]. Functional MRI has high spatial resolution but relatively poor temporal resolution (in seconds). EEG directly measures the brain's electrical activity, giving high temporal resolution (in milliseconds) but low spatial resolution. The two techniques are therefore complementary and may be used simultaneously to record brain activity. Recording an EEG signal inside an MRI system is technically challenging. The MRI system introduces artifacts into the EEG recording by inducing currents in the EEG. This can happen through several different mechanisms. An imaging sequence applies a series of short radiofrequency pulses which induce a signal in the EEG system. The pulses are short and relatively infrequent, so interference may be avoided by blanking the EEG system during the transmission. The EEG system also affects the MRI scan. Metal in the EEG

leads and electrodes can introduce susceptibility artifacts into MR images. Care must be taken to limit currents induced in the EEG leads via the MRI system, which could heat the leads sufficiently to burn the subject [3].

Preliminary investigations of human brain mapping with these procedures have yielded insights into the functional organization of various sensory, motor, and language systems. In fact, BOLD effects are measured using rapid volumetric acquisition of images, with moderately good spatial and temporal resolution. Images are usually taken every several seconds, and the voxels in the resulting image typically represent cubes of tissue about several millimeters. Practically, the course of a BOLD response to a briefly presented stimulus lasts about 15 seconds for the robust positive response [3].

A typical procedure of clinical applications includes these steps. (1) *Image Acquisition*. Images are acquired using a weighted gradient echo sequence. The system is equipped with echo planar options for rapid image acquisitions. Slice thickness is usually set at 3–5 mm. Simultaneous images are acquired on as many as tens of contiguous slices. (2) *Image Processing*. Some processing programs are developed as a stand-alone system outside of the scanner system. They provide the computational capability to reconstruct the large numbers of images and statistical analyses that identify the active areas. (3) *Task Procedure*. Patients are positioned in the scanner as for a conventional scan. During a functional imaging series, tens of images are obtained. To identify brain tissue involved in language, sensory, visual, auditory, hand movement, and other targeted functions, the patient acts accordingly during the activity epoch. The beginning and end of this activity period are cued by a visual or auditory signal. (4) *Data Analysis*. Statistical analyses are often used to identify areas of the brain activated by specific tasks and are based on a multistage comparison of stimulation and resting intensity levels as well as multiple replications. Figure 1 shows a reconstructed fMRI study for neural activity analysis in our laboratory.

The scope of this paper is restricted to the most recent fMRI research for imaging neural activity in the human brain, mostly to introduce the research progress in 2010–2011. Although fMRI has attracted many researchers as early as since the 1980s, this paper concentrates on the most recent contributions. Furthermore, we include only what we believe to be representatives of important works and trends from recent years. The paper has four more sections. Section 2 introduces the relevant aims and applications of neural activity analysis by fMRI. Section 3 summarizes typical technologies in development. Section 4 is a discussion and Section 5 is the conclusion.

2. Neural Activity from fMRI

2.1. Functional Connectivity. Functional connectivity measures based upon low-frequency BOLD fMRI signal fluctuations have become a widely used tool for investigating spontaneous brain activity. However, the precise relationship between neural activity, the hemodynamic response and fluctuations in the MRI signal is still unknown. Recent works

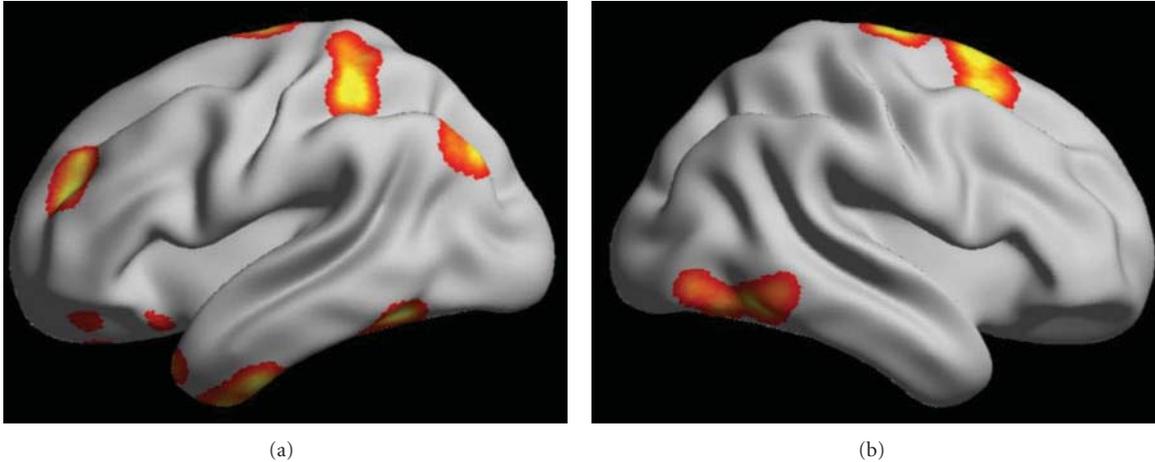


FIGURE 1: Reconstructed fMRI study for neural activity analysis.

had shown that correlated low-frequency fluctuations in the BOLD signal could be detected. Building on this preliminary work, Williams et al. demonstrate that functional connectivity observed in the rat depends strongly on the type of anesthesia used [11]. Lin et al. find that functional connectivity between cortical areas can be further revealed from the imaged source signals using phase synchrony measures [9]. Their approach is to image continuously oscillatory activities and their functional connectivity. Such ability promises to facilitate the investigation of the long-term neural behaviors and large-scale cortical interactions involved in spontaneous brain activity and cognition.

2.2. Biological Findings. Sparsity of the signal has been shown to be more promising in [12]. This coincides with biological findings such as sparse coding in the primary visual area (V1) simple cells, electrophysiological experiment results in the human medial temporal lobe, and so forth. A data-driven fMRI analysis is derived solely based upon the sparsity of the signals [12]. Comparative experiments have been done using canonical HRF, data-driven sparse GLM, sICA using Infomax, sICA using FastICA, and PCA (Figure 2). Khadka et al. attempted to find neural correlates between the performed cognitive tasks and hemodynamic signals detected by a diffuse optical tomography system [1]. The initial observation showed activation of oxyhemoglobin concentration in Brodmann area 10 (BA10), which is consistent with some results seen by positron emission tomography (PET) and fMRI.

Many studies assume a simple relationship between neural and BOLD activity, in spite of the fact that it is important to elucidate how the “when” and “what” components of neural activity are correlated to the “where” of fMRI data. Murayama et al. conducted simultaneous recordings of neural and BOLD signal fluctuations in V1 cortex of anesthetized monkeys. They explored the neurovascular relationship during periods of spontaneous activity. The results showed a positive neurovascular coupling with a lag of 4-5 seconds and a larger contribution from local

field potentials (LFPs) in the gamma range than from low-frequency LFPs or spiking activity. The method also detected a higher correlation around the recording site in the concurrent spatial map, even though the pattern covered most of the occipital part of V1 [6].

By integration of EEG source imaging and fMRI during continuous viewing of natural movies, it is found the most significant correlations in visual area V1. By calculating the impulse response function (IRF) between the BOLD signal and the estimated current density in area V1, it is found that the IRF is very similar to that observed using combined intracortical recordings and fMRI experiments in nonhuman primates. Taken together, these findings open an approach to noninvasive mapping of the brain. This is especially useful in combined EEG/fMRI experiments, where one can potentially study neural-hemodynamic relationships across the whole brain volume [13].

2.3. Vision and Hearing Research. In the community, fMRI has been applied for discovery of visual illusions, depth perceptions, hearing, or language-specific areas. Visual perceptual experiments have identified three types of neural pathways that represent color. It might be expected that there are neurons in the primary visual cortex that resemble the three perceptual pathways. Engel et al. used fMRI to examine responses in the human brain to a large number of colors. In visual cortical areas V1 and V2, the strongest response is to red-green stimuli, and much of this activity is from neurons receiving opposing inputs from L and M cones [14].

In [15], multivoxel pattern analysis is applied to investigate the specificity of brain activation patterns induced by acupuncture stimulations at a vision-related acupoint (GB37) and a nearby nonacupoint (NAP). Results showed that multiple areas could differentiate the neural response patterns induced by stimulation at the two sites with higher accuracy above the chance level [15]. Rauch et al. also studied effects of the local anesthetic Lidocaine on BOLD activity in V1 of nonhuman primates. Using independent component analysis (ICA), they describe and quantify the

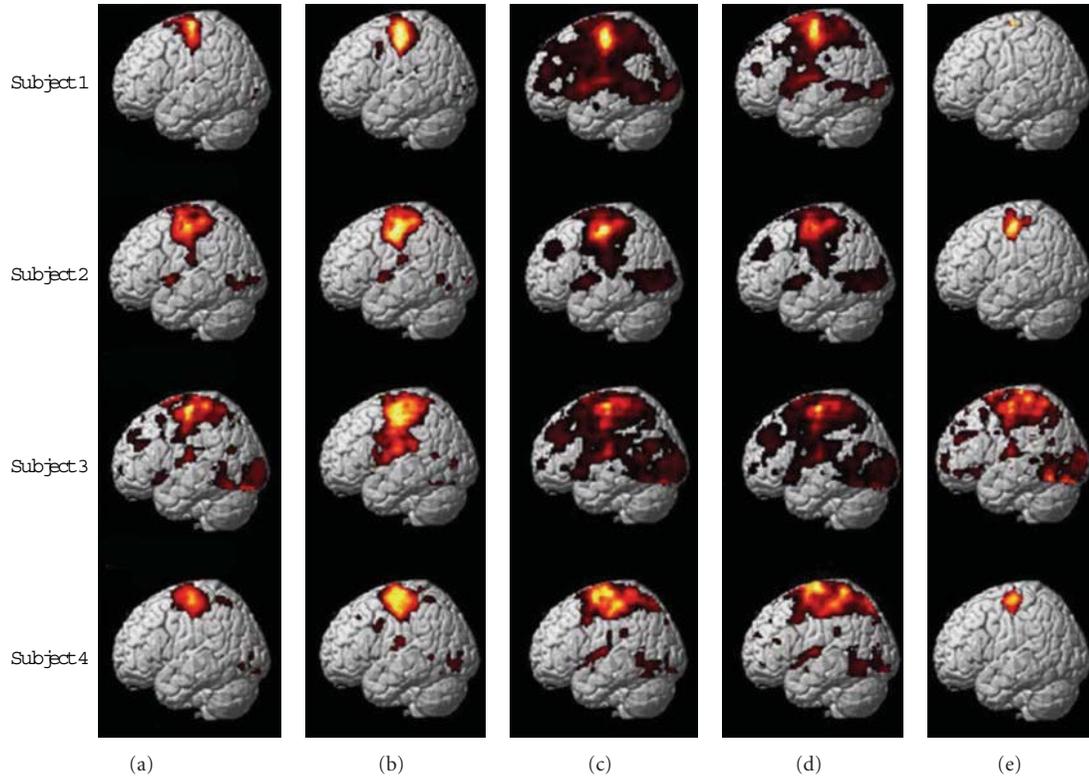


FIGURE 2: Comparative experiments using the design matrices constructed by (a) canonical HRF, (b) data-driven sparse GLM, (c) sICA using Infomax, (d) sICA using FastICA, and (e) PCA (Lee et al. 2011 [12] (2011 IEEE)).

pharmacodynamics and spatial distribution of Lidocaine effects on visually evoked V1 BOLD signal in a dose-dependent manner [16].

In [17], fMRI was used to investigate differences in neural activity between subjects that can modulate their tinnitus by jaw protrusion and normal hearing controls. Lanting et al. measured responses to bilateral sound and responses to jaw protrusion. The auditory system responded to both sound and jaw protrusion [17]. In the experiments by [18], participants identified speech sounds masked by varying noise levels while blood oxygenation signals were recorded with fMRI. Accuracy and response time were used to characterize the behavior of sensory and decision components of the perceptual system.

2.4. Emotional Research. Emotion plays a significant role in goal-directed behavior, but people yet know very little about its neural basis. In several psychological models the cardinal dimensions that characterize the emotion space are considered to be valence and arousal. In [19], 3T fMRI was used to reveal brain areas that show valence- and arousal-dependent BOLD signal responses. Influential theories of human emotion argue that subjective feeling states involve representation of bodily responses elicited by emotional events. Individual differences in intensity of emotional experience reflect variation in sensitivity to bodily responses. Critchley et al. measured regional brain activity by fMRI during an interoceptive task wherein subjects judged

the timing of their own heartbeats. They observed enhanced activity in insula, somatomotor, and cingulate cortices [20].

2.5. Neurosurgical Planning. Since neurosurgery relies on a precise delineation of the structural and functional aspects of brain, the role for fMRI in neurosurgical planning is very significant. The need for individualized maps of brain function is enhanced when the presence of a tumor alters the expected location of a function or when the location of the tumor is in an area with an uncertain function. fMRI does provide a source of precise functional and structural information for neurosurgery [2]. Many examples illustrated the potential advantage of functional and anatomical information for surgical treatment of brain tumors.

Recently, a cohort of neurosurgical patients are participants in a protocol at Columbia to evaluate the potential applications of fMRI for neurosurgical planning. These patients receive a standard battery of tasks targeted to localize language, sensory, motor, and visual areas both as candidates for surgery and as postsurgical patients. The objective of this investigation is to determine the potential role of functional mapping for neurosurgical procedures [2].

2.6. Pain Management. The experience of chronic and persistent pain is a debilitating condition for which the role of cortical processing is not well understood. The ability to use environmental stimuli to predict impending harm is critical for survival. Such predictions should be available as early

as they are reliable. Chains of successively earlier predictors are studied in terms of higher-order relationships and have inspired computational theories such as temporal difference learning. However, there is yet no adequate neurobiological account of how this learning occurs. In [8], by fMRI study of higher-order aversive conditioning, a computational strategy is described that humans use to learn predictions about pain.

People have focused on the identification of cortical areas that are modified by the reduction of pain following pain therapy. Recent studies indicate that the cortical representation of sympathetically maintained pain involves specific and identifiable cortical activity, as well as does the relief of that pain achieved by a peripheral nerve block procedure. The preliminary studies suggest a wide range of other approaches using fMRI to investigate cortical representations of specific pain types and, therefore, potential specific therapy options [2].

2.7. Others. With the ability to image the entire 3D volume of human brain, fMRI is capable of isolating many simultaneous and coordinated brain events. The multilevel view of brain activity can include executive functions and high-level cognitive tasks simultaneously with the inputs such as vision and audition. Methods can also be developed to identify brain structures involved with visual perception, language generation, comprehension of sequential information as in a video, the execution of visually guided responses, and problem solving. These aspects of brain function have not previously been scrutinized with such precision in neuroscience [2]. Many more applications can be explored. For example, Kannurpatti et al. studied neural and vascular variability and the fMRI-BOLD response in normal aging [21].

3. Processing Technology

3.1. Formulations and Basic Processing. Formulization and basic fMRI processing technologies are always useful to bring upgraded hardware and software. For example, fMRI is very sensitive to artifacts created by head motion and magnetic field deformation. It is thus necessary to attenuate these artifacts in order to obtain correct activation patterns. A model-based method is introduced in [22] to remove motion artifacts in short-duration movements. The algorithm can account for head movement and field deformations due to movements within and outside of the field of view [22]. On the other hand, the power of fMRI in assessing neural activities is hampered by intersubject variations in basal physiologic parameters, which may not be related to neural activation but has a modulatory effect on fMRI signals. Therefore, normalization of fMRI signals is useful in reducing variations and improving sensitivity [23].

The common gradient-echo echo-planar imaging technique in fMRI is sometimes hampered by macroscopic field inhomogeneities. This can affect the degree of signal change that occurs in the images as a response to neural activation and the subsequent blood oxygenation changes, that is, the BOLD sensitivity. In [24], quantitative sensitivity maps are calculated directly from gradient-echo field maps.

3.2. General Linear Model. To test for spatial heterogeneity, a direct statistical measure is proposed in [25] for the existence of distributed spatial patterns applicable to fMRI datasets. They extend the univariate general linear model (GLM) [12], typically used in fMRI analysis, to a multivariate case. Contrasting maximum likelihood estimations of different restrictions on this multivariate model can be used to estimate the extent of spatial heterogeneity. The test statistic is assessed using simulated time courses derived from real fMRI data followed by analyzing data from a real fMRI experiment [25].

3.3. Nonlinear Model. The signals and images acquired through this imaging technique demonstrate the brain's response to prescheduled tasks. Several studies on BOLD signal responses demonstrate nonlinear behavior for a stimulus. Taalimi and Fatemizadeh propose a mathematical approach for modeling BOLD signal activity, which is able to model nonlinear behaviors of physiological systems [26]. A nonlinear auto regressive moving average model is used to describe the mathematical relationship between output signals and predesigned tasks. Parameters can be used to distinguish between rest and active states of a brain region.

3.4. Generative Model. Generative models, such as the most typical ICA methods, can be used for the observed multivariate data in a large database of samples. ICA can separate a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals [16]. It is a special case of blind source separation and has been broadly applied to fMRI due to its capacity to separate spatially or temporally independent components. However, the assumption of independence has been challenged by recent studies, and, therefore, ICA does not guarantee independence of simultaneously occurring distinct activity patterns [12]. Missimer et al. compared two data-driven methods of statistical image analysis, principal component analysis (PCA) and ICA, in identifying neural networks related to the transient occurrence of phosphenes experienced by a patient subsequent to a brain infarct [27].

ICA of fMRI time series reveals distinct coactivation patterns in the resting brain representing spatially coherent spontaneous fluctuations of the fMRI signal. Among these patterns, the default-mode network has been attributed to the ongoing mental activity of the brain during wakeful resting state [28].

A data-driven approach is proposed in [9], which starts with using ICA to decompose the spatiotemporal EEG data into a linear combination of scalp potential maps and time courses. The time course of each independent component is used to construct a regressor to fit the fMRI time series. The resultant fMRI map then feeds back as a spatial constraint to the estimation of the source distribution underlying the corresponding component map. The estimated source distributions multiplied by the corresponding component time courses are summed across all components, giving rise to the reconstructed spatio-temporal activity [9].

3.5. Spatial Pattern Analysis. Much current works in fMRI employ multivariate machine-learning approaches (e.g., support vector machines, SVMs) to detect distributed spatial patterns from the temporal fluctuations of the neural signal. The aim is not classification, but investigation of multivariate spatial patterns, which pattern classifiers detect only indirectly [25]. These analyses demonstrate the utility of the measure of heterogeneity as well as considerations in its application. Measuring spatial heterogeneity in fMRI may have potential uses for better characterising neurological conditions such as stroke and Alzheimer's disease [25].

By combination of simultaneous EEG-fMRI imaging information and using multivariate machine-learning-based regression, De Martino et al. find that it is possible to predict EEG power oscillations from simultaneously acquired fMRI data during an eyes-open/eyes-closed task using either the original channels or the underlying cortically distributed sources as the relevant EEG signal for the analysis of multimodal data [10].

The cause of the detected brain activity relies on the anatomy. Diffusion tensor MR imaging as a noninvasive modality providing in vivo anatomical information allows determining neural fiber connections which leads to brain mapping. The main drawback of reliable fiber mapping is the correct detection of the orientation of multiple fibers within a single imaging voxel. Duru and Ozkan propose a method based on linear data structures to define the fiber paths regarding their diffusivity [29].

3.6. Statistical Analysis. Statistical methods are a valuable tool for decoding information from neural imaging data [30]. The noisy signal and the limited number of training patterns that are typically recorded from fMRI pose a challenge for the application of statistical learning methods in data analysis. For a typical fMRI scan, the 3D volume of the head is imaged every one or two seconds, producing a few hundred to a few thousand complete images per scanning session. Because of practical limitations of the scanner, small motions on the part of the subject and the subject's pulse and respiration will affect the images. After reconstruction, the scanning session consists of a series of 3D images. The most common tasks to perform on these images are corrections for motion and physiological effects. Outlier removal and spatial and temporal filtering may be performed. A variety of methods are used to correlate the voxel time series with the task to produce maps of task-dependent activation [3]. Multivariate statistical analysis is applied to compare multivariate data and establish the quantitative changes and differences between groups under investigation on their characteristics. PCA displays the original variables in a space, thus reducing the dimensionality of the data and allowing the visualization of a large number of variables.

Kangjoo et al. propose a statistical analysis method for fMRI to overcome the drawbacks of conventional data-driven ICA. A compressed sensing-based data-driven sparse GLM is proposed that enables estimation of spatially adaptive design matrix as well as sparse signal components

that represent synchronous, functionally organized and integrated neural hemodynamics. Furthermore, a minimum-description-length (MDL-) based model order selection rule is shown to be essential in selecting unknown sparsity level for sparse dictionary learning. Lee et al. propose a method that can adapt individual variation better than the conventional ICA methods do [12]. In fact, statistical parametric mapping is applied with a GLM expressed as

$$y_i = D \mathbf{x}_i + \varepsilon_i, \quad i \in [1, N], \quad (1)$$

where D denotes the regressors and \mathbf{x}_i denotes the corresponding response signal strength at the i th voxel. In the standard GLM model, the design matrix D is a predefined matrix. In [12], the data-driven sparse GLM assumes D as an unknown *global* dictionary, of which atom is assumed to indicate a principally dominant neural response in a small set of synchronous neural dynamics.

Danmei et al. propose using prior knowledge based on the behavioral performance of human observers to enhance the training of SVMs. They collect behavioral responses from human observers performing a categorization task during fMRI scanning and use the psychometric function generated based on the observers behavioral choices as a distance constraint for training an SVM. It is found that the behavior-constrained SVM outperforms SVM consistently [30], where the discrimination function is expressed as

$$f(\mathbf{x}) = \mathbf{w}\mathbf{x} + b. \quad (2)$$

Its parameters are optimized through minimizing a cost function. In nonlinear cases, the solution is reformulated as

$$f(\mathbf{x}) = \sum_{i \in SV} a_i y_i \mathcal{K}(\mathbf{x}_i, \mathbf{x}) + b, \quad (3)$$

where $\mathcal{K}(\mathbf{x}_i, \mathbf{x})$ is the kernel function.

The ultimate goal of fMRI data analysis is to detect correlations between brain activation and the task the subject performs during the scan. The BOLD signature of activation is relatively weak, however, so other sources of noise in the acquired data must be carefully controlled. A series of processing steps must be performed on the acquired images before the actual statistical search for task-related activation can begin [3].

3.7. Correlation Analysis. The precise relationship between neural signals and BOLD is an open problem. In general, changes in BOLD signal are well correlated with changes in blood flow. Numerous studies during the past decades have identified a coupling between blood flow and metabolic rate, that is, the blood supply is tightly regulated in space and time to provide the nutrients for brain metabolism [3].

Murayama et al. explored the neurovascular relationship during periods of spontaneous activity by using temporal kernel canonical correlation analysis, which is a multivariate method that can take into account any features in the signals that univariate analysis cannot. The method detects filters in voxel space and in frequency time space that maximize the neurovascular correlation without any assumption of a

hemodynamic response function. The results are consistent with those of previous studies and represent the multivariate analysis of intracranial electrophysiology and high-resolution fMRI [6].

To analyze EEG data in natural setting, Whittingstall et al. developed an ICA-based method to reject EEG artifacts due to blinks, subject movement, and so forth. They calculate the EEG source strength of the artifact-free data at each time point of the movie within the entire brain volume using low-resolution electromagnetic tomography. This provided for every voxel in the brain an estimate of the current density at every time point. A correlation is carried out between the time series of visual contrast changes with that of EEG voxels [13].

While current data indicate that local field potentials, an index of integrated electrical activity, form a marginally better correlation with blood flow than the spiking action potentials that are most directly associated with neural communication [15], no simple measure of electrical activity yet can provide an adequate correlation with metabolism and the blood supply across a wide dynamic range. Presumably, this reflects the complex nature of metabolic processes, which form a superset with regards to electrical activity. An initial small, negative dip before the main positive BOLD signal is more highly localized and also correlates with measured local decreases in tissue oxygen concentration. One problem with this technique is that the early negative BOLD signal is small and can only be seen using larger scanners with magnetic fields of at least 3 Tesla. Further, the signal is much smaller than the normal BOLD signal, making extraction of the signal from noise very difficult. Also, this initial dip occurs within 1-2 seconds of stimulus initiation, which may not be captured when signals are recorded at long repetition [3]. An event-related fMRI study was designed in [31] to dissociate the neural correlates of two putative key functions, volitional saccade generation and inhibition of reflexive saccades, and to investigate their interaction.

3.8. Multimodal Combination. Multimodal imaging techniques rely on the assumption of a common neuronal source for different recorded signals. In order to maximally exploit the combination of these techniques, we need to understand the coupling between EEG and fMRI BOLD signals. Beyond the correlation of the two measured brain signals, the ability of predicting the signal in one modality using information from the other modality is studied in [10]. Lin et al. report a data-driven approach to image spatio-temporal features of neural oscillatory activity and event-related activity from continuously recorded EEG and fMRI signals [9]. Lanting et al. studied multimodal integration of somatosensory jaw protrusion and sound [17]. Having simultaneously recorded EEG and fMRI data, the final hurdle is to coregister the two datasets, as each is reconstructed using a different algorithm, subject to different distortions [3].

By combining the neural and hemodynamic recordings in multiple modalities, we can get better insight into how and where the brain processes complex stimuli, which is especially useful for dealing with different neural diseases. However, due to different spatial and temporal resolutions,

the integration of EEG and fMRI recordings is not straightforward. One fundamental obstacle is that paradigms used for EEG experiments are usually event related, while fMRI is not limited in this regard. Furthermore, integration of the EEG signal directly with the BOLD signal is useful, by estimating the underlying IRF that relates the BOLD signal to the underlying current density [13].

4. Discussion

The main advantages of fMRI to image brain activity related to a specific action or sensory process include. (1) The signal does not require injections of radioactive isotopes. (2) The total scan time required can be very short, for example, in 1-2 minutes. (3) The in-plane resolution of the functional image is generally about $1 \times 1 \text{ mm}^2$. To put these advantages in perspective, functional images obtained by the earlier PET method require injections of radioactive isotopes, multiple acquisitions, and long imaging times. Further, the resolution of PET images is much larger than that in usual fMRI images. Additionally, PET usually requires that multiple individual images are combined in order to obtain a reliable signal. Consequently, information on a single patient is compromised and limited to a finite number of imaging sessions, and thus it is not optimally suitable to assist in a neurosurgical or treatment plan for specific individuals [2].

Many difficulties still exist in the fMRI technology. The images produced must be interpreted carefully, since correlation does not imply causality, and brain processes are complex and often nonlocalized. Statistical methods must be carefully used because they can produce false positives. The BOLD signal is an indirect measure of neural activity, which is susceptible to be influenced by nonneural changes in the body. It is difficult to interpret positive and negative BOLD responses. BOLD signals are most associated with the input to a given area rather than with the output. It is possible that a BOLD signal exists in an area even without a concrete activity there.

The BOLD response reaches to a peak in about 5 seconds after neuronal firing begins and fMRI has poor temporal resolution. It is, therefore, difficult to tell BOLD responses with different events occurring in a short period, although good experimental design can reduce this problem. Some researchers attempt to combine fMRI signals that have relatively high spatial resolution with signals recorded with other techniques, for example, EEG or magnetoencephalography (MEG), which have higher temporal resolution but worse spatial resolution. fMRI has often been used to show activation localized to specific regions, thus minimizing the distributed nature of processing in neural networks. The BOLD response can be affected by a variety of factors, for example, drugs, substances, age, brain pathology, local differences in neurovascular coupling, attention, amount of carbon dioxide in the blood, and so forth. For these reasons, functional imaging provides insights into neural processing that are complementary to insights of other studies in neurophysiology [3].

The aim of this paper is to introduce the most recent work of fMRI and potential future applications, for example,

neurosurgical planning and risk assessment, strategies for the treatment of chronic pain, seizure localization, and understanding of the physiology [2]. Although fMRI has been developed in many years as a relatively mature approach to estimation and diagnosis, problems still exist in its analysis in biomedical engineering. Researchers are exerting efforts in improving all simple and complex aspects.

5. Conclusion

Functional magnetic resonance imaging has been widely-used for detection of the brain's neural activity. This paper summarizes the annual progress for biomedical applications. Typical contributions are addressed for biological findings, functional connectivity, vision and hearing research, emotional research, neurosurgical planning, pain management, and so forth. Representative contributions are listed to show a general overview of new results in 2010-2011. Processing technology for solving fMRI problems is summarized. Particularly, introduced models and strategies include general linear model, nonlinear model, generative model, spatial pattern analysis, statistical analysis, correlation analysis, and multimodal combination.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC nos. 61025019, 61173096, 90820016, 60870002), Zhejiang Provincial S&T Department (2010R10006, 2010C33095), Zhejiang Provincial Natural Science Foundation of China (R1110679), and Natural Science Foundation of Hebei (F2009001638).

References

- [1] S. Khadka, S. R. Chityala, F. Tian, and H. Liu, "Identification of prefrontal cortex (BA10) activation while performing Stroop test using diffuse optical tomography," in *Photonic Therapeutics and Diagnostics VII*, vol. 7883 of *Proceedings of SPIE*, San Francisco, Calif, USA, January 2011.
- [2] Columbia University, "About Functional MRI(General)," 2011, <http://www.fmri.org/fmri.htm>.
- [3] Wikipedia, "Functional magnetic resonance imaging," 2011, http://en.wikipedia.org/wiki/Functional_magnetic_resonance_imaging.
- [4] J. W. Belliveau, D. N. Kennedy, R. C. McKinstry et al., "Functional mapping of the human visual cortex by magnetic resonance imaging," *Science*, vol. 254, no. 5032, pp. 716–719, 1991.
- [5] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," *Nature*, vol. 412, no. 6843, pp. 150–157, 2001.
- [6] Y. Murayama, F. Biessmann, F. C. Meinecke et al., "Relationship between neural and hemodynamic signals during spontaneous activity studied with temporal kernel CCA," *Magnetic Resonance Imaging*, vol. 28, no. 8, pp. 1095–1103, 2010.
- [7] G. Jaszdzewski, G. Strangman, J. Wagner, K. K. Kwong, R. A. Poldrack, and D. A. Boas, "Differences in the hemodynamic response to event-related motor and visual paradigms as measured by near-infrared spectroscopy," *NeuroImage*, vol. 20, no. 1, pp. 479–488, 2003.
- [8] B. Seymour, J. P. O'Doherty, P. Dayan et al., "Temporal difference models describe higher-order learning in humans," *Nature*, vol. 429, no. 6992, pp. 664–667, 2004.
- [9] L. Yang, Z. Liu., and B. He, "EEG-fMRI reciprocal functional neuroimaging," *Clinical Neurophysiology*, vol. 121, no. 8, pp. 1240–1250, 2010.
- [10] F. De Martino, G. Valente, A. W. de Borst et al., "Multimodal imaging: an evaluation of univariate and multivariate methods for simultaneous EEG/fMRI," *Magnetic Resonance Imaging*, vol. 28, no. 8, pp. 1104–1112, 2010.
- [11] K. A. Williams, M. Magnuson, W. Majeed et al., "Comparison of α -chloralose, medetomidine and isoflurane anesthesia for functional connectivity mapping in the rat," *Magnetic Resonance Imaging*, vol. 28, no. 7, pp. 995–1003, 2010.
- [12] K. Lee, S. Tak, and J. C. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1076–1089, 2011.
- [13] K. Whittingstall, A. Bartels, V. Singh, S. Kwon, and N. K. Logothetis, "Integration of EEG source imaging and fMRI during continuous viewing of natural movies," *Magnetic Resonance Imaging*, vol. 28, no. 8, pp. 1135–1142, 2010.
- [14] S. Engel, X. Zhang, and B. Wandell, "Colour tuning in human visual cortex measured with functional magnetic resonance imaging," *Nature*, vol. 388, no. 6637, pp. 68–71, 1997.
- [15] L. Linling, Q. Wei, B. Lijun, and T. Jie, "Exploring vision-related acupuncture point specificity with multivoxel pattern analysis," *Magnetic Resonance Imaging*, vol. 28, no. 3, pp. 380–387, 2010.
- [16] A. Rauch, G. Rainer, M. Augath, A. Oeltermann, and N. K. Logothetis, "Pharmacological MRI combined with electrophysiology in non-human primates: effects of Lidocaine on primary visual cortex," *NeuroImage*, vol. 40, no. 2, pp. 590–600, 2008.
- [17] C. P. Lanting, E. de Kleine, R. N. Eppinga, and P. van Dijk, "Neural correlates of human somatosensory integration in tinnitus," *Hearing Research*, vol. 267, no. 1-2, pp. 78–88, 2010.
- [18] J. R. Binder, E. Liebenthal, E. T. Possing, D. A. Medler, and B. D. Ward, "Neural correlates of sensory and decision processes in auditory object identification," *Nature Neuroscience*, vol. 7, no. 3, pp. 295–301, 2004.
- [19] M. Viinikainen, I. P. Jääskeläinen, Y. Alexandrov, M. H. Balk, T. Autti, and M. Sams, "Nonlinear relationship between emotional valence and brain activity: evidence of separate negative and positive valence dimensions," *Human Brain Mapping*, vol. 31, no. 7, pp. 1030–1040, 2010.
- [20] H. D. Critchley, S. Wiens, P. Rotshtein, A. Öhman, and R. J. Dolan, "Neural systems supporting interoceptive awareness," *Nature Neuroscience*, vol. 7, no. 2, pp. 189–195, 2004.
- [21] S. S. Kannurpatti, M. A. Motes, B. Rypma, and B. B. Biswal, "Neural and vascular variability and the fMRI-BOLD response in normal aging," *Magnetic Resonance Imaging*, vol. 28, no. 4, pp. 466–476, 2010.
- [22] T. Lemmin, G. Ganesh, R. Gassert, E. Burdet, M. Kawato, and M. Haruno, "Model-based attenuation of movement artifacts in fMRI," *Journal of Neuroscience Methods*, vol. 192, no. 1, pp. 58–69, 2010.
- [23] H. Lu, U. Yezhuvath, and G. Xiao, "Improving fMRI sensitivity by normalization of basal physiologic state," *Human Brain Mapping*, vol. 31, no. 1, pp. 80–87, 2010.
- [24] P. Mannfolk, R. Wirestam, M. Nilsson, D. van Westen, F. Ståhlberg, and J. Olsrud, "Assessment of spatial BOLD

- sensitivity variations in fMRI using gradient-echo field maps,” *Magnetic Resonance Imaging*, vol. 28, no. 7, pp. 947–956, 2010.
- [25] R. Leech and D. Leech, “Testing for spatial heterogeneity in functional MRI using the multivariate general linear model,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 6, pp. 1293–1302, 2011.
- [26] A. Taalimi and E. Fatemizadeh, “A new mathematical approach for detection of active area in human brain fMRI using nonlinear model,” *Biomedical Engineering—Applications, Basis and Communications*, vol. 22, no. 5, pp. 409–418, 2010.
- [27] J. H. Missimer, R. J. Seitz, and R. Kleiser, “Data-driven analyses of an fMRI study of a subject experiencing phosphenes,” *Journal of Magnetic Resonance Imaging*, vol. 31, no. 4, pp. 821–828, 2010.
- [28] T. Meindl, S. Teipel, R. Elmouden et al., “Test-retest reproducibility of the default-mode network in healthy individuals,” *Human Brain Mapping*, vol. 31, no. 2, pp. 237–246, 2010.
- [29] D. Duru and M. Ozkan, “Determination of neural fiber connections based on data structure algorithm,” *Computational Intelligence and Neuroscience*, vol. 2010, Article ID 251928, 6 pages, 2010.
- [30] C. Danmei, L. Sheng, Z. Kourtzi, and W. Si, “Behavior-constrained support vector machines for fMRI data analysis,” *IEEE Transactions on Neural Networks*, vol. 21, no. 10, pp. 1680–1685, 2010.
- [31] B. Reuter, C. Kaufmann, J. Bender, T. Pinkpank, and N. Kathmann, “Distinct neural correlates for volitional generation and inhibition of saccades,” *Journal of Cognitive Neuroscience*, vol. 22, no. 4, pp. 728–738, 2010.

Research Article

Abdominal Tumor Characterization and Recognition Using Superior-Order Cooccurrence Matrices, Based on Ultrasound Images

Delia Mitrea,¹ Paulina Mitrea,¹ Sergiu Nedevschi,¹ Radu Badea,² Monica Lupsor,² Mihai Socaciu,² Adela Golea,² Claudia Hagiu,² and Lidia Ciobanu²

¹ Department of Computer Science, Technical University of Cluj-Napoca, George Baritiu Street 26–28, 400027 Cluj-Napoca, Romania

² Department of Ultrasonography, Iuliu Hatieganu University of Medicine and Pharmacy, Victor Babeş Street 8, 400079 Cluj-Napoca, Romania

Correspondence should be addressed to Delia Mitrea, delia.mitrea@cs.utcluj.ro

Received 2 September 2011; Accepted 18 September 2011

Academic Editor: Maria Crisan

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

The noninvasive diagnosis of the malignant tumors is an important issue in research nowadays. Our purpose is to elaborate computerized, texture-based methods for performing computer-aided characterization and automatic diagnosis of these tumors, using only the information from ultrasound images. In this paper, we considered some of the most frequent abdominal malignant tumors: the hepatocellular carcinoma and the colonic tumors. We compared these structures with the benign tumors and with other visually similar diseases. Besides the textural features that proved in our previous research to be useful in the characterization and recognition of the malignant tumors, we improved our method by using the grey level cooccurrence matrix and the edge orientation cooccurrence matrix of superior order. As resulted from our experiments, the new textural features increased the malignant tumor classification performance, also revealing visual and physical properties of these structures that emphasized the complex, chaotic structure of the corresponding tissue.

1. Introduction

The hepatocellular carcinoma (HCC) is the most frequent malignant liver tumor, representing 75% of the liver cancer cases [1]. The colorectal tumors also represent a frequent disease for the population of the developed countries. The human observations are not enough in order to perform the detection of the malignant tumors, the resulted diagnosis accuracy being below 80%. The golden standard for cancer diagnosis is the biopsy, but this is an invasive, dangerous method that can lead to the spread of the tumor inside the human body. A non-invasive, subtle analysis is due, in order to detect the cancer in early evolution stages, when the tumor can be surgically removed. We perform this study by using computerized methods applied on ultrasound images. Other types of image acquisition techniques, such as computer tomography (CT), magnetic resonance imaging (MRI), and endoscopy are considered invasive or expensive.

The texture is an important feature, as it provides subtle information concerning the pathological state of the tissue, overcoming the accuracy of the human perception, through the statistical and multiresolution approaches. The texture-based methods in combination with classifiers were widely used in the domain of malignant tumor characterization and recognition from medical images. In [2], Raeth used the textural features in order to distinguish the normal liver from the diffuse liver diseases and from the malignant liver tumors. The features derived from the second-order grey levels cooccurrence matrix, from the edge cooccurrence matrix, as well as other edge and gradient-based features, speckle noise distribution parameters, and the Fourier power spectrum, provided satisfying results concerning the differentiation between the tumoral and nontumoral tissue. In [3] the authors computed the first-order statistics (the mean grey level and the grey level variance), the second-order grey level cooccurrence matrix parameters and run-length matrix

parameters which were used in combination with an artificial neural networks based classifier, as well as with a classifier based on linear discriminants in order to differentiate the malignant liver tumors from hemangioma and from the normal liver. The resulted recognition rate was 79.6%. The wavelet transform was also implemented [4], in order to perform a multi-resolution analysis of the textural features. The method provided satisfying results concerning the differentiation between malignant and benign liver lesions, the area under the ROC (receiver operating characteristic) curve being approximately 90%. In [5] the authors analyzed the fluorescent images of the colonic tissue based on textural parameters derived from the second order grey level cooccurrence matrix (GLCM), in order to distinguish the colonic healthy mucosa versus adenocarcinoma. However, a systematic study concerning the most relevant textural features that best characterize the malignant tumors and of the most appropriate methods that lead to an increased diagnosis accuracy is not done. We perform this in our work by building the imagistic textural model of the malignant tumors. We previously defined the imagistic textural model of the malignant tumors [6], consisting in the most relevant textural features able to separate the HCC tumor from the visually similar tissues (cirrhotic parenchyma, benign tumors), together with their specific values (mean, standard deviation, and probability distribution). In this work, we analyzed new methods for textural features computation, based on the superior order grey level cooccurrence matrix (GLCM) [7], respectively on the superior order edge orientation cooccurrence matrix (EOCM), the purpose being to improve the characterization of the abdominal malignant tumors, and to increase the automatic diagnosis accuracy. In this way, we expect to get a more subtle evaluation procedure than in the case of using the other textural features. The superior order GLCM was theoretically described by Akono in [7]. The third-order GLCM was experimented for the analysis of the trabecular bones in proximal femur radiographs [8], as well as for crop classification [9], but it was never implemented for tumor characterization and recognition. There are no important realizations in the image analysis domain involving the fifth-order GLCM matrix. The second order EOCM was implemented by Raeth in [2] for malignant tumor contour characterization and provided satisfying results in this domain. The third order EOCM was not previously implemented. Thus, we analyzed the role that the second-, third-, and fifth-order GLCM, respectively, the second- and third-order EOCM have, concerning both the subtle characterization of HCC and colonic tumor tissue, as well as the automatic diagnosis of these types of cancer. Extended Haralick features were defined for the characterization of the tumor texture, and the best orientations of the corresponding displacement vectors were determined in both cases of the superior order GLCM and EOCM. The edge orientation variability feature was also defined in order to characterize the complex structure of the tumor tissue. The malignant tumors were compared with visually similar tissues. The HCC tumor was compared with the cirrhotic liver parenchyma on which it had evolved and with the benign liver tumors. The colonic tumors were compared

with the inflammatory bowel diseases (IBD), as they share, in ultrasound images, many visual characteristics with these affections. The assessment of the relevant textural features for the characterization of the malignant tumors was also performed, through specific methods such as the correlation-based feature selection (CFS) [10] and through the evaluation of the individual attributes based on their information gain with respect to the class [10]. Powerful classifiers that gave the best results in our former experiments [6], such as the multilayer perceptron [11] and the support vector machines (SVM) [11], as well as the AdaBoost combination scheme [11], were adopted for the evaluation of the textural model and of the recognition accuracy. The correlation of the textural features with the internal structure and with the properties of the tumor tissue was also discussed.

2. Materials and Methods

2.1. Materials and Working Methodology. In our study, mainly the patients suffering from HCC and colonic tumors were taken into consideration. Patients affected by benign liver tumors such as hemangioma and focal nodular hyperplasia (FNH) were also considered, being known that these tumors have a similar visual aspect with HCC in many situations. Subjects suffering from inflammatory bowel diseases (IBD) were taken into account as well, because these affections provided a similar visual aspect of the bowel walls like those provided by the colorectal tumors. All these patients were previously biopsied. For each patient, multiple images were acquired, corresponding to various orientations of the transducer, using the same settings of the ultrasound machine. The same number of images was considered for each patient, as described in the experimental section. Thus, the study was independent from the patient's characteristics. B-mode ultrasonography was used, in order to preserve the textural properties of the tissues. Rectangular regions of interest were selected inside the tumors, on the liver tissue, or on the bowel wall, in areas which were not affected by artifacts. Then, the imagistic textural model of the malignant tumors was built according to the steps below, and the role of the new derived textural features in improving the accuracy of the malignant tumor characterization and recognition performance was analyzed.

2.2. The Imagistic Textural Model of the Malignant Tumors

2.2.1. The Imagistic Textural Model of the Malignant Tumors and the Phases Due for Model Building. The imagistic textural model of HCC consists of the set of relevant, independent textural features, able to distinguish this tumor from the cirrhotic liver parenchyma and from the benign tumors. The specific, statistical values of the textural features—mean, standard deviation, and probability distribution—are part of the model. The mathematical description of the imagistic textural model is given below. Let F be the space of the potentially relevant textural features, containing a number of n such features:

$$F = \{f_i\}_{i=1,\dots,n}. \quad (1)$$

The features from F are considered in their initial representation, as they appear after applying the image analysis methods. We define

$$F_R = \text{Dimensionality_reduction}(F) \quad (2)$$

as being the transformed feature space, obtained from the initial feature space, F , after applying dimensionality reduction methods—mainly feature selection techniques [10]. The imagistic textural model of the tumor (TM) consists of a collection of vectors V_{f_r} , associated with each relevant textural feature f_r , containing the specific values that characterize each analyzed class:

$$\begin{aligned} \text{TM} &= \{V_{f_r} \mid V_{f_r} \\ &= [\text{Relevance, Mean, Standard_deviation,} \\ &\quad \text{Probability_distribution}]\}. \end{aligned} \quad (3)$$

The vectors of the imagistic textural model are composed by the specific parameters described by (3), where mean (the arithmetic mean value) and standard deviation are real numbers; the Relevance, represented by an integer, quantifies the importance that the considered textural feature has in the differentiation between HCC and other kinds of tissues.

In order to generate a reliable imagistic textural model, first, the image selection for the training set building is due. For each considered type of tissue, a corresponding class is built. Then, an image analysis phase is necessary: the textural feature computation using specific methods for texture analysis is involved in this process. The values of the textural features are stored in the database and used for further evaluations. The learning phase is essential in order to perform the relevant feature selection, to eliminate the redundant features and to determine the specific, statistical values, and the corresponding probability distributions. Dimensionality reduction methods consisting of feature selection [10] and feature extraction techniques [11] are implemented in this phase. At the end, a validation phase is necessary, involving the evaluation of the generated model by providing the relevant features at the classifiers inputs and estimating the accuracy of each classifier. A new test set of images, different from the training set, is used in this phase. The phases due in order to build the imagistic textural model are described below.

2.2.2. Training Set Building. For each patient, three to five images were considered. On each image, rectangular regions of interest were selected on each type of tissue, inside HCC and the colonic tumors, respectively, on the cirrhotic parenchyma on which HCC evolved, as well as inside the benign liver tumors and on the superior part of the bowel wall affected by inflammatory bowel diseases. Pairs of classes were considered, and then the classes were combined in equal proportions inside the training set. The potentially relevant textural features were determined on the regions of interest, using specific methods for texture analysis, and the

corresponding values were stored. An instance of the training set consisted of the values of the considered textural features, computed inside a certain region of interest, followed by the class specification.

2.2.3. Methods Applied during the Image Analysis Phase. During the image analysis phase, noise reduction was initially performed, by using an averaging filter [12]. Then, specific methods for texture analysis were applied, providing the initial set of potentially relevant textural features. We previously computed 48 textural features, from the following categories: the mean value of the grey levels [12], the second order grey levels cooccurrence matrix (GLCM), and the associated Haralick parameters [13]—the energy, entropy, correlation, contrast, variance, and local homogeneity that emphasized the global properties of the texture. Edge and gradient-based statistics [12], respectively, the frequency and density of the textural microstructures, detected by using the Laws convolution filters were computed as well [12]. The Shannon entropy [14], computed after applying the wavelet transform [15], was also determined. The Haar wavelet transform was applied recursively at two levels of resolution: the low-low, low-high, high-low, and high-high components were derived at the first level, then, the wavelet transform was applied again on each of these components. The Shannon entropy was computed on each resulted component, at both first and second levels. The determined textural features were independent on orientation, as they were computed on multiple directions and the result was averaged. They were also independent of illumination and scaled with the size of the region of interest. In this work, we defined and experimented the third-and fifth order GLCM, respectively, the second-and third order EOCM, for obtaining more refined textural features. The effect of the new textural features on the improvement of the imagistic textural model of the malignant tumors was carefully analyzed.

2.2.4. Description of the Learning Phase. During the learning phase, the selection of the relevant textural features was performed. We considered a feature as being relevant if it emphasized the defining characteristics of the tumor tissue and it substantially contributed to the separation of the tumor tissue from the visually similar tissues. From a more technical point of view, a feature was considered relevant if, by including it in the feature set, it led to an increase in the classification accuracy. There are specific methods for feature selection, integrated in two main groups, filters and wrappers [10], which perform a reliable separation of the relevant features from the nonrelevant ones. We compared, in our previous research [6], various methods from these categories, as well as their combinations. The best results were obtained when using the methods of correlation-based feature selection (CFS), combined with genetic search [10], the information gain attribute evaluation [16], the consistency-based feature subset evaluation [10], respectively the wrapper that used the decision trees as classifier, and the best first search method [16] for subset finding. The specific

values of the relevant textural features were determined by using confidence intervals and probability distribution tables [11]. In this work, we assessed the relevance of the newly obtained textural features, by using the most powerful feature selection methods, being interested in the diagnosis accuracy improvement.

2.2.5. Description of the Validation Phase. The validation phase consisted of providing the final set of relevant textural features at the inputs of some powerful classifiers, and in analyzing their effect on the classification process improvement. Classifiers from different categories, as well as classifier combinations, were compared in order to obtain the best performance during this phase [6]. The best results were provided by the methods of support vector machines (SVM) [11] with polynomial kernel of 3rd degree, by the multilayer perceptron (MLP), decision trees (C4.5 method), respectively, by the AdaBoost combination scheme. The following parameters were used in order to assess the classification performance: the recognition rate (percent of correctly classified instances), the sensitivity (TP rate), the specificity (TN rate), the area under the ROC curve (AUC) [11], and the time due for model building [16]. The stratified cross-validation strategy [11] was implemented for classification performance evaluation, in order to preserve the original class proportions.

2.3. The Newly Defined Textural Features and Their Role in the Improvement of Imagistic Textural Model

2.3.1. The Description of the New Texture Analysis Methods

(1) The Grey Level Cooccurrence Matrix of Superior Order. The grey level cooccurrence matrix (GLCM), also called the Grey Tone Difference Matrix, was previously defined by Julesz et al. [17] and Haralick [18]. Julesz et al. [17] was the first who used grey tone spatial dependence cooccurrence statistics in his texture discrimination experiments. Haralick [18] defined the two-dimensional cooccurrence matrix of the grey levels as containing, in its elements, the number of pairs of pixels having two specific values of the intensity, g_1 and g_2 , being situated at a distance defined by a displacement vector:

$$\vec{d} = (\vec{dx}, \vec{dy}). \quad (4)$$

Haralick also defined and implemented statistical measures, such as the homogeneity, energy, entropy, correlation, variance, contrast [18], in order to emphasize the global properties of the texture. In [7], Akono et al. described the GLCM of order n and proposed a fast computation algorithm for this method, but did not state a corresponding definition. He also extended the mathematical expressions of several statistical (Haralick) measures from order two to order n , such as the sum of the GLCM elements, the inverse difference, the dissimilarity and the contrast.

We defined the GLCM of order n in the following manner:

$$\begin{aligned} C_D(g_1, g_2, g_3, \dots, g_n) &= \#\{((x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)) : \\ &f(x_1, y_1) = g_1, f(x_2, y_2) = g_2, \dots, \\ &f(x_n, y_n) = g_n, |x_2 - x_1| = |\vec{dx}_1|, \\ &|x_3 - x_1| = |\vec{dx}_2|, \dots, |x_n - x_1| = |\vec{dx}_{n-1}|, \\ &|y_2 - y_1| = |\vec{dy}_1|, |y_3 - y_1| = |\vec{dy}_2|, \dots, \\ &|y_n - y_1| = |\vec{dy}_{n-1}|, \\ &\text{sgn}((x_2 - x_1)(y_2 - y_1)) = \text{sgn}(\vec{dx}_1 \cdot \vec{dy}_1), \dots, \\ &\text{sgn}((x_n - x_1)(y_n - y_1)) = \text{sgn}(\vec{dx}_{n-1} \cdot \vec{dy}_{n-1})\}. \end{aligned} \quad (5)$$

In (5), $\#S$ is the number of the elements in the set S , while

$$\vec{d} = ((\vec{dx}_1, \vec{dy}_1), (\vec{dx}_2, \vec{dy}_2), \dots, (\vec{dx}_{n-1}, \vec{dy}_{n-1})) \quad (6)$$

is the set of the displacement vectors. Thus, the GLCM matrix of order n contains in its elements the number of n -tuples of pixels with the spatial coordinates (x_i, y_i) , $i \in \{1, \dots, n\}$, having the intensity values g_i , $i \in \{1, \dots, n\}$, and being in a spatial relation defined by the displacement vectors described in (6). In practice, we used the GLCM probability matrix:

$$p(g_1, g_2, \dots, g_n) = \frac{C_D(g_1, g_2, \dots, g_n)}{\sum_{g_1=0}^{N_g-1} \sum_{g_2=0}^{N_g-1} \dots \sum_{g_n=0}^{N_g-1} C_D(g_1, g_2, \dots, g_n)}. \quad (7)$$

In (7), N_g is the total number of the gray levels in the image. Based on the n th order GLCM, we computed the following parameters: energy, entropy, local homogeneity, correlation, contrast, variance, as described in Appendix B. The maximum probability for a certain combination of grey levels to appear within the texture is also computed, as indicated in Appendix B, while searching for a specific pattern of grey levels within each type of analyzed tissue. The second order GLCM was determined for the following directions of the displacement vectors: 0° , 45° , 90° , and 135° . The corresponding Haralick features were averaged for all the resulted matrices.

The Implementation of the Third Order GLCM. For the third order GLCM, we considered specific orientations of the displacement vectors. The corresponding three pixels were either collinear, or they formed a right angle triangle (as shown in Figure 1), the current pixel, of coordinates (x_1, y_1) , being situated in the central position. Thus, in the case of the collinearity of the pixels, the direction pairs were $(0^\circ$,

180°), (90°, 270°), (45°, 225°), (135°, 315°), while in the second case, the following direction pairs were considered: (0°, 90°), (90°, 180°), (180°, 270°), (0°, 270°), (45°, 135°), (135°, 225°), (225°, 315°), and (45°, 315°). The values of $|\vec{d}x_i|$ and $|\vec{d}y_i|$ were 0 or 2, with $i \in \{1, 2\}$.

The Implementation of the Fifth Order GLCM. For the fifth order GLCM, the following groups of directions were taken into account: (0°, 180°, 90°, 270°), respectively, (45°, 225°, 135°, 315°). The current pixel, of coordinates (x_1, y_1) , was situated in the central position. The values of $|\vec{d}x_i|$ and $|\vec{d}y_i|$ were 0 or 2, with $i \in \{1, 2, 3, 4\}$.

(2) *The Cooccurrence Matrix of Edge Orientations.* The generalized cooccurrence matrix (GCM), defined by Davis and Jones in [19], represents the natural extension of the gray level cooccurrence matrix (GLCM), by taking into consideration, instead of the grey levels of the pixels, local features such as edges (points of increased gradient value) or edge orientations detected in the image through specific methods [12].

The edge orientation cooccurrence matrix (EOCM) of order two was defined by Davis and Jones [19] and implemented by Raeth [2] in order to analyze the contour shape of the malignant tumors. We consider the following definition for the edge orientation cooccurrence matrix of order n :

$$\begin{aligned}
& C_D(o_1, o_2, o_3, \dots, o_n) \\
& = \#\left\{((x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)) : \right. \\
& \quad \text{edge_ori}(x_1, y_1) = o_1, \text{edge_ori}(x_2, y_2) = o_2, \dots, \\
& \quad \text{edge_ori}(x_n, y_n) = o_n, |x_2 - x_1| = |\vec{d}x_1|, \\
& \quad |x_3 - x_1| = |\vec{d}x_2|, \dots, |x_n - x_1| = |\vec{d}x_{n-1}|, \\
& \quad |y_2 - y_1| = |\vec{d}y_1|, \\
& \quad |y_3 - y_1| = |\vec{d}y_2|, \dots, |y_n - y_1| = |\vec{d}y_{n-1}|, \\
& \quad \text{sgn}((x_2 - x_1)(y_2 - y_1)) = \text{sgn}(\vec{d}x_1 \cdot \vec{d}y_1), \dots, \\
& \quad \left. \text{sgn}((x_n - x_1)(y_n - y_1)) = \text{sgn}(\vec{d}x_{n-1} \cdot \vec{d}y_{n-1})\right\}. \tag{8}
\end{aligned}$$

Thus, each element of this matrix is equal with the number of n -tuples of pixels with spatial coordinates (x_i, y_i) , $i \in \{1, \dots, n\}$, the values of the edge orientations in these points being o_i , $i \in \{1, \dots, n\}$. The spatial relation between the pixels is defined by the set of the displacement vectors, in a similar way with the case of the superior order GLCM. In practice, the EOCM probability matrix was used, being defined in a similar way with the GLCM probability matrix. The edge orientation parameter was computed in each edge point (point of nonzero gradient) by applying the arctangent function on the fraction $G_y(x, y)/G_x(x, y)$, G_y being the

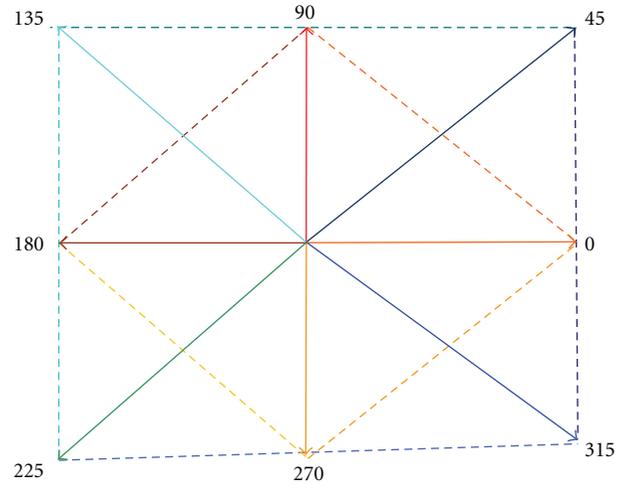


FIGURE 1: The main directions for the displacement vectors and their combinations in the case of third order GLCM.

vertical image gradient in the point (x, y) while G_x was the horizontal gradient in the same point. These gradient values were determined by using the Sobel convolution kernel for horizontal and vertical directions [12]. The extended Haralick features—contrast, variance, correlation, energy, and entropy—were defined in the same way as those corresponding to the superior order GLCM, detailed in Appendix B. The second- and third order EOCM were considered in our analysis. The directions of the displacement vectors and their combinations were chosen in a similar way with those corresponding to the case of second- and third order GLCM. The maximum probability parameters were also determined.

2.3.2. The Role of the New Textural Features in Improving the Textural Model of the Malignant Tumors. During the Image Analysis Phase, the old textural features were first computed. Then, the newly defined texture analysis methods were applied, and the corresponding textural features were determined in different conditions, by varying the displacement vector directions and the combinations of these directions. The groups of features corresponding to different values of the intrinsic parameters were assessed separately by powerful classifiers. The discrimination ability of the newly defined textural features was assessed as well, by feature selection methods and appropriate classifiers.

Thus, during the learning phase, feature selection methods were applied in order to estimate the relevance of the textural attributes. First, the new textural features were evaluated individually, by considering only the group of the new textural attributes, obtained for various instances of the intrinsic parameters. Then, the new textural features, corresponding to the most successful configuration of the intrinsic parameters, were considered in combination with the old textural features, in order to assess the increase in accuracy and to derive the final set of relevant textural features by applying the feature selection methods.

In this work, the selection of the relevant textural features was implemented by using the correlation-based feature

selection (CFS) method, in combination with genetic search [10], for retaining those textural features that were mostly correlated with the class parameter, and less correlated with the other textural features. For each group of features, a merit was computed:

$$\text{Merit}_s = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}. \quad (9)$$

In (9), Merit_s is the heuristic merit of the subset S , containing k features, \bar{r}_{cf} represents the average correlation of the features with the class parameter, while \bar{r}_{ff} is the average correlation between the features. These correlations were established using the symmetrical uncertainty formula [10]. This method was implemented in combination with genetic search, in order to obtain a complete set of attributes subsets to be analyzed [16].

Another feature selection method, the information gain attribute evaluation, that performed the assessment of the individual attributes was also used. Each attribute was assigned a score based on the information gain between itself and the class:

$$IG_i = H(C) - H(C | A_i), \quad (10)$$

where $H(C)$ is the entropy of the class before observing the attribute A_i , respectively, $H(C | A_i)$ is the entropy of the class after observing the attribute A_i . The amount by which the entropy of the class decreased after observing the attribute A_i revealed the additional information about the class and constituted the information gain which was due to the attribute A_i .

During the Validation Phase, we adopted the classifiers of multilayer perceptron (MLP) [11] and support vector machines (SVM) [11], as they led to the best results in our former experiments [6]. The AdaBoost combination scheme [11], having the methods of MLP and SVM as basic classifiers, was also implemented.

3. Results and Discussions

3.1. Description of the Experiments. We considered a number of 300 patients suffering from HCC, 100 patients with hemangioma, 70 patients with colonic tumors, and 70 patients with inflammatory bowel diseases. Each of the considered type of disease corresponded to a class in the training, respectively, test set. These classes were combined in equal proportions inside the dataset. For each patient, three to five images were considered, acquired for various orientations of the transducer. The images were acquired using a Logiq 7 ultrasound machine, at the frequency of 5.5 MHz, the depth being 16 cm. During the image analysis phase, rectangular regions of interest, having 50×50 pixels in size, were selected on each type of analyzed tissue. After performing noise reduction using an averaging filter, the old textural features were computed independently of the orientation and illumination conditions. The new textural features were computed for various values of the parameters, as described previously. The values of the textural features, for each region of interest, were stored in specific files, for further analysis.

During the learning phase, the feature selection and classification experiments were performed using the methods of the Weka 3.5 library [16]. For feature selection, the method of correlation-based feature selection (CFS) was used, in conjunction with genetic search. For the genetic search method, the seed had the value 1, the crossover probability was 0.6, the mutation probability was 0.033, the population size was 20, and the number of generations to be evaluated was 20. The feature selection method that performed feature evaluation based on the information gain of the attributes with respect to the class, the information gain attribute evaluation method of the Weka 3.5 library, was also implemented during the learning phase, in conjunction with the Ranker search method.

During the validation phase, the Weka 3.5 versions of the support vector machines (SVM) method, the multilayer perceptron (MLP) classifier, and the AdaBoost combination scheme that used the SVM and MLP classifiers as basic learners were implemented. In the case of the SVM classifier, John's Platt sequential minimal optimization (SMO) algorithm for training a support vector classifier was used [16]. The polynomial kernel of 3rd degree, which provided the best result in our former experiments [6], was adopted for the SVM method. In the case of the MLP method, the learning rate was 0.2 in order to obtain a refined learning process and to avoid overtraining. The momentum was 0.8 in order to achieve a fast crossing over the plane areas of the learning surface. The number of nodes from the hidden layer was the arithmetic mean between the number of the input features and the number of classes. The AdaBoost M1 combination procedure of Weka 3.5, with 10 iterations, was implemented as well. The stratified cross-validation method strategy with 5 folds was used for classifier evaluation. Thus, for each iteration of the cross-validation method, the training set was formed by considering 80% of the data instances, while the test set consisted of 20% of the data instances.

3.2. Results

3.2.1. Performing the Differentiation between the HCC Tumor and the Cirrhotic Liver Parenchyma on Which the Tumor Evolved. Figure 2 illustrates the classification performance achieved by using the group of third order GLCM textural features, in comparison with that achieved by using the group of second order GLCM textural features. In this situation, the second- and the third order GLCM features were averaged after considering all the adopted directions. The following feature sets were taken into account: the second order GLCM features combined with the other textural features, represented with red color in Figure 2; the third order GLCM features combined with the other textural features, represented with yellow; the entire set of textural features, consisting of the third order GLCM features, the second order GLCM features, and the other textural features, represented with blue. Figure 2 illustrates the recognition rates obtained for these sets of features using the adopted classifiers. From Figure 2, it results that the third order GLCM led to a better classification performance than the second order GLCM in most of the situations. However, the

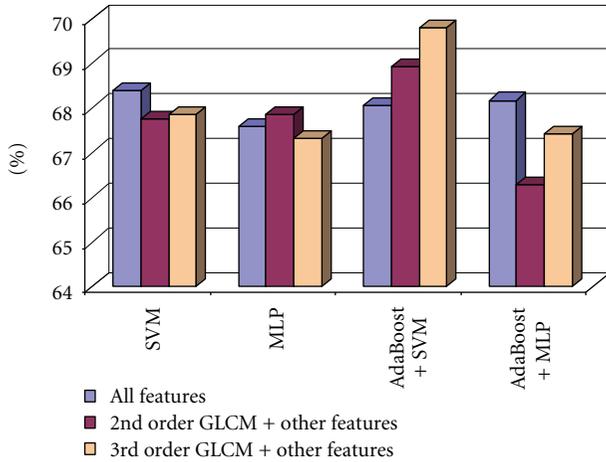


FIGURE 2: The evaluation of the 3rd order GLCM performances compared with those of the 2nd order GLCM, when considering the averaged values for all the directions, in the case of the differentiation between HCC and cirrhotic parenchyma.

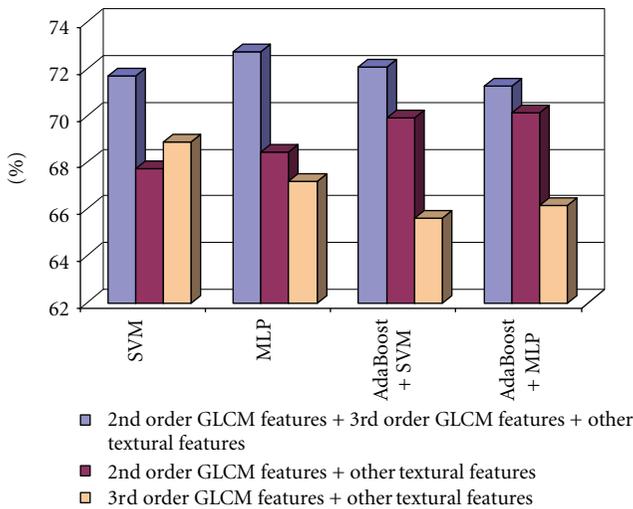


FIGURE 3: The recognition rate obtained when considering the 2nd order GLCM features combined with the other textural features, the 3rd-order GLCM textural features and the other textural features, respectively the 2nd-order GLCM, the 3rd-order GLCM, and the other textural features, in the case of the differentiation between HCC and cirrhotic parenchyma.

best results were obtained when considering all the textural features, involving the features derived from both types of GLCM, combined with the other textural features.

Further experimental steps consisted of the assessment of the combinations between the directions of the displacement vectors, in order to detect the combination that leads to the best classification accuracy. The best classification accuracy was obtained when considering only the (0°, 270°) combination of directions. When combining the third order GLCM Haralick features obtained for the (0°, 270°) directions with all the other textural features (including the second order GLCM features), we obtained the best classification

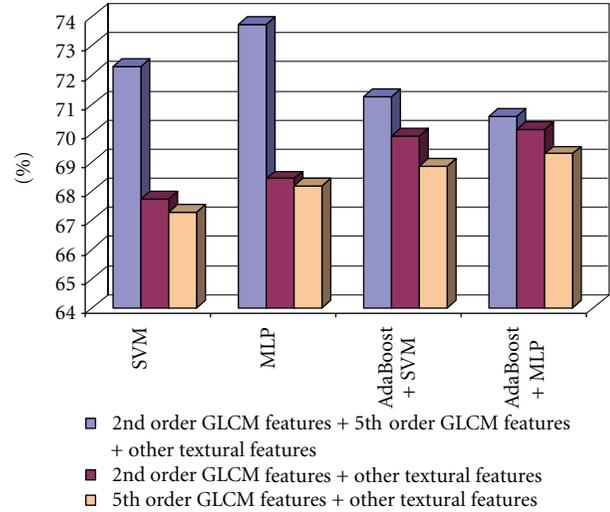


FIGURE 4: The recognition rate obtained when considering the 2nd order GLCM features combined with the other textural features, the 5th-order GLCM textural features and the other textural features, respectively, the 2nd order GLCM, the 5th-order GLCM, and the other textural features, in the case of the differentiation between HCC and cirrhotic parenchyma.

performance. Figure 3 illustrates the results concerning the improvement in the recognition rate.

Considering the case of the fifth order GLCM, the best classification results were obtained for the directions (0°, 180°, 90°, and 270°). The classification accuracy, resulted after combining the fifth order GLCM parameters obtained for the above group of displacement vector directions, combined with the other textural feature, is depicted in Figure 4.

As we can notice, the classification accuracy was more increased when considering all the textural features, than in the case when we used only the old textural features, achieving the maximum value of 73.75% in the case of the multilayer perceptron (MLP) classifier. We also considered the features derived from the second- and third order cooccurrence matrices of edge orientations. The second order cooccurrence matrix of edge orientations (EOCM) was computed for the directions 0°, 45°, 90°, and 135°, and the values of the resulted features were averaged. The second order EOCM features, combined with the old textural features, led to a recognition rate of 75%, for the AdaBoost metaclassifier that used MLP as a basic classifier, and to a value of AUC above 80%.

The third order EOCM matrix was computed in a similar way with the third order GLCM matrix. The relevant textural features, obtained after applying the methods of correlation-based feature selection (CFS) and information gain attribute evaluation, indicated the prevalence of the (45°, 315°) and (0°, 90) pairs of the displacement vector directions. After providing the extended Haralick features to the inputs of the SVM and MLP classifiers, the (0°, 90°) pair of directions provided the best results, as illustrated in Table 1.

After combining the second- and third order EOCM features with the old textural features, we obtained a

TABLE 1: The classification performance obtained by using the third order EOCM features for the most important combinations of displacement vector directions.

| Comb. of dir. | Classif. meth. | Recog. rate | TP rate | TN rate | AuC | Time |
|---------------|----------------|-------------|---------|---------|-------|----------|
| (45°, 315°) | SVM | 65.27% | 48.9% | 84.6% | 63.3% | 129.25 s |
| | MLP | 61.176% | 58.2% | 64.2% | 63.9% | 19.08 s |
| (0°, 90°) | SVM | 65.81% | 46.6% | 85% | 65.8% | 157.95 s |
| | MLP | 62.51% | 60.9% | 64.2% | 67.6% | 17.74 s |

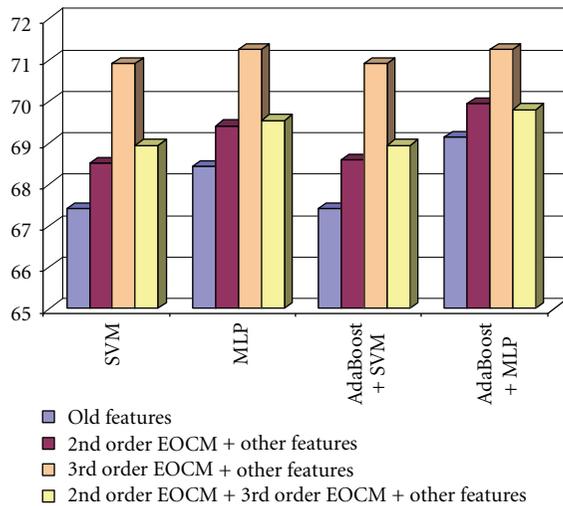


FIGURE 5: The recognition rate obtained when considering the 2nd order EOCM features combined with the old textural features, the 3rd order EOCM textural features and the old textural features, respectively, the 2nd order EOCM, the 3rd order EOCM, and the old textural features, in the case of the differentiation between HCC and cirrhotic parenchyma.

recognition rate situated above 71%. The combinations between the second order EOCM textural features and the old textural features, respectively, between the third order EOCM textural features and the old textural features, always led to an accuracy improvement compared with the case when only the old features were used. The combination between the second order EOCM textural features, the third order EOCM textural features, and the old textural features led, in most of the situations, to an accuracy improvement, compared with the cases when only the old textural features, or the combination between the second order EOCM textural features and the old textural features were used. The combination between the third order EOCM textural features and the old textural features provided the best recognition rates, in all the situations. These results can be visualized in Figure 5.

The final set of the relevant textural features, for the case of differentiation between HCC and cirrhotic parenchyma, resulted after performing feature selection on the group formed by the old textural features, by the third order GLCM features, by the fifth order GLCM features, by the second order EOCM features, and by the third order EOCM features. This set consisted of the union between the features selected by the CFS method, and those selected by the method of

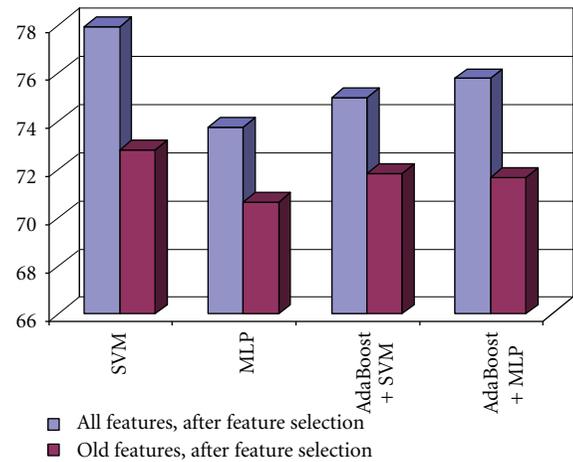


FIGURE 6: The increase of the recognition rate, obtained when using the new relevant textural features, compared with those obtained when using the old relevant textural features in the case of the differentiation between HCC and cirrhotic parenchyma.

information gain attribute evaluation. After we provided the values of the relevant textural features at the classifier inputs, we obtained a recognition rate of almost 78% for the MLP classifier and an AUC of above 82% for the same classifier, as we can observe in Table 2.

The increase in the recognition rate, obtained by using the final set of the new and old relevant textural features, compared with the accuracy due only to the old relevant textural features is depicted in Figure 6. Thus, an increase in accuracy from 71% to almost 78% was achieved, due to the new textural features, in the case of differentiation between HCC and the cirrhotic parenchyma on which the tumor evolved.

The Relevant Textural Features for the Differentiation between HCC and the Cirrhotic Parenchyma on Which HCC Had Evolved. After performing feature selection using the CFS and information gain attribute evaluation methods, the most important textural features contained in the union of the two feature subsets were the mean of the grey levels, indicating differences in echogenicity between the HCC tumor and the cirrhotic liver parenchyma, because, as it is well known, the HCC tumor, in advanced evolution phases, is hyperechogenic in most of the cases. The third- and fifth order GLCM correlation and the autocorrelation index indicated differences in granularity between the HCC tumor and the cirrhotic liver parenchyma. The second- and third

TABLE 2: The accuracy results obtained by considering the final set of relevant textural features.

| Classifier | Recognition rate | TP rate | TN rate | AUC | Time |
|----------------|------------------|---------|---------|-------|---------|
| SVM | 77.94% | 78.3% | 77.6% | 77.9% | 78.21 s |
| MLP | 73.75% | 70.1% | 77.4% | 82.3% | 54.32 s |
| AdaBoost + SVM | 75% | 79% | 71% | 78.6% | 86.25 s |
| AdaBoost + MLP | 75.79% | 74.6% | 76.7% | 82% | 87.55 s |

order GLCM homogeneity, second- and third order GLCM contrast, and the fifth order GLCM variance, provided information about the inhomogeneous aspect of the tumor tissue. The fifth order GLCM energy and entropy, the third order EOCM entropy, and the entropy computed after applying the Wavelet transform at the first level and at the second level, on the low-low component, were increased in the case of the tumor tissue, indicating its chaotic structure. The edge orientation variability and the frequency of the Laws textural microstructures indicated the complexity of the malignant tumor, which was constituted by multiple types of tissues.

The Values of the Maximum Probability Parameters. In the case of the third order GLCM, the maximum value of the probability for a given combination of three grey levels to appear within the considered class of tissue was around 0.01 in both cases of HCC and cirrhotic parenchyma on which HCC had evolved. This result was derived as an arithmetic mean of the maximum probability parameters computed on all the images belonging to the 300 patients included in the dataset. This probability was decreased in comparison with the same parameter computed in the case of the second order GLCM, when the mean value of the maximum probability was 0.05. The experimental results also revealed that groups of three hypoechoic pixels (57, 57, 57) corresponding to pure tumor regions with active growth, or to regions affected by necrosis, appeared frequently inside HCC. These values appeared rarely inside the cirrhotic parenchyma and inside the benign liver tumors.

In the case of the fifth order GLCM, the probability for a given combination of five gray level values to occur in the region of interest was computed separately for HCC and the cirrhotic parenchyma, for the (0°, 180°, 90°, 270) group of displacement vector directions, which provided the best accuracy results. The maximum probability had the value of 0.0035 in the case of HCC, respectively, 0.0036 in the case of the cirrhotic parenchyma. Thus, the probability was higher in the case of the cirrhotic parenchyma and lower in the case of HCC. This was a normal result, if we take into consideration the chaotic structure of the HCC tissue. For the second order EOCM, the maximum probability for a pair of two edge orientations to appear inside the tissue of the cirrhotic parenchyma was 0.132 while the value of the same parameter in the case of HCC was 0.131. This also emphasized the chaotic character of the HCC tissue, and the more regular character of the cirrhotic parenchyma. The most frequently met pair of two edge orientation values was (0°, 89°) inside both HCC and cirrhotic parenchyma regions,

corresponding to the directions of the liver tissue fibers and of the separating walls. For the third order EOCM, the maximum probability for a combination of three values of the edge orientation feature to appear within the HCC tissue was 0.00115 while in the case of the cirrhotic parenchyma, the value of this parameter was 0.00129. The most frequently met combination of three edge orientation values was (90°, 45°, 45°) inside the HCC tissue, respectively, (45°, 0°, 45°) inside the cirrhotic parenchyma on which HCC had evolved.

3.2.2. Performing the Differentiation between HCC and the Benign Liver Tumors. The third order GLCM was also experimented in the case of the differentiation between HCC and the benign liver tumors. After performing relevant feature selection using the CFS and information gain attribute evaluation methods, the features corresponding to the (45°, 225°), (45°, 135°), respectively, (0°, 90°) direction pairs appeared to be relevant. After applying the SVM and MLP classifiers for the final assessment of the efficiency of the displacement vector direction pairs, we noticed that the (0°, 90°) combination provided the best results.

The comparison between the recognition rates obtained in the cases of using the second order GLCM textural features combined with the other textural features, the third order GLCM textural features combined with the other textural features, respectively, the second order GLCM, the third order GLCM and the other textural features, is illustrated in Figure 7. As we can notice, the combination between the third order GLCM textural features and the other textural features outperformed the two other groups of features in most of the situations. The best recognition rate, of 76.88%, was obtained in the case of applying the AdaBoost combination scheme that used the MLP classifier as a basic learner.

Concerning the fifth order GLCM, the assessment of the two considered directions groups revealed that using all the textural features, provided by both versions of the fifth order GLCM, led to the best results. After the combination of the fifth order GLCM features with the other textural features (except the second order GLCM features), the recognition rates were always higher than in the case of using only the second order GLCM features and the other textural features. The best recognition rate was achieved when considering both the second order GLCM and the fifth order GLCM textural features, together with the old textural features. This result can be visualized in Figure 8.

The highest values of the accuracy parameters were obtained for the combination between the second order GLCM textural features, the fifth order GLCM textural features and

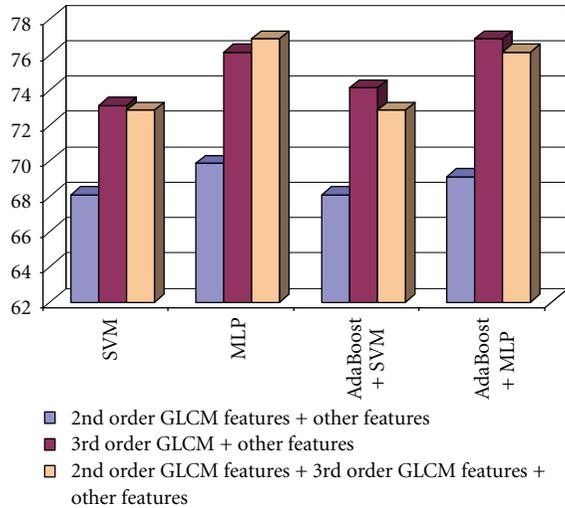


FIGURE 7: The recognition rates obtained when considering the group of 2nd order GLCM features combined with the other textural features, those of the 3rd order GLCM textural features together with the other textural features, respectively, the groups formed by 2nd order GLCM, the 3rd order GLCM, and the other textural features in the case of the differentiation between HCC and the benign liver tumors.

the old textural features, in the case of the MLP classifier, respectively, AdaBoost metaclassifier that used the MLP method as a basic classifier. In the latter case, the best value of the recognition rate, of 75.90%, was achieved.

The second order EOCM textural features were computed in a similar way as in the case of differentiating between the HCC tumor and the cirrhotic parenchyma. Concerning the textural features derived from the third order EOCM matrix, the CFS and information gain feature evaluation methods were applied again for relevant feature selection. The two sets of important textural features emphasized the frequency of those attributes corresponding to the $(0^\circ, 90^\circ)$, $(0^\circ, 180^\circ)$, and $(0^\circ, 270^\circ)$ combinations of directions. After the assessment through the MLP and SVM classifiers, the $(0^\circ, 270^\circ)$ direction group was found to be the best. The results obtained after combining the second- and third order EOCM textural features, corresponding to the $(0^\circ, 270^\circ)$ pair of displacement vector directions, with the old textural features are illustrated in Figure 9. It results, from Figure 9, that the best recognition rates were achieved when combining both the second- and third order EOCM features with the old textural features. Also, the combination between the third order EOCM textural features and the old textural features led to a better recognition rate that in the cases when using the combination between the second order EOCM features and the old textural features, respectively, only the old textural features.

The best recognition rate, of 72.75%, was achieved in the case of combining the third order EOCM features with the old textural features, and using the AdaBoost combination scheme in conjunction with the MLP method. The old textural features, the features derived from the third- and fifth order GLCM matrix, respectively from the second- and

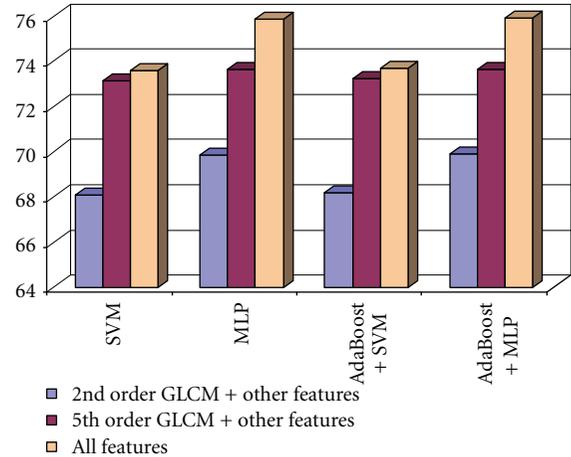


FIGURE 8: The recognition rate obtained when considering the 2nd order GLCM features and the other textural features, the 5th order GLCM textural features and the other textural features, respectively the 2nd order GLCM, the 5th order GLCM, and the other textural features in the case of the differentiation between HCC and the benign liver tumors.

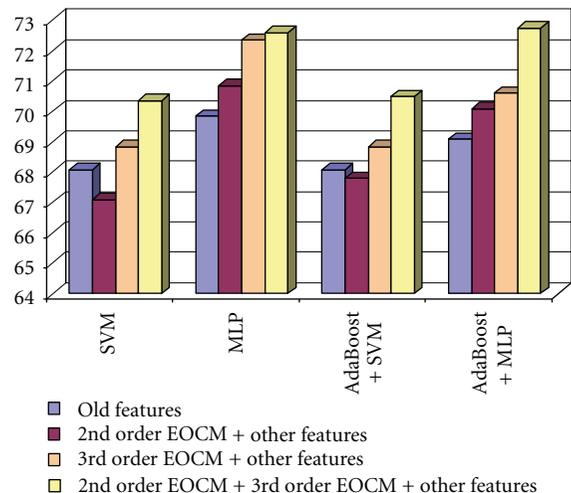


FIGURE 9: The comparison of the recognition rates obtained by combining the 2nd and 3rd order EOCM features with the old textural features in the case of the differentiation between HCC and the benign liver tumors.

third order EOCM matrix, were finally combined and a single group of textural features was obtained. After applying the CFS and information gain attribute evaluation methods, the final set of relevant textural features for differentiating HCC from hemangioma resulted as the union between the two resulted feature subsets. The values of the accuracy parameters resulted after providing the final set of relevant textural features at the classifiers inputs is illustrated in Table 3. As we can notice, a recognition rate of 83.66% was obtained in the case of AdaBoost combination scheme that used the MLP as basic classifier, and also an increased AUC, of 89.9%, was obtained for the MLP classifier. We also remark

TABLE 3: The values of the accuracy parameters obtained by using the final set of relevant textural features appropriate for the differentiation between HCC and the benign liver tumors.

| Classifier | Recognition rate | TP rate | TN rate | AUC | Time |
|----------------|------------------|---------|---------|-------|---------|
| SVM | 83.16% | 78.9% | 87.4% | 83.2% | 81.32 s |
| MLP | 82.66% | 78.4% | 86.9% | 89.9% | 62.27 s |
| AdaBoost + SVM | 83.21% | 79% | 88.3% | 83.9% | 89.22 s |
| AdaBoost + MLP | 83.66% | 80.4% | 86.9% | 84.3% | 89.16 s |

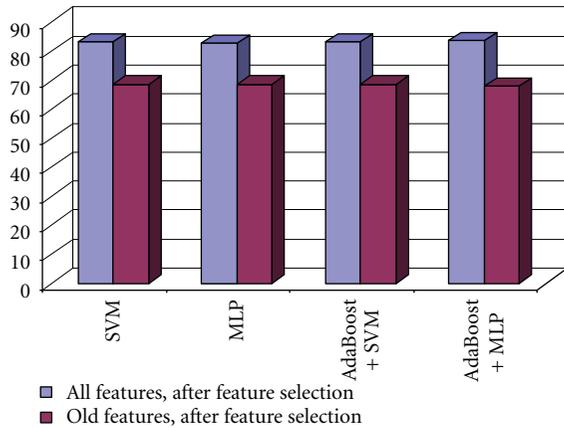


FIGURE 10: Comparison between the recognition rates obtained by using the entire set of relevant textural features, respectively, the old set of relevant textural features in the case of the differentiation between HCC and the benign liver tumors.

on the increased specificity (TN rate), always situated above 86%.

In Figure 10, we can also visualize the increase in the recognition rate, due to the new derived textural features, compared with that obtained by providing the original set of relevant textural features at the classifier inputs. Thus, an accuracy increase, from 70%, to 80% can be noticed.

The Relevant Textural Features for the Differentiation between HCC and the Benign Tumors. In the final set of relevant textural features, we noticed the presence of the second-, third-, and fifth order GLCM features, which played an important role in the differentiation between the HCC tumor and the benign tumors. Features like the third order GLCM homogeneity, third- and fifth order GLCM contrast, respectively, the third order GLCM variance, emphasized the difference in homogeneity and complexity in the grey level structure between the HCC tumor and the benign tumors. The fifth order GLCM correlation, together with the autocorrelation index, revealed differences in granularity between the malignant tumors and the benign tumors. The energy and entropy of the third order EOCM matrix were important as well, putting into evidence the uniformity of edge orientations present in the case of the benign tumors, and the lack of it in the case of the malignant tumors, where the entropy parameter has higher values. The features computed after applying the Wavelet transform were also important. The entropy was more emphasized at the

second level of resolution in the case of HCC. The spot textural microstructures, determined after applying the Laws convolution filters, were frequently met inside HCC and sparsely met inside the benign tumor tissue. Thus, the spots contributed to the differentiation of the HCC tumor from the cirrhotic liver parenchyma and from the benign tumors as well.

The Values of the Maximum Probability Parameters. The value of the maximum probability parameter determined in the case of the third order GLCM, equivalent with the probability to encounter a combination of three grey levels inside the benign liver tumors, was 0.05, which was higher than the value of the same parameter computed in the case of HCC (0.01), emphasizing, once more, the chaotic structure of the malignant tumor tissue. The average value of the maximum probability parameter computed in the case of the fifth order GLCM matrix was 0.007 for the class of benign tumors, being again more increased than the value of the same parameter computed in the case of HCC, 0.003. Groups of three hypoechoic pixel values (53, 53, 53) were frequently encountered inside the benign tumors, corresponding to vascular lakes. In the case of the second order EOCM, the maximum probability for a pair of two edge orientation values to occur inside the benign tumor region was 0.138, this being larger than the value of the same parameter computed in the case of HCC (0.131). The pair of edge orientation values that most often appeared inside the benign tumor regions was $(0^\circ, 89^\circ)$, being similar with the edge orientation pair that was met in the case of HCC and cirrhotic parenchyma. The average value of the maximum probability parameter, computed inside the third order EOCM matrix in the case of the benign liver tumors, was 0.0021, being more increased than the same value obtained in the case of the HCC tumor, of 0.0013. The most frequently met combination of three edge orientation values inside the benign tissue was $(90^\circ, 90^\circ, 90^\circ)$, denoting the more regular structure of the tissue.

3.2.3. Performing the Differentiation between the Colorectal Tumors and the Inflammatory Bowel Diseases (IBD). In the case of the comparison between the colo-rectal tumors and the inflammatory bowel diseases, the best improvement in the classification accuracy was provided by the textural features derived from the third order GLCM, respectively, by those resulted from the third order EOCM. The best combination of displacement vector directions was $(0^\circ, 270^\circ)$ in the case of the third order GLCM, and $(0^\circ, 180^\circ)$ in the

TABLE 4: The values of the accuracy parameters obtained by using the final set of relevant textural features appropriate for the differentiation between the colo-rectal tumors and the inflammatory bowel diseases.

| Classifier | Recognition rate | TP rate | TN rate | AUC | Time |
|----------------|------------------|---------|---------|-------|---------|
| SVM | 94.93% | 94.9% | 94.9% | 94.9% | 77.18 s |
| MLP | 94.3% | 93.9% | 94.1% | 98.3% | 64.28 s |
| AdaBoost + SVM | 94.93% | 94.9% | 94.9% | 94.9% | 85.34 s |
| AdaBoost + MLP | 92.50% | 91.1% | 89.9% | 97.7% | 83.11 s |

case of the third order EOCM. The comparison between the recognition rates obtained in these cases, and in the case of the original textural features, is illustrated in Figure 11. The combination between the third order GLCM textural features, the second order GLCM textural features, and the other textural features always provided the best recognition rate, situated above 90%. The combination between the second order GLCM features, the third order EOCM features, and the old textural features also provided an increase in accuracy, compared with the set of old textural features, in all the situations.

Finally, the second order GLCM textural features, the third order GLCM textural features, the third order EOCM textural features, and the other textural features were combined, and then the relevant textural features were selected. The final set of relevant textural features resulted after performing the union operation between the subsets of important textural features provided by each of the feature selection methods. The comparison between the recognition rates obtained by using the final set of relevant textural features, respectively, the initial set of relevant textural features, obtained by considering only the old textural features, is depicted in Figure 12. An increase in accuracy from 85% to almost 94% can be noticed.

The values of all the considered accuracy parameters, in the case of the final set of relevant textural features, are illustrated in Table 4. The best recognition rate, of 94.93%, was obtained in the case of the SVM classifier, respectively, in the case of AdaBoost combination scheme that used the SVM as basic classifier. We also noticed the increased value of AUC, of 98.3%, obtained in the case of the MLP classifier.

The Relevant Textural Features for the Differentiation between the Colorectal Tumors and the Inflammatory Bowel Diseases. The third order GLCM homogeneity, as well as the third order EOCM homogeneity resulted to be important in order to distinguish between the colo-rectal tumors and the IBD, due to the heterogeneous structure of the tumor tissue. The energy and the entropy features were also relevant when derived from the second order GLCM, from the third order GLCM, as well as from the third order EOCM, emphasizing the chaotic structure and the irregular aspect of the colorectal tumor tissue, respectively, the more regular aspect of the bowel wall that correspond to the IBD case. The entropy computed at the first level after applying the wavelet transform was also important in this context. Concerning the textural microstructures obtained after applying the Laws convolution filters, the spots and the waves appeared to

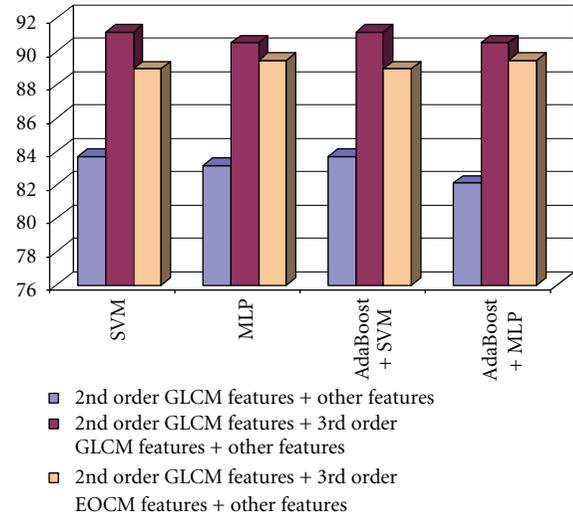


FIGURE 11: The recognition rate obtained when considering the 2nd order GLCM features and the other textural features, the 2nd order GLCM features, the 3rd order GLCM, and the other textural features, respectively, the 2nd order GLCM features, the 3rd order EOCM features, and the other textural features in the case of the differentiation between the colorectal tumors and IBD.

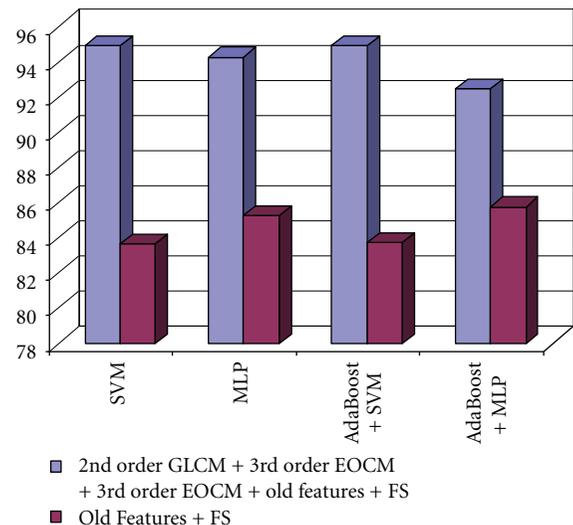


FIGURE 12: Comparison between the recognition rates obtained by using the entire set of relevant textural features, respectively, the old set of relevant textural features in the case of the differentiation between the colorectal tumors and IBD.

be more emphasized in the tumor region, suggesting the presence of severe fibrosis and also the complex structure of the tumor.

The Values of the Maximum Probability Parameters. The maximum probability for a pair of three grey level values to occur in the region of interest was 0.009 inside the colo-rectal tumors, respectively, 0.014 on the bowel wall affected by IBD. Groups of hyperechogenic values, corresponding to tissue regions strongly affected by fibrosis, were often met inside the colo-rectal tumors. The maximum probability for a pair of three edge orientation values to occur was 0.0013 inside the colo-rectal tumors and 0.0026 on the bowel wall affected by IBD. The most frequent group of edge orientations that appeared inside the colo-rectal tumor was $(90^\circ, 45^\circ, 90^\circ)$, respectively, on the bowel wall, this group was $(90^\circ, 90^\circ, 90^\circ)$. This fact confirms the more complex aspect of the bowel wall that exists in the presence of the colo-rectal tumor.

3.3. Discussions. As the experiments revealed, the new implemented methods for texture analysis, based on the superior order cooccurrence matrices led to a considerable accuracy increase and to a better emphasis of the malignant tumors characteristics, in comparison with those of the benign tumors and the tissue of the visually similar diseases. Concerning the orientations of the displacement vectors, the combinations between the horizontal and vertical directions led to the best accuracy results. The best orientations of the displacement vectors were also parallel or perpendicular on the direction of the ultrasound signal propagation. The combination between the old textural features, the third- and fifth order GLCM features, respectively, the second- and third order EOCM features, followed by feature selection, led to an increase of the recognition rate from 70% to 80% in both cases of differentiation between HCC and cirrhotic liver parenchyma, and between HCC and the benign tumors, respectively, to an accuracy improvement from 80% to 90% in the case of the colo-rectal tumor recognition. The probability of a certain combination of grey levels to appear inside the tissue, determined by the maximum probability of the grey level cooccurrence matrix, had higher values in the case of the benign tumors and of the visually similar tissues, and lower values in the case of the malignant tumors, putting into evidence the complex structure of the malignant tissue. In the case of the cooccurrence matrix of edge orientations, the same situation appeared. Also, the third order EOCM energy was higher inside the benign tumors and lower inside the malignant tumors, which indicated an increased uniformity of the edge orientations inside the benign tumors and the irregularity of the values of this feature in the case of the malignant tumors. The most frequently met edge orientation values inside the malignant tumors and the cirrhotic parenchyma are $0^\circ, 45^\circ$, and 90° , while in the case of the benign tumors and of the bowel wall affected by IBD, only the $0^\circ, 90^\circ$ values were met more often. This emphasized the complexity of the malignant tumors, respectively, of the tissue affected by diseases that precedes cancer, such as cirrhosis. We also noticed that the value of the maximum

probability parameter decreased while the order of the cooccurrence matrix increased. However, the cooccurrence matrices of order $n, n > 2$ always led to accuracy improvements and to a more refined characterization of the analyzed tissues, as shown in the experiments section. Concerning the relevant textural features that differentiated the malignant tumors from the other kinds of tissues, the homogeneity, the variance, and the contrast computed from the cooccurrence matrices indicated the heterogeneous structure of the malignant tumor tissue. The correlation, together with the autocorrelation index, emphasized a difference in granularity between the restructuring areas of the malignant tumors and the tissue zones corresponding to less aggressive diseases. The entropy computed after applying the wavelet transform revealed the presence of the chaotic character at multiple resolutions, in the case of the malignant tumors. The textural microstructures, determined after applying the Laws convolution filters, were also important in order to distinguish the malignant tumors from the visually similar tissues, emphasizing the complexity of the tissue affected by malignancy.

4. Conclusions and Future Work

The superior order grey level cooccurrence matrices, as well as the edge orientation cooccurrence matrices of superior order, led to an improvement of the classification performance, in comparison with the case when only the old textural features were used. The probability for a certain combination of grey levels or edge orientations to occur in the region of interest was lower in the case of the tumor tissue and higher in the case of the visually similar tissues. This fact reflected the irregular structure of the malignant tissues. The value of the maximum probability parameter decreased while the order n of the superior matrix increased, as the number of possible combinations of feature values increased and the evaluation became more refined. The final set of relevant textural features revealed, in each case, the presence of the new textural features, derived from superior order matrices, and emphasized the inhomogeneous, complex, chaotic structure of the malignant tumor tissue. The smaller classification accuracy obtained in the case of HCC tumor recognition is due mainly to the variations in the aspect of the HCC tumor, and also to the small differences that exist between the HCC and cirrhotic parenchyma tissues, both diseases involving a restructuring process. In the case of colorectal tumor recognition, the classes were more homogeneous, so the classification accuracy was higher. In our future work, we aim to divide the HCC tumor into subclasses and to improve the classification accuracy through multiclass classification. The specific groups of grey levels or edge orientation values that appeared inside each subclass of malignant tumors will be further analyzed and their correspondence with the tissue microstructures will be established. We will also implement more complex classifier combination schemes, such as stacking, in order to improve the automatic diagnosis performance. The computation of the extended Haralick features at multiple resolutions is a future research objective as well.

Appendices

A. The Hepatocellular Carcinoma and the Colonic Tumors: Medical Considerations and Visual Aspect in Ultrasound Images

The hepatocellular carcinoma (HCC) is the most frequent malignant liver tumor, representing 75% of the liver cancer cases, besides hepatoblastoma (7%), cholangiocarcinoma, and cystadenocarcinoma (6%). The most relevant oncogenic agent for HCC development is the chronic viral infection with the hepatitis B virus (HBV), or hepatitis C virus (HCV), the next evolution phase, preceding HCC, being cirrhosis [1]. HCC evolves from cirrhosis, after a restructuring phase, at the end of which dysplastic nodules (future malignant tumors) result. Concerning the visual aspect in ultrasound images, in incipient phase, HCC appears like a small region having a different texture than the other parts of the tissue and a diameter of about 1.5 cm to 2 cm. In the case of an evolved HCC, the essential textural attribute is that of heterogeneity, due to the coexistence of fatty regions, of regions with necrosis, fibrosis, and, respectively, active growths. HCC is also characterized through a complex structure of vessels [1]. Hemangioma is the most frequent form of benign liver tumors, consisting of a mass of abnormal blood vessels. Up to 5 percent of adults in the United States have small hemangiomas in their liver. Concerning the aspect in ultrasound images, most of the hemangiomas are isoechogetic and homogeneous [20]. The focal nodular hyperplasia is another frequent benign liver tumor. The necrosis and hemorrhage are rarely met inside these structures. The histological studies detected the existence of some fibrous bands, of multiple biliary ducts, and of some fibrous scars, of stellar shape. In ultrasound images, the FNH tumors appear as isoechogetic or hypoechogetic lesions, having a homogeneous aspect, but containing a small, stellar scar inside [20]. The colo-rectal tumors represent a frequent disease for the population of the developed countries, being the third cause of cancer-related death in the Western world. They arise from the adenomatous polyps which are present on the bowel wall. Like every tumor, they are characterized by the heterogeneity of the tissue structure and by the complexity and irregularity of the vessel structure [5]. In ultrasound images, they have an inhomogeneous, mixed aspect, the parietal delimitation being linear, but interrupted by the tumor invasion. The adenopathy could also be present inside the tumor, having rounded shape. Although distinct from IBD, they share a lot of characteristics with the latter, like wall thickening and increased vascularity [21]. Some eloquent examples of all the described affections can be visualized in Figure 13.

In Figure 13, the liver tumors are illustrated: an instance of HCC in focal encephalic form, evolved phase; an instance of hemangioma benign tumor; an instance of the FNH benign tumor. In Figure 14, an example of a colo-rectal tumor is illustrated. The shape modifications of the bowel wall, due to the colo-rectal tumor, are also visible in Figure 14(a). The inflammatory bowel diseases—Crohn's disease and ulcerohemorrhagic rectocolitis are also depicted.

B. Description of the Superior Order Haralick Features

The contrast of order n for the generalized cooccurrence matrices was computed as follows:

$$\text{Contrast} = \sum_{f_1=0}^{M-1} \sum_{f_2=0}^{M-1} \cdots \sum_{f_n=0}^{M-1} \left(\sum_{u=1}^{n-1} \sum_{v=u+1}^n \text{dif}(f_u, f_v)^2 \right) * p(f_1, f_2, \dots, f_n). \quad (\text{B.1})$$

In (B.1), f_1, f_2, \dots, f_n are the values of the descriptors for the considered features: in the case of the GLCM of order n , they are the grey levels of the pixels, while in the case of superior order EOCM, the values of the edge orientations are taken into account; $\text{dif}(f_u, f_v)$ is the absolute value of the difference between any two values of the local features involved in a certain n -tuple that gives the coordinates of a superior order matrix element. M is the maximum value that a certain feature (grey value of edge orientation) can achieve. The contrast, also called dissimilarity, estimates the difference that exists, in the region of interest, between the values of the considered local features, corresponding to pixels that are in a spatial relation defined by the displacement vectors. It measures the local variations in the grey level cooccurrence matrix.

Concerning the other measures, we defined the following mathematical formulas:

$$\text{Entropy} = \sum_{f_1=0}^{M-1} \sum_{f_2=0}^{M-1} \cdots \sum_{f_n=0}^{M-1} p(f_1, f_2, \dots, f_n) \times (-\ln p(f_1, f_2, \dots, f_n)). \quad (\text{B.2})$$

The entropy expresses the disorder of the texture, with respect to the considered feature (grey level or edge orientation value), the uncertainty for a certain combination of n values (f_1, f_2, \dots, f_n) to appear,

$$\text{Energy} = \sum_{f_1=0}^{M-1} \sum_{f_2=0}^{M-1} \cdots \sum_{f_n=0}^{M-1} p(f_1, f_2, \dots, f_n)^2. \quad (\text{B.3})$$

The energy, also called angular second moment, is the opposite of the entropy, measuring the order, the uniformity within the texture with respect to the considered local feature.

Local homogeneity

$$= \sum_{f_1=0}^{M-1} \sum_{f_2=0}^{M-1} \cdots \sum_{f_n=0}^{M-1} \frac{p(f_1, f_2, \dots, f_n)}{1 + \sum_{u=1}^{n-1} \sum_{v=u+1}^n \text{dif}(f_u, f_v)^2}. \quad (\text{B.4})$$

The local homogeneity characterizes the texture from the point of view of the similarity of the pixels with respect to the considered feature, having increased values when the difference between the feature values of the corresponding pixels is decreased. This feature measures the closeness of

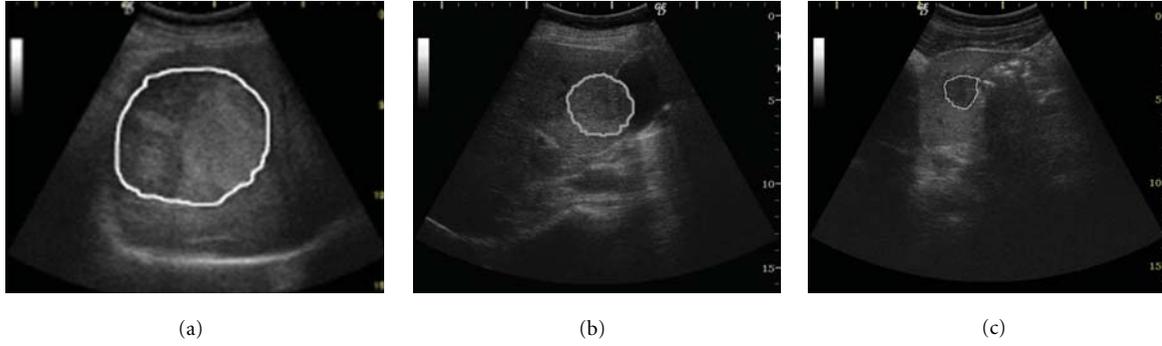


FIGURE 13: The malignant and benign liver tumors: (a) hepatocellular carcinoma, encephalic form, (b) hemangioma; (c) focal nodular hyperplasia.

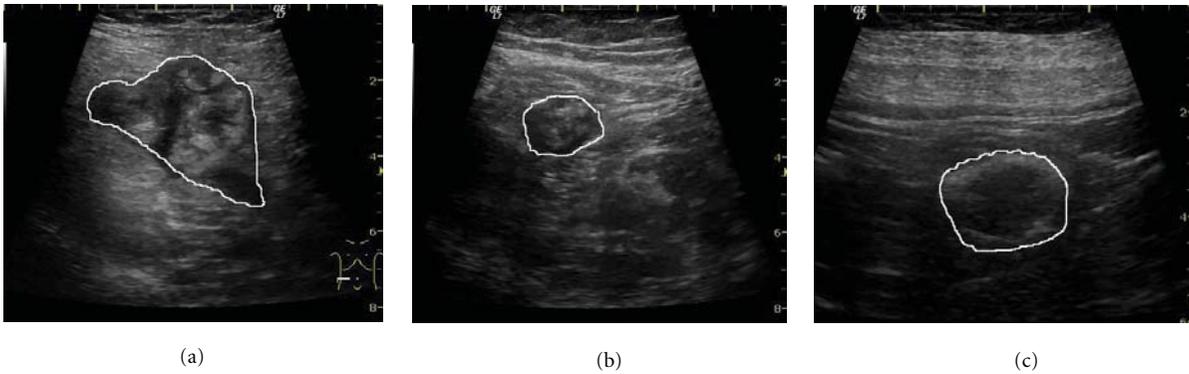


FIGURE 14: The colo-rectal tumors and the inflammatory bowel diseases: (a) colo-rectal tumor; (b) Crohn's disease; (c) ulcero-hemorrhagic recto-colitis.

the distribution of the elements in the GLCM to the GLCM diagonal.

The variance was computed as indicated;

$$\text{Variance} = \sqrt{\sigma_1^2 \sigma_2^2 \cdots \sigma_n^2}. \quad (\text{B.5})$$

The generalized variance, defined in (B.5), characterizes the texture from the point of view of the spreading of the considered feature values inside the cooccurrence matrix.

Correlation

$$\begin{aligned} &= \sum_{f_1=0}^{N_g-1} \sum_{f_2=0}^{N_g-1} \cdots \sum_{f_n=0}^{N_g-1} \left(\sum_{u=1}^{n-1} \sum_{v=u+1}^n \frac{(f_u - \mu_u)(f_v - \mu_v)}{\sqrt{\sigma_u^2 \sigma_v^2}} \right) \\ &\quad \times p(f_1, f_2, \dots, f_n). \end{aligned} \quad (\text{B.6})$$

The *correlation*, mathematically defined in (B.6), expresses the linear dependence, between every two values of the local features, that are met inside the same n -tuple, respecting the spatial relation established by the displacement vectors. The mean and variance of the superior order cooccurrence

matrices, with respect to the reference pixel (μ_1, σ_1) , are described:

$$\mu_1 = \sum_{f_1=0}^{M-1} f_1 \times \sum_{f_2=0}^{M-1} \cdots \sum_{f_n=0}^{M-1} p(f_1, f_2, \dots, f_n). \quad (\text{B.7})$$

The mean of the superior order cooccurrence matrices with respect to the neighboring pixels (σ_k, σ_n) was computed according to the definitions described below:

$$\begin{aligned} \mu_k &= \sum_{f_k=0}^{M-1} f_k \\ &\quad \times \sum_{f_1=0}^{M-1} \cdots \sum_{f_{k-1}=0}^{M-1} \sum_{f_{k+1}=0}^{M-1} \cdots \sum_{f_n=0}^{M-1} p(f_1, f_2, \dots, f_n), \\ &\quad k \geq 2, k < n, \\ \mu_n &= \sum_{f_n=0}^{M-1} f_n \times \sum_{f_1=0}^{M-1} \sum_{f_2=0}^{M-1} \cdots \sum_{f_{n-1}=0}^{M-1} p(f_1, f_2, \dots, f_n). \end{aligned} \quad (\text{B.8})$$

The variance of the superior order cooccurrence matrices, with respect to the reference pixel (σ_1) , is described:

$$\sigma_1^2 = \sum_{f_1=0}^{M-1} (f_1 - \mu_1)^2 \times \sum_{f_2=0}^{M-1} \cdots \sum_{f_n=0}^{M-1} p(f_1, f_2, \dots, f_n). \quad (\text{B.9})$$

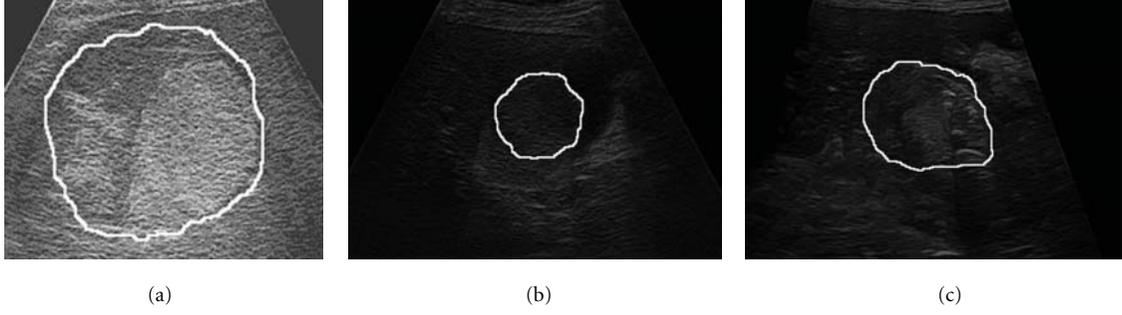


FIGURE 15: The properties of the edges inside the tumoral tissue (a) edges inside HCC; (b) edges inside hemangioma; (c) edges inside the colo-rectal tumor.

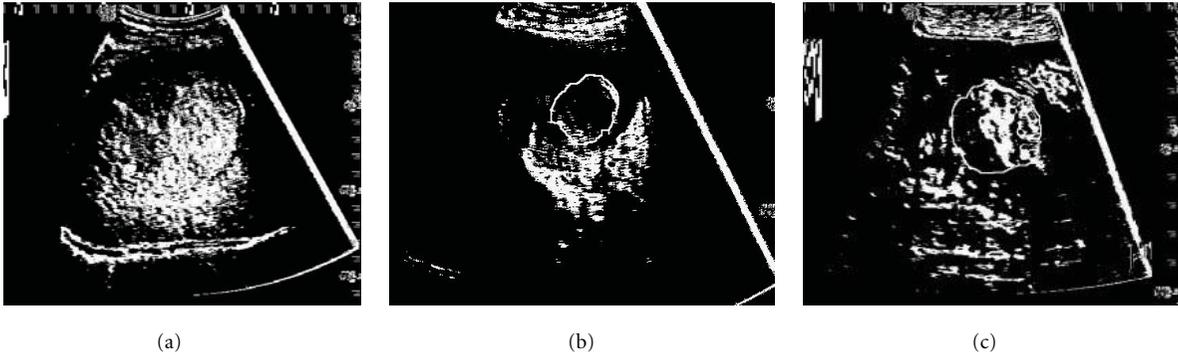


FIGURE 16: The spot textural microstructures inside the tumoral tissue (a) HCC image after the detection of the spot microstructures; (b) hemangioma image after the detection of the spot microstructures; (c) colo-rectal tumor after the detection of the spot microstructures.

The variance of the superior order cooccurrence matrices with respect to the neighboring pixels (σ_k, σ_n) was computed according to the definitions described below:

$$\begin{aligned} \sigma_k^2 &= \sum_{f_k=0}^{M-1} (f_k - \mu_k)^2 \\ &\times \sum_{f_1=0}^{M-1} \cdots \sum_{f_{k-1}=0}^{M-1} \sum_{f_{k+1}=0}^{M-1} \cdots \sum_{f_n=0}^{M-1} p(f_1, f_2, \dots, f_n), \\ &k \geq 2, k < n, \\ \sigma_n^2 &= \sum_{f_n=0}^{M-1} (f_n - \mu_n)^2 \\ &\times \sum_{f_1=0}^{M-1} \sum_{f_2=0}^{M-1} \cdots \sum_{f_{n-1}=0}^{M-1} p(f_1, f_2, \dots, f_n). \end{aligned} \quad (\text{B.10})$$

The maximum probability, defined below, represents the maximum value that appears in the cooccurrence matrix and also highlights the pair, or n -tuple of feature values, that appears most frequently in the region of interest, the corresponding pixels respecting the spatial relation specified by the displacement vectors,

$$p_{\max} = \text{Max}(p(f_1, f_2, \dots, f_n), 0 < f_k < M, 0 \leq k \leq n). \quad (\text{B.11})$$

C. The Role of the Texture Analysis Methods in Emphasizing the Tumor Characteristics

In Figure 15, we can visualize the specific properties of the edges inside the tumoral tissue. Thus, the edge contrast and edge orientation variability are higher inside the malignant tumors while inside the benign tumors the edge contrast and density are more decreased. In Figure 16, the density of the spot microstructures inside the tumoral region is put into evidence. The spot frequency is higher inside the malignant tumors, emphasizing tissue regions severely affected by fibrosis, or containing an increased number of fatty cells. Inside the benign tumors, the spot microstructures are sparse.

Acknowledgments

The paper was funded by Romanian Ministry of Education and Research, CNCSIS Project PNII-ID-PCE-833/2008, (*SonoDig*). All the authors have contributed equally to the current paper.

References

- [1] J. Bruix, "Hepatitis B virus and hepatocellular carcinoma," *Journal of Hepatology*, vol. 39, no. 1, pp. S59–S63, 2003.
- [2] U. Raeth, D. Schlaps, and B. Limberg, "Diagnostic accuracy of computerized B-scan texture analysis and conventional

- ultrasonography in diffuse parenchymal and malignant liver disease,” *Journal of Clinical Ultrasound*, vol. 13, no. 2, pp. 87–99, 1985.
- [3] H. Sujana, S. Swarnamani, and S. Suresh, “Application of artificial neural networks for the classification of liver lesions by image texture parameters,” *Ultrasound in Medicine and Biology*, vol. 22, no. 9, pp. 1177–1181, 1996.
- [4] H. Yoshida, D. D. Casalino, B. Keserci, A. Coskun, O. Ozturk, and A. Savranlar, “Wavelet-packet-based texture analysis for differentiation between benign and malignant liver tumours in ultrasound images,” *Physics in Medicine and Biology*, vol. 48, no. 22, pp. 3735–3753, 2003.
- [5] V. Atlamazoglou, D. Yova, N. Kavantzias, and S. Loukas, “Texture analysis of fluorescence microscopic images of colonic tissue sections,” *Medical and Biological Engineering and Computing*, vol. 39, no. 2, pp. 145–151, 2001.
- [6] D. Mitrea, S. Nedevschi, M. Lupsor, M. Socaciu, and R. Badea, “Advanced classification methods for improving the automatic diagnosis of the hepatocellular carcinoma, based on ultrasound images,” in *17th IEEE International Conference on Automation, Quality and Testing, Robotics, (AQTR ’10)*, pp. 265–270, Cluj-Napoca, Romania, May 2010.
- [7] A. Akono, N. T. Tankam, and E. Tonye, “Nouvel algorithme d’évaluation de texture d’ordre n sur la classification de l’occupation des sols de la region volcanique du mont Cameroun,” *Teledetection*, vol. 5, no. 1, pp. 227–244, 2005.
- [8] M. B. Huber, J. Carballido-Gamio, K. Fritscher et al., “Development and testing of texture discriminators for the analysis of trabecular bone in proximal femur radiographs,” *Medical Physics*, vol. 36, no. 11, pp. 5089–5098, 2009.
- [9] H. Anys and D. C. He, “Evaluation of textural and multipolarization radar features for crop classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 5, pp. 1170–1181, 1995.
- [10] M. A. Hall and G. Holmes, “Benchmarking attribute selection techniques for discrete class data mining,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 3, pp. 1–16, 2003.
- [11] R. Duda, *Pattern Classification*, Wiley Interscience, 2nd edition, 2003.
- [12] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 1989.
- [13] D. A. Clausi, “An analysis of co-occurrence texture statistics as a function of grey level quantization,” *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45–62, 2002.
- [14] E. J. Stollnitz and T. D. Rose, “Wavelets for computer graphics,” *IEEE Computer Graphics and Applications*, vol. 15, no. 3, pp. 76–84, 1995.
- [15] J. R. Parker, *Algorithms of Image Processing and Computer Vision*, Wiley Computer, 1996.
- [16] Weka 3, “Data mining software in java,” 2004, <http://www.cs.waikato.ac.nz/ml/weka/>.
- [17] B. Julesz, E. N. Gilbert, and J. D. Victor, “Visual discrimination of textures with identical third-order statistics,” *Biological Cybernetics*, vol. 31, no. 3, pp. 137–140, 1978.
- [18] R. M. Haralick, “Statistical and structural approaches to texture,” *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.
- [19] L. Davis and S. Jones, “Texture analysis using generalized cooccurrence matrices,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 3, pp. 251–259, 1979.
- [20] American Liver Foundation, <http://www.liverfoundation.org/abouttheliver/info/benigntumors/>.
- [21] L. Ruess, “Inflammatory bowel diseases in children and young adults: correlation of sonographic and clinical parameters during treatment,” *American Journal on Roentgenology*, vol. 175, no. 1, pp. 79–84, 2000.

Research Article

Approximating Ideal Filters by Systems of Fractional Order

Ming Li^{1,2}

¹ School of Information Science & Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China

² Department of Computer and Information Science, University of Macau, Avenue Padre Tomas Pereira, Taipa 1356, Macau, China

Correspondence should be addressed to Ming Li, ming_lihk@yahoo.com

Received 16 September 2011; Accepted 10 October 2011

Academic Editor: Carlo Cattani

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

The contributions in this paper are in two folds. On the one hand, we propose a general approach for approximating ideal filters based on fractional calculus from the point of view of systems of fractional order. On the other hand, we suggest that the Paley and Wiener criterion might not be a necessary condition for designing physically realizable ideal filters. As an application of the present approach, we show a case in designing ideal filters for suppressing 50-Hz interference in electrocardiogram (ECG) signals.

1. Introduction

Filters have wide applications in various fields, ranging from medical engineering to electrical engineering; see, for example, Hussain et al. [1], Bhattacharyya et al. [2], Fieguth [3], Bendat and Piersol [4], Gray and Davisson [5], and Li [6], just mentioning a few. In the field, the theory and techniques to approximate ideal filters are desired. There are some methods about approximating ideal filters, such as Butterworth filters, Chebyshev filters, Cauer-Chebyshev filters, and Bessel ones (Wanhammar [7], Mitra and Kaiser [8]).

Recall that the conventional filters of Butterworth type, Chebyshev type, Cauer-Chebyshev type, or Bessel one are discussed in the domain of systems of integer order. More precisely, the frequency response of a filter that is denoted by $H(\omega)$ is a rational function. Both the denominator and the numerator of the rational function are polynomials of integer order; see [7, 8], Vegte [9], Dorf and Bishop [10], and Li [11]. From the point of view of mathematical analysis, conventional filters are in the domain of calculus of integer order.

This paper aims at providing an approach to approximate ideal filters by using frequency responses of fractional-order. The basic idea is like this. Denote by ω_c the cutoff frequency of a filter. Then, $H(\omega_c) = 0$ from a view of ideal filters. In this case, we present the following approximation:

$$\lim_{r \rightarrow 0} |H(\omega)|^r = \begin{cases} 1, & \omega \neq \omega_c, \\ 0, & \omega = \omega_c, \end{cases} \quad (1)$$

where $|H(\omega)|$ is the amplitude of $H(\omega)$.

An obvious advantage of the present approach is that the above always holds no matter what the concrete structure of $H(\omega)$ is. However, theoretically speaking, $H^r(\omega)$ has to be explained from the point of view of fractional calculus.

The remaining paper is organized as follows. Section 2 explains the research background. The problem statement is described in Section 3. The present approximation is given in Section 4. A case study is stated in Section 5, which is followed by conclusions.

2. Research Background

2.1. Glimpse at Ideal Filters. The ideal lowpass filter implies that the amplitude of the frequency response is given by

$$|H(\omega)| = \begin{cases} 1, & \omega < \omega_c \\ 0, & \text{elsewhere,} \end{cases} \quad (2)$$

where $H(\omega_c) = 0$. One says that $H(\omega)$ is the frequency response of an ideal highpass filter if

$$|H(\omega)| = \begin{cases} 1, & \omega > \omega_c, \\ 0, & \text{elsewhere.} \end{cases} \quad (3)$$

The ideal bandpass filter has the frequency response expressed by

$$|H(\omega)| = \begin{cases} 1, & \omega_{cl} < \omega < \omega_{ch}, \\ 0, & \text{elsewhere,} \end{cases} \quad (4)$$

where ω_{c_l} and ω_{c_h} are cut-off frequencies. A filter is said to be ideal band stop if its frequency response function is given by

$$|H(\omega)| = \begin{cases} 0, & \omega_{c_l} < \omega < \omega_{c_h}, \\ 1, & \text{elsewhere.} \end{cases} \quad (5)$$

2.2. Paley and Wiener Criterion. For facilitating the discussions, we write

$$H(\omega) = R(\omega) + jX(\omega) = A(\omega)e^{-j\vartheta(\omega)}, \quad (6)$$

where $\vartheta(\omega)$ is the phase response of a filter. Note that the condition for $F^{-1}[H(\omega)] = h(t)$ to be zero for negative t , where F^{-1} implies the inverse of the Fourier transform, is that $A(\omega)$ must be square-integrable. That is,

$$\int_{-\infty}^{\infty} A^2(\omega)d\omega < \infty. \quad (7)$$

The above implies the causality of a filter; see, for example, Papoulis [12]. A necessary and sufficient condition for $A(\omega)$ to satisfy (7) is explained by Paley and Wiener [13]. That condition is called the Paley and Wiener condition or the Paley and Wiener criterion. It is expressed by

$$\int_{-\infty}^{\infty} \frac{|\ln A(\omega)|}{1 + \omega^2} d\omega < \infty. \quad (8)$$

The Paley and Wiener criterion implies that ideal filters are not physically realizable because $A(\omega) = 0$ in a certain frequency range for each type of ideal filters. Therefore, approximations of ideal filters are desired.

2.3. Some Filters of Integer Order for Approximating Ideal Filters. Various methods in the approximations are studied, such as Butterworth filters, Chebyshev's, Cauer-Chebyshev's, and Bessel's filters; see, for example, [2], and Lam [14].

Taking lowpass filtering as an example, the system function of the Butterworth filters of order n is given by

$$A(\omega) = \frac{1}{\sqrt{1 + (\omega/\omega_c)^{2n}}}, \quad n = 1, 2, \dots \quad (9)$$

Denote the Chebyshev polynomial of the first kind by $C_n(\omega)$. Then,

$$C_n(\omega) = \begin{cases} \cos\left(n \cos^{-1} \frac{\omega}{\omega_c}\right), & |\omega| \leq \omega_c, \\ \text{ch}\left(n \text{ch}^{-1} \frac{\omega}{\omega_c}\right), & |\omega| > \omega_c, \end{cases} \quad n = 1, 2, \dots \quad (10)$$

The frequency response of the Chebyshev type lowpass filters for $\varepsilon > 0$ is given by

$$A(\omega) = \frac{1}{\sqrt{1 + \varepsilon^2 C_n^2(\omega)}}, \quad n = 1, 2, \dots \quad (11)$$

Denote the Chebyshev rational function of degree n by $R_n(\omega)$. Then,

$$R_n(\omega) = C_n\left(\frac{\omega - 1}{\omega + 1}\right), \quad n = 1, 2, \dots \quad (12)$$

One of the applications of $R_n(\omega)$ is to design an elliptic filter, which is also known as a Cauer filter, named after Wilhelm Cauer. An elliptic filter has the property of equalized ripple (equiripple) behavior in both the passband and the stopband. The frequency response of the elliptic type lowpass filters for $\varepsilon > 0$ is given by

$$A(\omega) = \frac{1}{\sqrt{1 + \varepsilon^2 R_n^2(\xi, \omega/\omega_c)}}, \quad n = 1, 2, \dots, \quad (13)$$

where ε is the ripple factor, and ξ is the selectivity factor [15, 16].

3. Problem Statement

The Butterworth filters obviously correspond to linear differential equations of integer order [17, 18].

Note that the Chebyshev polynomial of the first kind is the solution to the Chebyshev equation that is the second-order linear differential equation given by

$$(1 - x^2) \frac{d^2 y}{dx^2} - x \frac{dy}{dx} + n^2 y = 0. \quad (14)$$

Therefore, a consequence we note hereby is that the Chebyshev type filters as well as the elliptic type filters are corresponding to linear differential equations of integer order.

Recently, filters of fractional-order attract much attention in the field of circuits, systems, and signals; see, for example, Podlubny [19], Ortigueira [20], MacHado et al. [21], Lim et al. [22], Chen and Moore [23], and Zhang [24], simply citing a few. However, the literature regarding approximating ideal filters from a view of filters of fractional-order is rarely seen. For that reason, we propose a question like this. May ideal filters be approximated by filters or equations of fractional-order? We will give the affirmative answer to it in the next section.

4. Approximating Ideal Filters by Systems of Fractional Order

A linear filter can be expressed by a linear differential equation given by

$$\sum_{i=0}^p a_i \frac{d^{p-i} y(t)}{dt^{p-i}} = \sum_{i=0}^q b_i \frac{d^{q-i} e(t)}{dt^{q-i}}, \quad (15)$$

where $y(t)$ is the response and $e(t)$ excitation. Denote the Fourier transforms of $y(t)$ and $e(t)$ by $Y(\omega)$ and $E(\omega)$, respectively. Then, the system function is given by

$$H(\omega) = \frac{Y(\omega)}{E(\omega)}. \quad (16)$$

Denote $|H(\omega)|$ by $A(\omega)$. Then, $H(\omega) = A(\omega)e^{-j\vartheta(\omega)}$. This is the basic principle regarding linear filters. In this case, we say that $H(\omega)$ is the system function or frequency response of a filter of integer order; see, for example, Monje et al. [25].

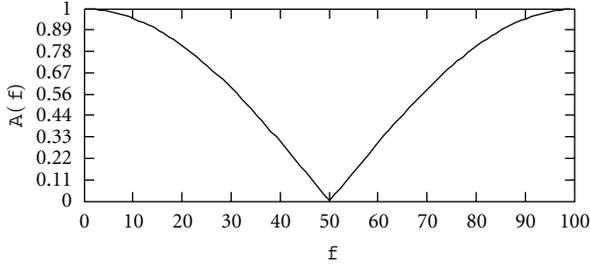


FIGURE 1: Amplitude response of filter.

We now consider a filter of fractional-order presented by

$$H_1(\omega) = [H(\omega)]^r = [A(\omega)]^r e^{-jr\vartheta(\omega)} \quad (r > 0). \quad (17)$$

Denote

$$H_1(\omega) = A_1(\omega)e^{-j\vartheta_1(\omega)}. \quad (18)$$

Then,

$$A_1(\omega) = [A(\omega)]^r, \quad \vartheta_1(\omega) = r\vartheta(\omega). \quad (19)$$

Since $\vartheta_1(\omega)$ is similar to $\vartheta(\omega)$, the key difference between $H(\omega)$ and $H_1(\omega)$ is in the aspect of amplitude response, namely, $A(\omega)$ and $A_1(\omega)$.

It can be seen from (19) that

$$\lim_{r \rightarrow 0} A_1(\omega) = \lim_{r \rightarrow 0} [A(\omega)]^r = 1 \quad \text{for } A(\omega) \neq 0. \quad (20)$$

In addition,

$$\lim_{r \rightarrow 0} A_1(\omega_c) = \lim_{r \rightarrow 0} [A(\omega_c)]^r = 0 \quad \text{if } A(\omega_c) = 0. \quad (21)$$

Denote $B_{0.7}$ the 3-dB bandwidth of $H_1(\omega)$ by

$$A_1(f) \big|_{f=B_{0.7}} = 0.707, \quad (22)$$

where $f = \omega/2\pi$ is frequency. Denote $B_{0.1}$ the bandwidth for

$$A_1(f) \big|_{f=B_{0.1}} = 0.1. \quad (23)$$

Then, the rectangular coefficient defined by

$$\text{Rec} = \frac{B_{0.7}}{B_{0.1}} \quad (24)$$

is always ideal for $A_1(\omega)$. That is,

$$\text{Rec} = \frac{B_{0.7}}{B_{0.1}} = 1 \quad \text{for } A_1(\omega), \quad (25)$$

because of (20).

On the other hand,

$$\lim_{r \rightarrow 0} \vartheta_1(\omega) = 0. \quad (26)$$

The expression (26) implies that $H_1(\omega)$ always has a linear phase response.

Remark 1. Equation (25) does not relate to any concrete forms of $H_1(\omega)$. Thus, the present results, namely, (20) and (21), stand for a general approach for approximating ideal filters based on systems of fractional-order.

Remark 2. Let

$$H_0(\omega) = \lim_{r \rightarrow 0} [H(\omega)]^r, \quad A_0(\omega) = \lim_{r \rightarrow 0} [A(\omega)]^r. \quad (27)$$

Then, $A_0(\omega)$ does not satisfy the Paley and Wiener criterion expressed by (8) because

$$A_0(\omega) = \begin{cases} 1 & \text{if } A_0(\omega) \neq 0, \\ 0 & \text{if } A_0(\omega) = 0. \end{cases} \quad (28)$$

That is,

$$\int_{-\infty}^{\infty} \frac{|\ln A_0(\omega)|}{1 + \omega^2} d\omega = \infty. \quad (29)$$

Therefore, this remark suggests a theoretical significance that the Paley and Wiener criterion might not be a necessary condition for designing physically realizable ideal filters of fractional-order.

5. Case Study

We consider a finite impulse response filter (FIR) given by

$$\begin{aligned} H(f) &= \frac{1 + \cos(2\pi fT) - j \sin(2\pi fT)}{2} \\ &= \frac{1}{2} [1 + \exp(-j2\pi fT)], \quad j = \sqrt{-1}, \end{aligned} \quad (30)$$

where T is the sampling period. Figure 1 indicates $A(f)$ for $T = 0.01$.

For $A(f) = |H(f)|$ and $T = 0.01$, we have

$$A(f) \big|_{f=50} = 0. \quad (31)$$

Note that

$$A(f) \big|_{f=25} = 0.707, \quad (32)$$

$$A(f) \big|_{f=46.8} = 0.1,$$

Thus, the rectangular coefficient of $H(f)$ is

$$\frac{B_{0.7}}{B_{0.1}} = \frac{25}{46.8} = 0.534. \quad (33)$$

The rectangular coefficient of 0.534 exhibits that $H(f)$ is not a satisfactory filter in general. Nevertheless, one is able to easily modify it to be such that it is an ideal filter by

$$\lim_{r \rightarrow 0} [H(f)]^r = \lim_{r \rightarrow 0} \left[\frac{1 + \exp(-j2\pi fT)}{2} \right]^r = \begin{cases} 1, & f < 50, \\ 0, & f = 50. \end{cases} \quad (34)$$

Figure 2 shows the approximations of $[H(f)]^r$ for $r = 0.1, 0.01, 0.001, \text{ and } 0.0001$, respectively. It exhibits that the

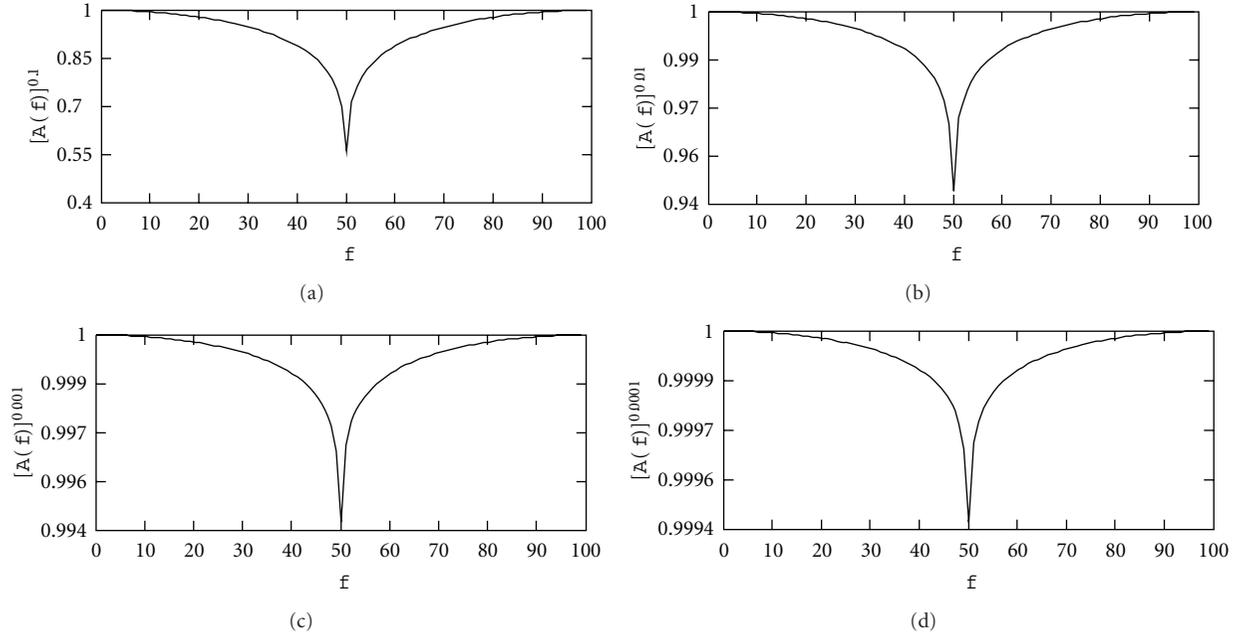


FIGURE 2: Approximations of $[H(f)]^r$ for different values of r . (a) $r = 0.1$. (b) $r = 0.01$. (c) $r = 0.001$. (d) $r = 0.0001$.

present method well approximates the ideal filter. As a matter of fact, in the sense of $0.9994 \approx 1 [H(f)]^r$ for $r = 0.0001$, see Figure 2(d), can be regarded as an ideal filter in practice.

The following is called a binomial series:

$$(1+x) = \sum_{k=0}^{\infty} \binom{r}{k} x^k \quad \text{for } |x| < 1, \quad (35)$$

where $\binom{r}{k} = \frac{r(r-1)\dots(r-k+1)}{k!}$ is binomial coefficient [26]. By using binomial series, (34) can be expanded by

$$\begin{aligned} \lim_{r \rightarrow 0} [H(f)]^r &= \lim_{r \rightarrow 0} \left[\frac{1 + \exp(-j2\pi f T)}{2} \right]^r \\ &= \frac{1}{2} \lim_{r \rightarrow 0} \sum_{k=0}^{\infty} \frac{\binom{r}{k}}{\binom{r}{0}} \exp(-jk2\pi f T). \end{aligned} \quad (36)$$

Therefore, in general, $[H(f)]^r$ should be taken as a filter of fractional-order from a view of fractional-order systems [25]; see the Appendix for the meaning of $[H(f)]^r$ in fractional-order systems.

It is worth noting that (34) may yet be an ideal FIR notch filter used for suppressing 50-Hz interference in electrocardiogram (ECG) signals, which is a key component in processing ECG signals in medical engineering; see, for example, Talmaon [27], Levkov et al. [28], Martens et al. [29], Dotsinsky and Stoyanov [30], and Li [31], though $H(f)$ is not a satisfactory filter for this purpose. Finally, it is noted that the research though reflected in this paper might be used for studying other topics, such as those in [32–35].

6. Conclusions

We have presented a general approach for approximating ideal filters from a view of fractional-order systems. This

approach is based on fractional calculus. The theoretical significance of the present approach is that the Paley and Wiener criterion might be no longer a necessary condition for designing physically realizable ideal filters. We have showed a case that can be used for designing ideal filters for suppressing 50-Hz interference in ECG signals.

Appendix

The fractional derivative of Caputo type of a function $f(t)$ is defined by

$$D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t \frac{f^{(n)}(u) du}{(t-u)^{\alpha-n+1}}, \quad n-1 \leq \alpha \leq n, \quad (A.1)$$

where Γ is the Gamma function [19]. For simplicity, we write ${}_0D_t^\alpha$ by D^α . Without generality losing, we take a system of second-order as a case:

$$\left(\frac{d^2}{dt^2} + \omega_0^2 \right) x(t) = e(t), \quad \omega_0 > 0. \quad (A.2)$$

There are two types of fractional-order systems based on (A.2). One is given by (see [36])

$$\left(\frac{d^{2-\varepsilon}}{dt^{2-\varepsilon}} + \omega_0^2 \right) x(t) = e(t), \quad 0 < \varepsilon < 1. \quad (A.3)$$

Another is expressed by

$$\left(\frac{d^2}{dt^2} + \omega_0^2 \right)^\beta x(t) = e(t), \quad \beta > 0; \quad (A.4)$$

see, for example, [37, 38].

Denote the impulse response function of (A.5) by $h(t)$. Then,

$$(D^2 + \omega_0^2)^\beta h(t) = \delta(t), \quad (\text{A.5})$$

where $\delta(t)$ is the Dirac- δ function.

Denote the Fourier transform of $h(t)$ by $H(\omega)$. Then, we have

$$H(\omega) = \frac{1}{(\omega^2 + \omega_0^2)^\beta}. \quad (\text{A.6})$$

Therefore, if one denotes the frequency response of (A.2) by $H_0(\omega)$, which is a system of 2-order,

$$H_0(\omega) = \frac{1}{(\omega^2 + \omega_0^2)}, \quad (\text{A.7})$$

then,

$$H(\omega) = [H_0(\omega)]^\beta. \quad (\text{A.8})$$

The expression (A.8) gives the explanation of $[H(f)]^r$ in fractional-order systems discussed in the body text of the paper.

Acknowledgments

This work was supported in part by the 973 Plan under the Project no. 2011CB302802, by the National Natural Science Foundation of China under the Project Grant nos. 61070214 and 60873264.

References

- [1] Z. M. Hussain, A. Z. Sadik, and P. O'Shea, *Digital Signal Processing: An Introduction with MATLAB and Applications*, Springer, New York, NY, USA, 2011.
- [2] S. S. Bhattacharyya, R. Leupers, and J. Takala, Eds., *Handbook of Signal Processing Systems*, Springer, New York, NY, USA, 2010.
- [3] P. Fieguth, *Statistical Image Processing and Multidimensional Modeling*, Springer, New York, NY, USA, 2011.
- [4] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedure*, John Wiley & Sons, City, State, USA, 3rd edition, 2000.
- [5] R. M. Gray and L. D. Davisson, *Introduction to Statistical Signal Processing*, Cambridge University Press, Cambridge, Mass, USA, 2004.
- [6] M. Li, "Generation of teletraffic of generalized Cauchy type," *Physica Scripta*, vol. 81, no. 2, Article ID 025007, 2010.
- [7] L. Wanhammar, *Analog Filters Using MATLAB*, Springer, New York, NY, USA, 2009.
- [8] S. K. Mitra and J. F. Kaiser, *Handbook for Digital Signal Processing*, John Wiley & Sons, New York, NY, USA, 1993.
- [9] J. van de Vegte, *Fundamentals of Digital Signal Processing*, Prentice Hall, New York, NY, USA, 2003.
- [10] R. C. Dorf and R. H. Bishop, *Modern Control Systems*, Prentice Hall, New York, NY, USA, 9th edition, 2002.
- [11] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [12] A. Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, New York, NY, USA, 1965.
- [13] R. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, American Mathematical Society Colloquium Publication, New York, NY, USA, 19th edition, 1934.
- [14] H. Y.-F. Lam, *Analog and Digital Filters: Design and Realization*, Prentice-Hall, New York, NY, USA, 1979.
- [15] R. W. Daniels, *Approximation Methods for Electronic Filter Design*, McGraw-Hill, New York, NY, USA, 1974.
- [16] M. D. Lutovac, D. V. Tomic, and B. L. Evans, *Filter Design for Signal Processing Using MATLAB and Mathematica*, Prentice Hall, Upper Saddle River, NJ, USA, 2001.
- [17] G. Bianchi and R. Sorrentino, *Electronic Filter Simulation & Design*, McGraw-Hill Professional, New York, NY, USA, 2007.
- [18] S. Butterworth, "On the theory of filter amplifiers," *Wireless Engineer*, vol. 7, pp. 536–541, 1930.
- [19] I. Podlubny, *Fractional Differential Equations*, Academic Press, San Diego, Calif, USA, 1999.
- [20] M. D. Ortigueira, "An introduction to the fractional continuous-time linear systems: the 21st century systems," *IEEE Circuits and Systems Magazine*, vol. 8, no. 3, Article ID 4609961, pp. 19–26, 2008.
- [21] J. A. Tenreiro MacHado, M. F. Silva, R. S. Barbosa et al., "Some applications of fractional calculus in engineering," *Mathematical Problems in Engineering*, vol. 2010, Article ID 639801, 34 pages, 2010.
- [22] S. C. Lim, M. Li, and L. P. Teo, "Langevin equation with two fractional orders," *Physics Letters A*, vol. 372, no. 42, pp. 6309–6320, 2008.
- [23] Y. Q. Chen and K. L. Moore, "Discretization schemes for fractional-order differentiators and integrators," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 3, pp. 363–367, 2002.
- [24] X. Zhang, "Maxflat fractional delay IIR filter design," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 2950–2956, 2009.
- [25] C. A. Monje, Y.-Q. Chen, B. M. Vinagre, D. Xue, and V. Feliu, *Fractional Order Systems and Controls—Fundamentals and Applications*, Springer, New York, NY, USA, 2010.
- [26] G. Arfken, *Mathematical Methods for Physicists*, Academic Press, Orlando, Fla, USA, 3rd edition, 1985.
- [27] J. L. Talmaon, *Pattern recognition of the ECG: a structured analysis*, Ph.D. thesis, 1983.
- [28] C. Levkov, G. Michov, R. Ivanov, and I. K. Daskalov, "Subtraction of 50 Hz interference from the electrocardiogram," *Medical and Biological Engineering and Computing*, vol. 22, no. 4, pp. 371–373, 1984.
- [29] S. M. M. Martens, M. Mischi, S. G. Oei, and J. W. M. Bergmans, "An improved adaptive power line interference canceller for electrocardiography," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2220–2231, 2006.
- [30] I. Dotsinsky and T. Stoyanov, "Power-line interference cancellation in ECG signals," *Biomedical Instrumentation and Technology*, vol. 39, no. 2, pp. 155–162, 2005.
- [31] M. Li, "Comparative study of IIR notch filters for suppressing 60 Hz interference in electrocardiogram signals," *International Journal of Electronics and Computers*, vol. 1, no. 1, pp. 7–18, 2009.
- [32] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3–4, pp. 207–217, 2010.
- [33] C. Cattani and G. Pierro, "Complexity on acute myeloid leukemia mRNA transcript variant," *Mathematical Problems in Engineering*, vol. 2011, Article ID 379873, 16 pages, 2011.

- [34] E. G. Bakhoun and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.
- [35] E. G. Bakhoun and C. Toma, "Dynamical aspects of macroscopic and quantum transitions due to coherence function and time series events," *Mathematical Problems in Engineering*, vol. 2010, Article ID 428903, 13 pages, 2010.
- [36] B. N. N. Achar, J. W. Hanneken, and T. Clarke, "Damping characteristics of a fractional oscillator," *Physica A*, vol. 339, no. 3-4, pp. 311-319, 2004.
- [37] S. C. Lim and S. V. Muniandy, "Self-similar Gaussian processes for modeling anomalous diffusion," *Physical Review E*, vol. 66, no. 2, Article ID 021114, pp. 021114-1-021114-14, 2002.
- [38] M. Li, S. C. Lim, and S. Y. Chen, "Exact solution of impulse response to a class of fractional oscillators and its stability," *Mathematical Problems in Engineering*, vol. 2011, Article ID 657839, 9 pages, 2011.

Research Article

High-Resolution Remotely Sensed Small Target Detection by Imitating Fly Visual Perception Mechanism

Fengchen Huang, Lizhong Xu, Min Li, and Min Tang

College of Computer and Information Engineering, Hohai University, Nanjing 210098, China

Correspondence should be addressed to Lizhong Xu, lzhu@hhu.edu.cn

Received 25 July 2011; Accepted 10 October 2011

Academic Editor: Carlo Cattani

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

The difficulty and limitation of small target detection methods for high-resolution remote sensing data have been a recent research hot spot. Inspired by the information capture and processing theory of fly visual system, this paper endeavors to construct a characterized model of information perception and make use of the advantages of fast and accurate small target detection under complex varied nature environment. The proposed model forms a theoretical basis of small target detection for high-resolution remote sensing data. After the comparison of prevailing simulation mechanism behind fly visual systems, we propose a fly-imitated visual system method of information processing for high-resolution remote sensing data. A small target detector and corresponding detection algorithm are designed by simulating the mechanism of information acquisition, compression, and fusion of fly visual system and the function of pool cell and the character of nonlinear self-adaption. Experiments verify the feasibility and rationality of the proposed small target detection model and fly-imitated visual perception method.

1. Introduction

With the fast development of sensor technique, the high spatial/spectral resolution remote sensing systems are more and more deployed. For example, the spatial resolution of American space remote sensing platform QuickBird, Worldview has achieved meter level or submeter level of resolution. The spatial resolution of military reconnaissance satellite KH-12 and 8X-1 has achieved a resolution of 0.1 meters level. The American space AVIRIS spectrum resolution amounts to 10 nanometers while the Trwis-3 spectrum resolution is even higher, amounting to 6 nanometers, and the band number surpasses 380. The China's 12th five-year plan endeavors to push the development and implementation of the observation project of the high resolution to the ground level (implemented in the 11th five-year plan) and concentrates on the emerging research focus on the basic theories and key techniques of the high-resolution remote sensing target, space environment feature analysis, and high reliable automatic interpretation to meet the requirement of national security and massive demand of the socioeconomic development.

Differing from the observation for the larger-scale and high-resolution remote sensing at ground level, the present small target detection technology has faced huge challenges due to increasing types of recognizable ground object, high density, complex target detection environment caused by synonyms spectrum of same object (or the same spectrum of different object) and scarce target prior information.

Remote sensed small targets in high resolution occupy little pixel areas while they contain rich detail texture features (such as airplane, suspicious items, and military installations). In contrast the precise and abnormal detection of hyperspectral remotely sensed materialized parameters barely has texture and shape structure information. Many researchers and institutes have made extensive efforts on this topic. The methods can be divided into three types: subspace-based methods, machine learning-based methods, and object-oriented methods.

The subspace transformation-based target detection method is a mathematical method based on image spectrum analysis. Hyvärinen and Oja proposed independent component analysis method in 2000 [1], and Chang [2] proposed target detection algorithm based on orthogonal

transformation in 2005, both of which are based on subspace transformation. However, because this kind of linear space transformation method is not good for the abstraction of the high-dimensional feature information in image background, the sub-space method based on Kernel function is proposed. For example, Capobianco et al. proposed orthogonal subspace mapping method based on Kernel function [3] and principal component analysis based on Kernel function, and so forth in 2009. Li and Yu adopted low-dimensional hyperplane structure to realize hyperspectral remote sensing image of anomaly target detection with better abstraction effect being achieved [4]. However because this type of method is confined with the problems of the selection of the Kernel function and the parameters estimation, it has a certain limitation.

Based on the machine learning method, the nonlinear no parameters' estimation of the background is achieved by abstracting background spectral high-dimensional statistical features according to the limited samples learning. Bruzzone and Carlin [5] in 2006 and Camps-Valls et al. [6] in 2010 proposed high-resolution remote sensing image target extraction method based on support vector machine, respectively. In 2009, aimed to the complex changeable feature of hyperspectral remote sensing image background, Mei et al. [7] proposed target detection method based on adaptive support vector machine. At present, this kind of method needs to be researched further because of the limitation of the selection of the sample data and the amount of the training samples.

Object-oriented target recognition and detection method is a kind of feature-level target detection method, which is paid much attention for expressing and applying semantic information features. In 2001, Blaschke and Strobl [8] first proposed object-oriented remote sensing ground object classification method, which is used by small target detection of high-resolution remote sensing image. In 2009, Sirmacek and Unsalan [9] proposed the method combining scale-invariant feature transform with graph theory. After that, Sirmacek and Unsalan [10] further put forward object-oriented detection method for building target in IKONOS image, but it has a certain difficulty in abstracting effective feature and combining reasonably to achieve the exact description of the targets for high-resolution remote sensing image of complex texture feature and rich details when it lacks prior information. Besides, Di et al. come up with the target detection method based on fuzzy integral [11].

In summary, the aforementioned methods have some limitations, and even they are in the face of great difficulties. In addition, vast amounts of image data also increase the computational cost of detection algorithm, which is not conducive to real-time requirements.

In recent years, our research group has been working on the bionic compound eye information processing and visual detection. In 2008 [12], for multiremote sensing platform monitoring, bionic compound eye information fusion system model and computational method was proposed. In 2009 [13], inspired by the fly compound eye the sequence image superresolution reconstruction and integration method is proposed. In 2010 [14], inspired by

the fly the small-target detector algorithm in complex background is put forward.

The fly compound eyes composed of many small closely spaced eyes in the nature are taken into account. Flies can rely on the visual system with low-resolution and very limited computing power to accurately detect and track targets while flying in complex natural scenes with high speed. However this is still a big challenge for human who have mastered high-resolution imaging technology and the high-end data processing technology.

In order to provide a theoretical basis for designing the "low-order" target detector to percept the complex, "high-order" remote sensing image, this paper builds up a fly-imitated visual perception information expression model inspired by the fly visual system information acquisition and processing mechanisms, which is completely different from previous research strategy from a new perspective. The advantages of fly visual system, small target detection, and identification of high-resolution optical remote sensing make it unnecessary of precise modeling and priori information binding.

Aiming to solve the limitations of high-resolution remotely sensed small target detection methods, fly imitated visual information processing system pattern of high-resolution remotely sensed small target detection is built through the engineering simulation of perceptual mechanism of the fly visual system. Based on this pattern, fly visual information acquisition, compression, and integration mechanisms are simulated and the information expression of the fly-imitated visual perception based on "cartridges" is analyzed. Moreover, the fly visual "pool cell" function and nonlinear self-adaptations of neurons arrays are simulated, and then the small target detector and fly-imitated visual perception algorithm are proposed and designed.

At present, it is lack of the comparative study about fly visual neurons arrays, biological mechanism of pool cell, and engineering simulation of fly internal mechanism. The theory behind small target detector and algorithm design is also limited. Therefore, it is necessary to carry out the work on fly visual perceptual information expressing modeling by engineering simulation and the internal mechanism of the pool cell to support target detector design and fly visual neurons arrays and improve detection algorithm eventually.

2. The Basic Structure of Fly Visual System

Each compound eyes of flies are composed of about 3000–3200 small eyes, and each small eye is self-contained formed by the imaging system with cornea and crystalline dimension, light-sensitive retinal visual cells, and the optic nerve leading to the brain. Therefore each individual small eye can see things. The compound eye of fly is shown in Figure 1. The neural pathway sections from the primary visual information to advanced brain information processed from fly visual nervous system are shown in Figure 2.

Retina, lamina, medulla, lobular, lobular plate, central brain, and other tissues are mainly distributed in the neural pathway. The key of the fly visual system to achieve target



FIGURE 1: Fly compound eye.

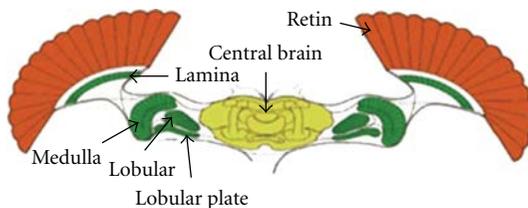


FIGURE 2: Section of fly visual channel.

detection is that the neural pathways formed by these organizations have highly nonlinear filtering properties.

The photoreceptor cells mainly complete information acquisition of the entire visual nervous system before the lamina and retina form the primary visual system processing stage of the compound eye. The cartridge structure compresses and fuses the obtained information and then spreads the information to the next high-order neural unit for processing by the LMC cells, which is composed by photoreceptor cells wrapped together with the LMC cell on lamina. LPTC, LGMD, STMD, and other high-order neuronal cell populations and pool cell populations are distributed in lobular, lobular plate, and brain. These nerve cells through the pool cell scheduling achieve the target perception and detection by background inhibition and target enhancement.

3. Fly-Imitated Visual Detection and Information Processing System Pattern

The fly-imitated visual detection and information processing system are shown in Figure 3. This system is built through comparative analysis among the fly visual system structure, mechanism, and the engineering simulation of internal mechanism in fly system. In addition, the system uses the following as the reference: the information processing mechanism of the visual neural pathways from the compound eyes retina imaging to brain center determining information, engineering simulated fly visual system mechanisms of acquisition, representation, and processing for natural

scenes information. The system model mainly includes three parts: cartridge information expression, pool cell shunt and inhibition, and small-target detector [15].

The “cartridge information representation” section of fly visual system is different from the traditional “dynamic equation” analysis and accurate modeling. The so-called fly-imitated perceptual information expression models attempt to simulate the cartridge system for modeling, combining with background characteristics, targets and environmental characteristics, and performance requirements of target detection. The limited resolution of flies’ compound eyes and the very limited computing resources are used for reference, which provide a theoretical basis for designing the “low-order” target detector to perceive complex, “high-order” remote sensing images. By simulating biological cartridge mechanism, thousands of small eye information can be fused, the information processing and computing transition from compound eye optical imaging to high-order neural array can be achieved, and the full-color data of high-resolution remote sensing image and multispectral data can be fused.

“Cartridge” information processing procedure mainly includes the following. (1) Information fusion: the total potential V of the cartridge is exponential adjusted by the difference between the membrane potential v_{k-1} of the single photoreceptor cell and the average membrane potential v_{mean} of the all photoreceptor cells in the cartridge. (2) Information compression: the information transmission way of ion diffusion is produced by the concentration difference of charged ions between cartridge structure and second-order neural LMC cells, and the information compression and feature extraction are achieved through the offset and confrontation of the luminance information within the local area.

“Pool cell scheduling” which is based on the characteristic information input achieves adaptive enhancement and inhibition effects of the feature by shunt inhibition to the two unipolar pool cells and one pair of bipolar pool cells in two sides of the compound eyes. The engineering simulation of the pool cell mechanisms is implemented by two steps. (1) Shunt inhibition: for different contrast polarity features channels, the fusion results of the features within the local area are used as inhibit components in the shunt inhibition procedure, which is good for the inhibition of the background texture feature. (2) Feature extraction: for the visible spectrum of the static image, the remote sensing image data coding method with comprehensive airspace and “map-in-one” spectral information is built to extract the characteristic information of remote sensing images.

“Small target detector” achieves target detection based on the biological mechanism of LPTC, LGMD, STMD, and other fly visual system target detection which refers to nonlinear adaptive filtering process of high-order neural arrays, the correlation of opposite polarity, and so on. It mainly includes the following. (1) Nonlinear adaptive filtering mechanism: it simulates the biological mechanism of fast polarization and slow depolarization of high-order neural arrays, which can enhance the signals with low frequent and large magnitude of changes, and adaptively suppresses the

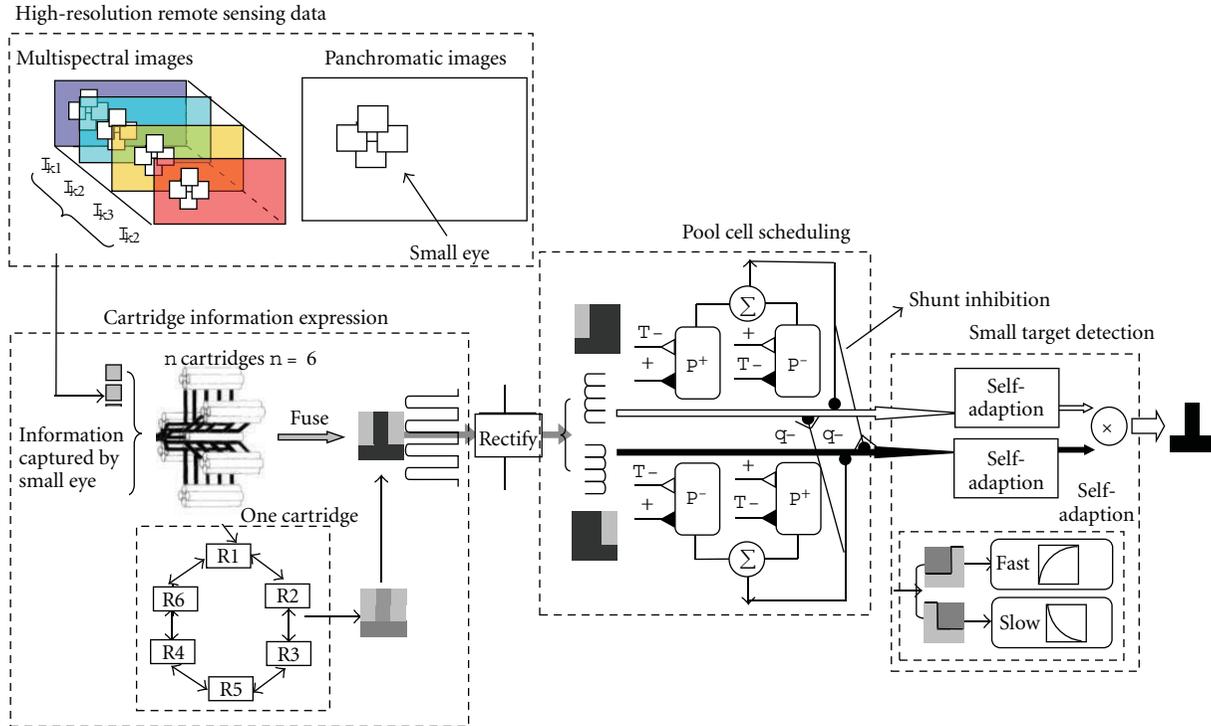


FIGURE 3: Fly-imitated visual detection and information processing system.

texture signals with high frequent and small magnitude of changes. (2) Correlation of opposite polarity mechanisms: the correlation computation the edge features with opposite polarity in different spatial locations is used to obtain the effect of nonlinear adaptive target extraction without the high-accuracy description of target feature.

4. Fly Imitated Visual Perception Calculation of Small Target Detection

4.1. The Design of Virtual Compound Eye. The selection and design of virtual small eye should be set according to the requirements of fields applications. The satellite that equipped imaging equipment is regarded as a compound eye imaging system, while the panchromatic images and the multispectral images obtained can be mapped to “small eye” images in a logical concept.

For example, QuickBird is used as the research object to design the virtual small eyes: panchromatic and multispectral remote sensors equipped on QuickBird satellite use push-broom imaging method, which can get one panchromatic band and four multispectral bands. The spatial resolution of panchromatic band is 0.61 meters and for multispectral band it is 2.44 meters while its spectral range is 450 nm to 900 nm. The local window containing only a small number of pixels is selected as single “small eye”. Five “small eyes” like the aforementioned ones are overlapping bundled, one of which is the center and the other four are overlapping with central “small eye” to form a sliding “small eyes group”. Multispectral bands use the sliding local window to extract the “small eye” images similar to panchromatic image.

Virtual “compound eye” system is divided into two types, one is composed of “small eye image” panchromatic image data, and another is composed of “small eye image” multi-spectral image data. In a certain phase, the virtual QuickBird compound eye system is composed of four groups of visible light “small eyes group” in low spatial resolution and one group of panchromatic “small eye group” in high spatial resolution.

4.2. Image Data Preprocessing. Because of the differences between panchromatic images and multi-spectral images in the spatial resolution, in fly-imitated compound eye system, remote sensing data always needs preprocessing, which requires different images registering each other; namely, it requires the image locations of the same ground object in different images overlapping. Many studies have proven that the phase-coherent model successfully explains the effectiveness of phase information sensed by biological vision and the stability for noise, brightness, and contrast changes. Therefore, the design is registered by phase coherence, which includes two steps: the first step is to obtain local energy function and each harmonic amplitude of spatio-temporal data by Gabor wavelet; the second step is to extract feature points to find the corresponding relationship between reference image and the registered image [16, 17].

4.3. The Biological Mechanism and Engineering Simulation of Small Target-Detected Neuron. In 1985, Egelhaaf [18] found that on the lobular plate of the fly visual system, there is a high-order neuron, small target motion detection (STMD), and pointed that the neuron has highly nonlinear

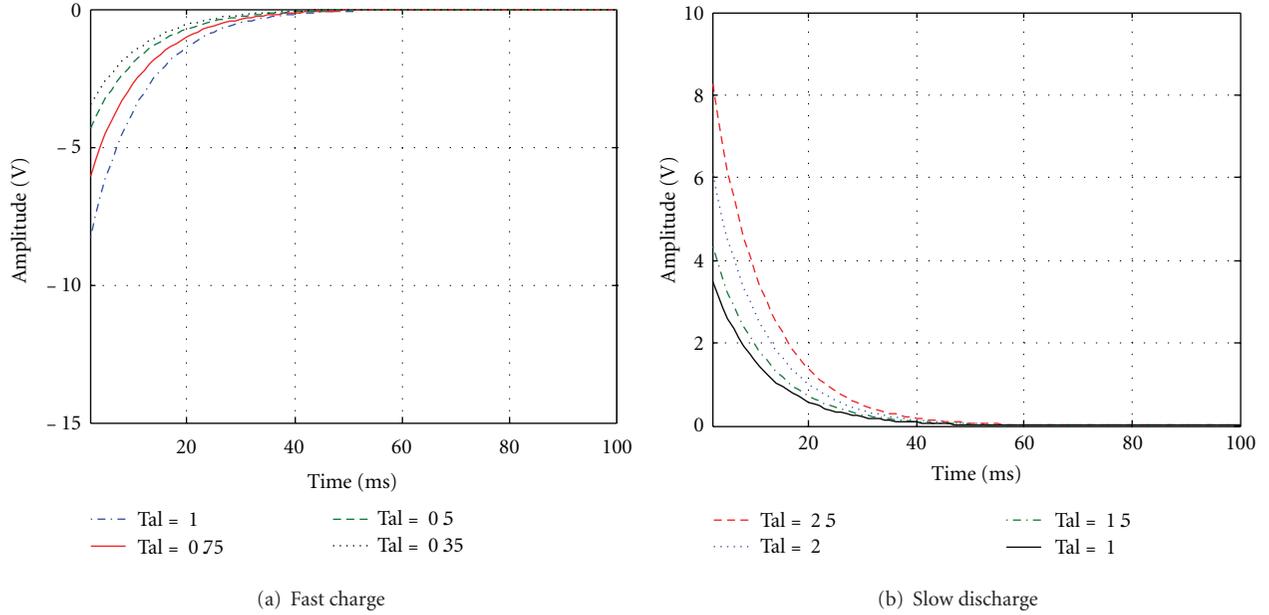


FIGURE 4: Nonlinear self-adaptive response characteristic circle.

filtering characteristics and high sensitivity for the mutant signals. In 2008, Wiederman et al. [19] built a small target detected neuron model according to the results of previous studies. The nonlinear adaptive mechanism, the central lateral inhibition mechanism, and the correlation of opposite polarity mechanism in the model can enhance the target feature while they can inhibit the background texture.

4.3.1. Small Target-Detected Neuron Algorithm

(1) *Nonlinear Adaptive Mechanism.* Nonlinear adaptive mechanism enhances the mutant signals with low frequency and large amplitude of changes, inhibiting texture information with high frequency and low amplitude of changes. Taking the horizontal direction, for example (similarly, the vertical direction), for the output signal $on_h(i, j)$, the discrete form of the adaptive mechanism is represented as

$$\begin{aligned} & \text{if } on_h(i, j) < on_h(m, n), \\ & on'_{on}(i, j) = on_h(i, j) - \exp\left(-\frac{\Delta s}{1}\right), \\ & \text{if } on_h(i, j) \geq on_h(m, n), \\ & on'_h(i, j) = on_h(i, j) + \exp\left(\frac{\Delta s}{2}\right), \end{aligned} \quad (1)$$

where $on_h(m, n)$ is the characteristic response intensity for the (i, j) pixel in the channel on, and Δs is the Euclidean distance between the two pixels which represents the interaction between the two pixels related to the distance. ξ_1 and ξ_2 are the response attenuation (enhancement) factor. Formula (1) shows that the signal intensity in position (i, j) is lower than that in the field around the signal, and the

response will decay at the speed of $\exp(\Delta s/1)$; otherwise, it will increase at the speed of $\exp(\Delta s/2)$, namely, the rapid depolarization and slow repolarization in biological neurons, and the engineering simulated results are shown in Figure 4.

(2) *Central Lateral Inhibition Mechanism.* Central lateral inhibition mechanism can enhance the contrast between the signals. After the non-linear adaptive processing, the characteristic information of background is suppressed, and the feature information of target is retained. At this time, the central lateral inhibition mechanism can be used to enhance the retained target features, so the false alarm rate of the test results is reduced. Taking the channel on for example, in the local area $N(i, j)$ with the field radius r (taking into account that the small target size in the horizontal or vertical direction is limited to 1 or 2 pixels, the radius of the local area is $1 \leq r \leq 2$), the output of the channel on in the pixel position (i, j) is

$$on''_h(i, j) = on'_h(i, j) - \sum_{m,n \in N(i,j)} w_{m,n} \times on'_h(m, n), \quad (2)$$

where weighting factor $w_{m,n}$ is

$$w_{m,n} = \frac{1}{\left(\sqrt{(i-m)^2 + (j-n)^2} + \varepsilon\right)}, \quad (3)$$

where ε is a constant.

Central lateral inhibition mechanism of off channel is similar to on channel.

(3) *Correlation of Opposite Polarity Channels.* Channel on and channel off of small target-detected neurons correspond to the two edges of small targets, respectively; by shifting the

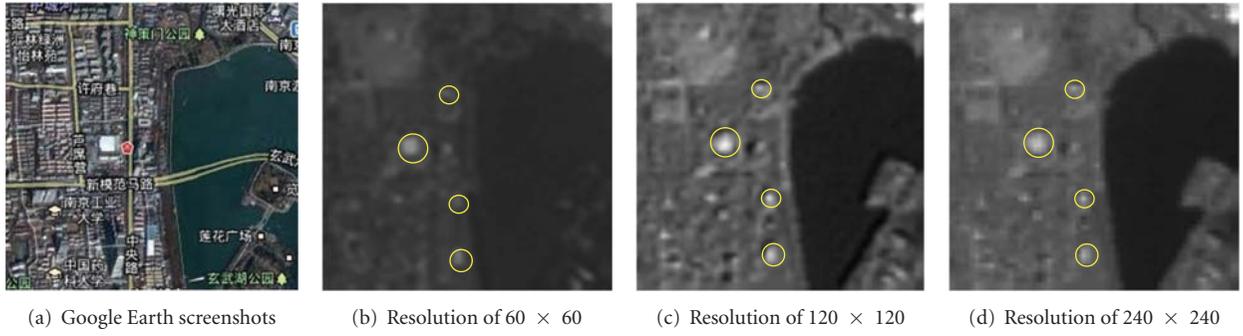


FIGURE 5: Experiment results of multispectral remote sensing image of Xuanwu district in Nanjing city.

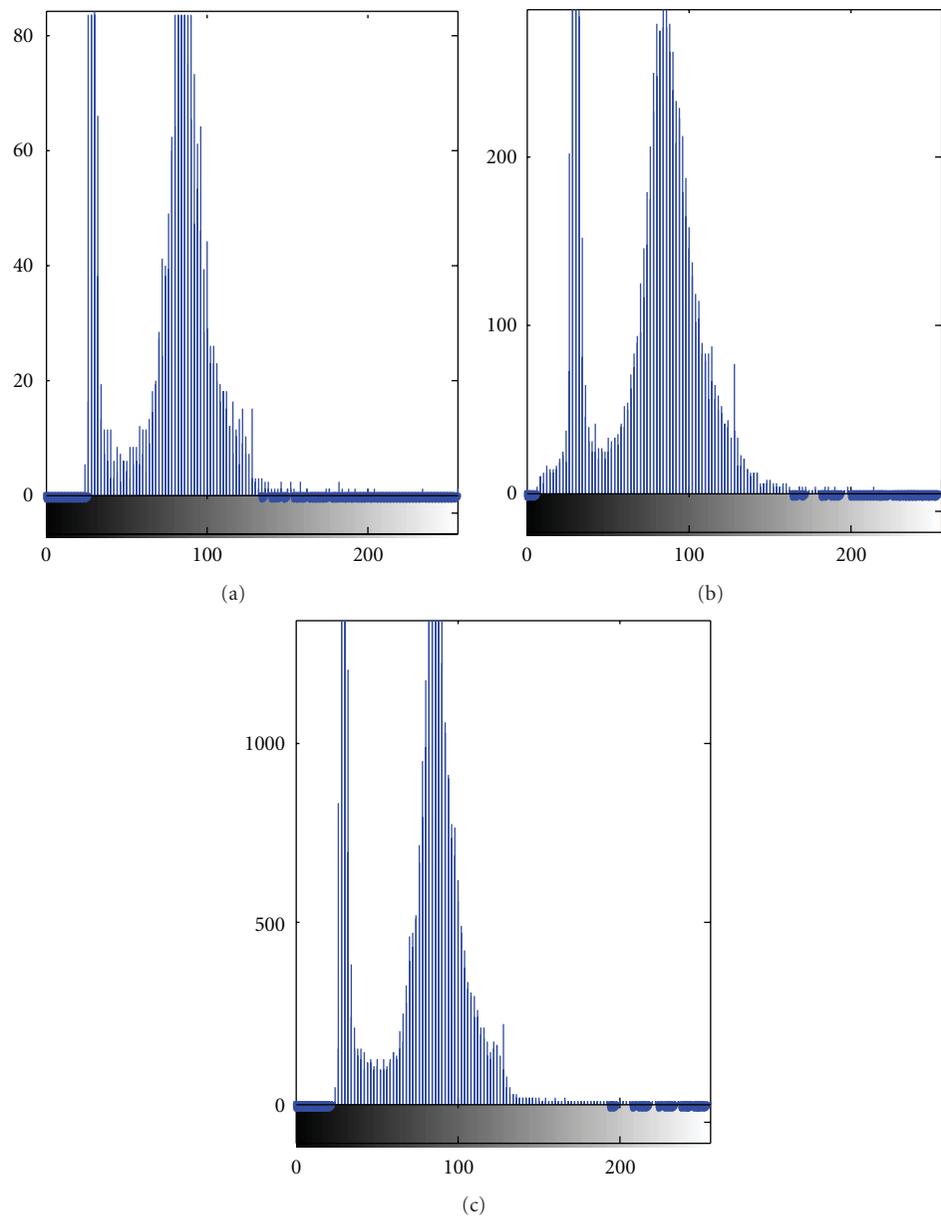


FIGURE 6: Background histogram under three different kinds of spatial resolution.

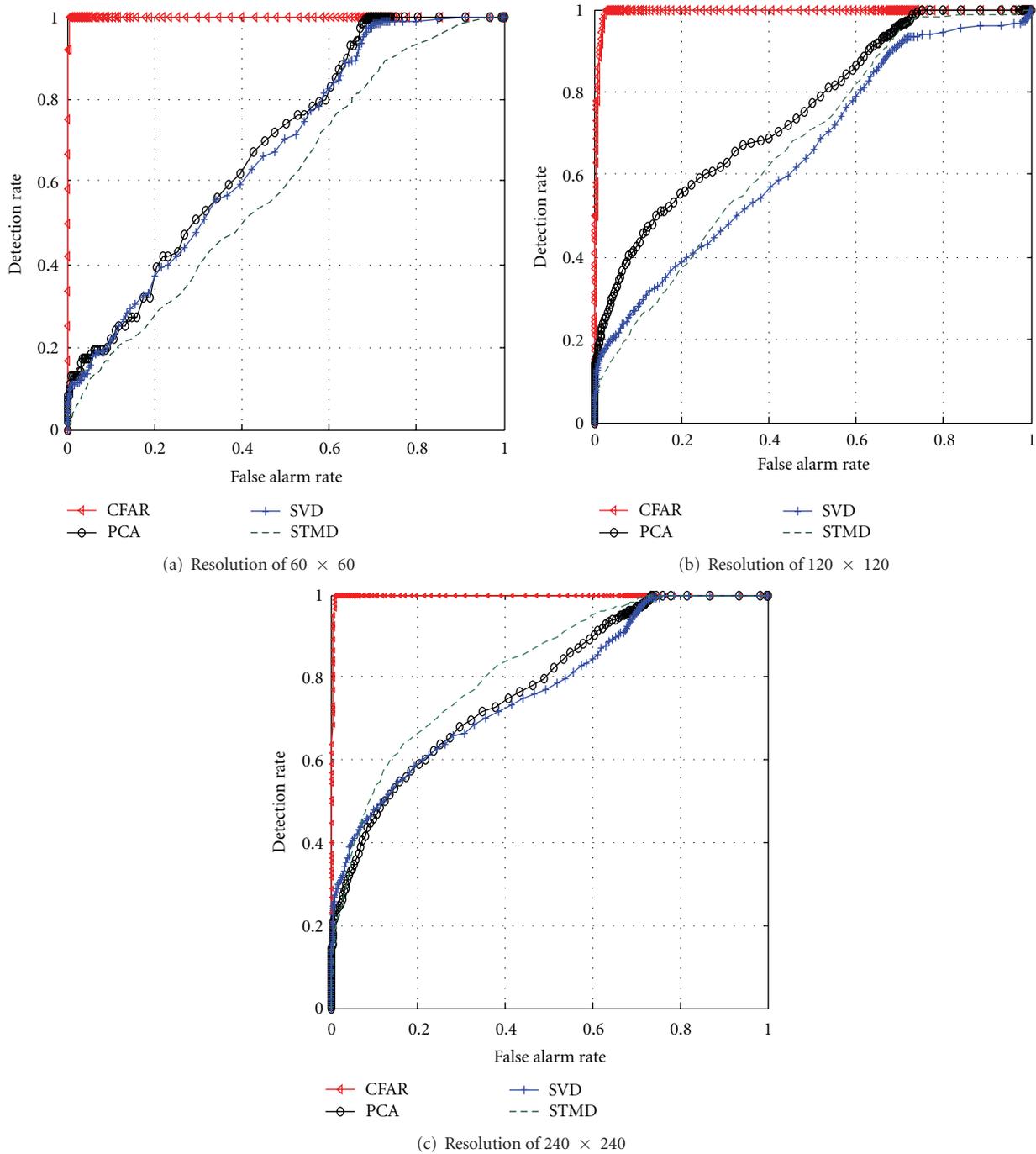


FIGURE 7: Comparison of false alarm rate of four detection algorithms under different spatial resolutions.

opposite polarity channel, the small target detection results can be obtained after the correlation of opposite polarity channels. According to the small target definition in SPIE [20], the two edges of the small target are separated by 1 or 2 pixels. For the targets of different types (light or dark targets), the channel polarity which is selected to shift is different. The small targets in multi-spectral remote sensing images often have the characteristics of high spectral intensity and low probability of emergence and can be approximated as the light target, namely, $I(i, j) > I(i + \Delta, j + \Delta)$. Thus, the

signal polarity in the horizontal direction from left to right is channel on (increased brightness) and channel off (decreased brightness). Then, the channel-correlation processing of the remote sensing image is

$$\text{output}(i, j) = \text{off}'_h(i, j) \times \text{shift}(\text{on}''_h(i, j), 2\Delta), \quad (4)$$

where $\text{shift}(\cdot)$ is the shifting function, Δ is the shifting amount, $\Delta > 0$ represents shifting right, and $\Delta < 0$ represents shifting left.

4.3.2. Small Target Detected Neuron Simulation and Analysis. The local area with the size of 60×60 on the 3-band of the multispectral remote sensing image in Xuanwu District, Nanjing, obtained by Landsat satellite is used for the research. The area is the city's commercial district. A large number of buildings are concrete structures having poor absorption capacity and high reflection capacity, which are the light targets in remote sensing image. Figure 5(a) is the corresponding region satellite images intercepted from the Google Earth as a reference for target location. Figures 5(b), 5(c), and 5(d) are remote sensing images with different spatial resolution, and the yellow circles mark the four preselected targets. Figure 6 is the spectral distribution histogram in three types of spatial resolutions. There are two peaks in the spectral distribution of the image: the single peak of brightness has less energy in the lower brightness region, and the single peak appearing in the high brightness region is similar to the Gaussian distribution model but contains more energy.

Figure 7 is the ROC curve under the conditions of four kinds of different spatial resolution detection algorithms. Green dashed line is the algorithm proposed by this paper, the red triangle line is CFAR algorithm, the black line \circ is PCA algorithm, and blue line + is SVD algorithm. Figures 7(a), 7(b), and 7(c) show comparison with the false alarm rates, respectively.

Spatial resolution CFAR algorithm always maintains a low level of false alarm rate, but SVD, PCA, and the proposed algorithms have almost similar false alarm rates. Combined with the spectral distribution histogram of the region, we can see that the background spectral distribution is similar to Gaussian distribution according to the theoretical basis of CFAR algorithm. So a lower false alarm rate of detection result can be obtained. As the spatial resolution increases, the false alarm rates of PCA, SVD, and the proposed algorithm decrease, where the false alarm rate of the proposed method declines fastest, although this algorithm shows that the false alarm rate is still higher than that of the CFAR algorithm. With the continuous improvement of spatial resolution images, background complexity increases, and the adaptive processing advantages will be fully reflected.

5. Conclusion

As the high precision and resolution remote sensing images in spatial and spectral observation scale result in complex background spectrum changes and target features diversity, the traditional large-scale-based target detection methods are difficult to transplant and apply. Current computer vision-based high-resolution remote sensing small target detection methods are either inhibiting complex background features from the background spectral analysis or describing and detecting targets by machine learning from the target features. Both aspects need to be improved in terms of the false alarm rate, real-time efficiency, robustness, and complexity of the algorithm.

Fly visual system has their unique advantages on small target detection in identification and tracking of natural

scenes. Inspired by the information acquisition and processing mechanism of fly compound eye, this paper proposes small target detection fly-imitated visual information processing system pattern of high-resolution remote sensing images. Based on this model, the small target detector and fly-imitated visual perception algorithm are designed, providing a new strategy to further solve the difficulties existing in the automatic interpretation of high-resolution remote sensing image.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (no. 60901003).

References

- [1] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, 2000.
- [2] C. I. Chang, "Orthogonal Subspace Projection (OSP) revisited: a comprehensive study and analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 502-518, 2005.
- [3] L. Capobianco, A. Garzelli, and G. Camps-Valls, "Target detection with semisupervised kernel orthogonal subspace projection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, Article ID 5159487, pp. 3822-3833, 2009.
- [4] Z. Y. Li and W. X. Yu, "Application of low-dimension hyperplane in anomaly detection of hyperspectral image," *Infrared and Laser Engineering*, vol. 38, no. 2, pp. 194-199, 2009.
- [5] L. Bruzzone and L. Carlini, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, Article ID 1677767, pp. 2587-2600, 2006.
- [6] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio-spectral remote sensing image classification with graph kernels," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, Article ID 5462854, pp. 741-745, 2010.
- [7] F. Mei, C. H. Zhao, L. G. Wang, and J. You, "Support vector data description based on adaptive anomaly detection method in hyperspectral imagery," *Acta Photonica Sinica*, vol. 38, no. 11, pp. 2820-2825, 2009.
- [8] T. Blaschke and J. Strobl, "What's wrong with pixels? Some recent developments interfacing remote sensing and GIS," *Geo-Information-Systeme*, vol. 14, no. 6, pp. 12-17, 2001.
- [9] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, Article ID 4799121, pp. 1156-1167, 2009.
- [10] B. Sirmacek and C. Unsalan, "A probabilistic framework to detect buildings in aerial and satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, Article ID 5523977, pp. 211-221, 2011.
- [11] W. Di, Q. Pan, Y. Q. Zhao, and L. He, "Anomaly target detection in hyperspectral imagery based on band subset fusion by fuzzy integral," *Journal of Electronics and Information Technology*, vol. 30, no. 2, pp. 267-271, 2008.
- [12] L. Z. Xu, A. Y. Shi, F. C. Huang et al., "A perceptual and computational mechanism for bionic compound eye systems with multi-source information fusion," *CAAI Transactions on Intelligent Systems*, vol. 3, no. 4, pp. 328-335, 2008.

- [13] M. Li, L. Z. Xu, A. Y. Shi et al., "Image reconstruction algorithm inspired by a bionic compound eye system," *CAAI Transactions on Intelligent Systems*, vol. 4, no. 2, pp. 180–187, 2009.
- [14] M. Li, H. Wang, C. Huang, L. Tu, and J. Zhou, "A self-adaptive algorithm for small targets detection in clutter scene inspired by insects compound eye," in *Proceedings of the 8th World Congress on Intelligent Control and Automation*, pp. 5268–5273, Jinan, China, July 2010.
- [15] 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, 2011.
- [16] S. Y. Chen, Y. F. Li, Q. Guan, and G. Xiao, "Real-time three-dimensional surface measurement by color encoded light projection," *Applied Physics Letters*, vol. 89, no. 11, Article ID 111108, 2006.
- [17] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 2012.
- [18] M. Egelhaaf, "On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. II. Figure- detection cells, a new class of visual interneurons," *Biological Cybernetics*, vol. 52, no. 3, pp. 195–209, 1985.
- [19] S. D. Wiederman, P. A. Shoemaker, and D. C. O'Carroll, "A model for the detection of moving targets in visual clutter inspired by insect physiology," *PLoS ONE*, vol. 3, no. 7, Article ID e2784, 2008.
- [20] J. Xu, *Small Target Detection Technology Research of SAR Image in Different Noise Wave Backgrounds*, Huhan University, Huhan, China, 2008.

Research Article

Separable Transition Density in the Hybrid Model for Tumor-Immune System Competition

Carlo Cattani¹ and Armando Ciancio²

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

²Department of Mathematics, University of Messina, Viale Ferdinando Stagno d'Alcontres 31, 98166 Messina, Italy

Correspondence should be addressed to Armando Ciancio, aciancio@unime.it

Received 16 September 2011; Accepted 8 October 2011

Academic Editor: Sheng-yong Chen

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

A hybrid model, on the competition tumor cells immune system, is studied under suitable hypotheses. The explicit form for the equations is obtained in the case where the density function of transition is expressed as the product of separable functions. A concrete application is given starting from a modified Lotka-Volterra system of equations.

1. Introduction

The competition between tumor cells and the immune system is mainly due to a significant presence of the proliferation and/or destructive events. In particular, cancer cells have the ability of expressing their biological activity to escape from the immune system which, in principle, have to challenge the progressing cells. The biological activity is not generally the same for all cells since it is statically distributed.

Several authors [1–7] have applied the methods of the classical mathematical kinetic theory of gases to study the immune competition with special attention to cancer phenomena. In this approach, one has to take account of statistical averages and stochastic parameters, typical of macro-models.

Other authors [8–15] have proposed mathematical models based on nonlinear differential equations, which generalize the classical Lotka-Volterra equations. These equations, as known, follow from a deterministic approach on a micro-scale.

In some recent papers [16–20], a hybrid model was proposed which can be considered as an alternative method between the above two approaches, aiming to mix the two scales into a unique set of equations, the hybrid model. In this model, a system of nonlinear ordinary differential equations are coupled with a stochastic parameter generated by

the (kinetic) interaction between the tumor cells and the immune system.

This time-depending stochastic parameter was linked [17] to the hiding-learning information process which underlies the cells competition. In particular [17], the hiding-learning dynamics appears between two populations (tumor cells-immune system) in which the first one has an uncontrolled proliferating and hiding ability and the second one has higher destructive ability and the need of learning about the presence of the first population.

In this paper, we study the above hybrid model by assuming a particular form of the stochastic coefficient. There follow interesting results on the model and, moreover, the classical model of Lotka-Volterra modified by the hiding-learning process can be derived as a special case.

2. Modelling the Immune Competition of Complex Systems

Let us consider a system of two interacting and competing populations. Each population is constituted by a large number of individuals called active particles; their microscopic state is called (biological) activity. This activity enables the particle to organize a suitable response with respect to any information process. In absence of prior information, the

activity reduces either to a minimal loss of energy or to a random process.

In active particle competitions, the simplest model of binary interaction is based on proliferation-destructive competition. So that, when the first population get aware of the existence of the other challenging population, it starts to proliferate and destroy the competing cells. However, in this process the most important step is the ability of cells to hide themselves and to learn about the activity of the competing population.

In details consider a physical system of two interacting populations each one constituted by a large number of active particles with sizes:

$$n_i = n_i(t), \quad (n_i(t) : [0, T] \longrightarrow \mathbb{R}_+; i = 1, 2). \quad (1)$$

Particles are homogeneously distributed in space, while each population is characterized by a microscopic state, called activity, denoted by the variable u . The physical meaning of the microscopic state may differ for each population. We assume that the competition model depends on the activity by a function of the overall distribution:

$$\mu = \mu[f_i(t, u)], \quad (\mu[f_i(t, u)] : \mathbb{R}_+ \longrightarrow \mathbb{R}_+). \quad (2)$$

The description of the overall distribution over the microscopic state within each populations is given by the probability density function:

$$\begin{aligned} f_i &= f_i(t, u), \\ (f_i(t, u) : [0, T] \times D_u \longrightarrow \mathbb{R}_+, D_u \subseteq \mathbb{R}; i = 1, 2) \end{aligned} \quad (3)$$

such that $f_i(t, u)du$ is the probability that the activity u of particles of the i th population, at the time t , ranges in the interval $[u, u + du]$.

Moreover, it is

$$\boxed{\forall i, \quad \forall t \geq 0 : 0 \leq f_i(t, u) \leq 1, \quad \int_{D_u} f_i(t, u)du = 1.} \quad (4)$$

We will see in Sections 3 and 4 how the microscopic structure influences the macroscopic system.

3. Hybrid Model

We consider, in this section, the competition between two cell populations: the first one with uncontrolled proliferating ability and with hiding ability; the second one with higher destructive ability, but with the need of learning about the presence of the first population. The analysis developed in what follows is referring to a specific case where the second population attempts to learn about the first population which, instead, escapes by modifying its appearance. Specifically, the hybrid evolution equations can be formally written as follows:

$$\begin{aligned} \frac{dn_i}{dt} &= G_i(n_1, n_2; \mu[f]), \\ \frac{\partial f_i}{\partial t} &= \mathbb{A}_i[f], \end{aligned} \quad (5)$$

where

- (1) G_i , for $i = 1, 2$, is a function of $n = \{n_1, n_2\}$,
- (2) μ , acts over $f = \{f_1, f_2\}$,
- (3) \mathbb{A}_i , for $i = 1, 2$, is a nonlinear operator acting on f ,
- (4) $\mu[f]$ is a functional ($0 \leq \mu \leq 1$) which describes the ability of the second population to identify the first one.

As a consequence, (5) denotes a hybrid system of a deterministic system coupled with a microscopic system statistically described by a kinetic theory approach. In the following, the evolution of density distribution will be taken within the kinetic theory.

The derivation of (5)₂ can be obtained starting from a detailed analysis of microscopic interactions. Specifically, consider binary interactions between a test, or candidate, particle with state u_* belonging to the i th population, and field particle with state u^* belonging to the j th population. We assume that microscopic interactions are characterized by the following quantities.

- (i) The encounter rate, which depends, for each pair of interacting populations on a suitable average of the relative velocity η_{ij} , with $i, j = 1, 2$.
- (ii) The transition density function $\varphi_{ij}(u_*, u^*, u)$, denotes the probability density that a candidate particle with activity u_* belonging to the i th population, falls into the state $u \in D_u$, of the test particle, after an interaction with a field entity, belonging to the j th population, with state u^* . The probability density $\varphi_{ij}(u_*, u^*, u)$ fulfills the condition

$$\boxed{\forall i, j, \quad \forall u_*, u^* : \int_{D_u} \varphi_{ij}(u_*, u^*, u)du = 1, \quad \varphi_{ij}(u_*, u^*, u) > 0.} \quad (6)$$

Then, by using the mathematical approach, developed in [17], it yields the following class of evolution equations:

$$\boxed{\begin{aligned} \frac{\partial f_i}{\partial t}(t, u) &= \sum_{j=1}^2 \int_{D_u \times D_u} \eta_{ij} \varphi_{ij}(u_*, u^*, u) \\ &\quad \times f_i(t, u_*) f_j(t, u^*) du_* du^* \\ &\quad - f_i(t, u) \sum_{j=1}^2 \int_{D_u} \eta_{ij} f_j(t, u^*) du^*, \end{aligned}} \quad (7)$$

which can be formally written as (5)₂.

Since our model is based on the hiding-learning dynamics, one has to introduce the functional which takes into account the ‘‘distance’’ between the two distribution so that μ in (5) is defined as

$$\mu[f_i, f_j](t) = \mu(|f_i - f_j|)(t) \quad (8)$$

with

$$\begin{aligned} 0 \leq \mu[f_i, f_j](t) \leq 1, \quad \forall u \in D_u \wedge t \in T, \\ \mu[f_i, f_j](t) = 1 \iff f_i = f_j, \\ \mu[f_i, f_j](t) = 0 \iff f_i = 0 \vee f_j = 0, \end{aligned} \quad (9)$$

where the maximum learning result is obtained when the second population is able to reproduce the distribution of the first one: $f_1 = f_2$, while the minimum learning is achieved when one distribution is vanishing.

In some recent papers [5–7, 17], it has been assumed that

$$\mu[f_i, f_j](t) = \mu(|f_i - f_j|)(t) = 1 - \int_{D_u} (f_1 - f_2)^2(t, u) du. \quad (10)$$

In this case, it is $\mu = 1$, when $f_1 = f_2$, otherwise $\mu \neq 1$ with $\mu \downarrow 0$, depending on the time evolution of the distance between f_1 and f_2 . There follows that this parameter could have an infinite value range.

Thus, we have

$$0 \leq \mu[f](t) \leq 1 \implies 0 \leq \int_{D_u} (f_1 - f_2)^2(t, u) du \leq 1, \quad \forall t \in [0, T]. \quad (11)$$

Notice that μ is the coupling term which links the macroscopic model (5)₁ to the microscopic model (5)₂.

4. Transition Density Function Based on Separable Functions

In order to find some classes of solutions of (7), we assume that the transition density is the product of separable density functions as

$$\varphi_{ij}(u_*, u^*, u) = (1 - \delta_{ij}) \psi_i(u_*, u) \xi_j(u^*, u), \quad (12)$$

that is,

$$\begin{aligned} \varphi_{11} = \varphi_{22} = 0, \\ \varphi_{12}(u_*, u^*, u) = \psi_1(u_*, u) \xi_2(u^*, u), \\ \varphi_{21}(u_*, u^*, u) = \psi_2(u_*, u) \xi_1(u^*, u), \end{aligned} \quad (13)$$

and using (10) one has

$$\begin{aligned} \int_{D_u} \psi_i(u_*, u) \xi_j(u^*, u) du = 1 \quad (i \neq j), \\ \psi_i(u_*, u) > 0, \quad \xi_j(u^*, u) > 0 \quad (i, j = 1, 2). \end{aligned} \quad (14)$$

By a substitution of the above terms into (7) we get

$$\begin{aligned} \frac{\partial f_1}{\partial t}(t, u) &= \sum_{j=1}^2 \int_{D_u \times D_u} \eta_{1j} \varphi_{1j}(u_*, u^*, u) f_1(t, u_*) f_j(t, u^*) du_* du^* \\ &\quad - f_1(t, u) \sum_{j=1}^2 \int_{D_u} \eta_{1j} f_j(t, u^*) du^*, \\ \frac{\partial f_2}{\partial t}(t, u) &= \sum_{j=1}^2 \int_{D_u \times D_u} \eta_{2j} \varphi_{2j}(u_*, u^*, u) f_2(t, u_*) f_j(t, u^*) du_* du^* \\ &\quad - f_2(t, u) \sum_{j=1}^2 \int_{D_u} \eta_{2j} f_j(t, u^*) du^*, \end{aligned} \quad (15)$$

from where, by taking into account (13), we obtain

$$\begin{aligned} \frac{\partial f_1}{\partial t}(t, u) &= \eta_{12} \int_{D_u} \psi_1(u_*, u) f_1(t, u_*) du_* \int_{D_u} \xi_2(u^*, u) f_2(t, u^*) du^* \\ &\quad - f_1(t, u) \left[\eta_{11} \int_{D_u} f_1(t, u^*) du^* + \eta_{12} \int_{D_u} f_2(t, u^*) du^* \right], \\ \frac{\partial f_2}{\partial t}(t, u) &= \eta_{21} \int_{D_u} \psi_2(u_*, u) f_2(t, u_*) du_* \int_{D_u} \xi_1(u^*, u) f_1(t, u^*) du^* \\ &\quad - f_2(t, u) \left[\eta_{21} \int_{D_u} f_1(t, u^*) du^* + \eta_{22} \int_{D_u} f_2(t, u^*) du^* \right]. \end{aligned} \quad (16)$$

According to (4) and (13), we have the more general system for the transition density based on separable functions

$$\begin{aligned} \frac{\partial f_1}{\partial t}(t, u) &= \eta_{12} \int_{D_u} \psi_1(u_*, u) f_1(t, u_*) du_* \\ &\quad \times \int_{D_u} \xi_2(u^*, u) f_2(t, u^*) du^* - (\eta_{11} + \eta_{12}) f_1(t, u), \\ \frac{\partial f_2}{\partial t}(t, u) &= \eta_{21} \int_{D_u} \psi_2(u_*, u) f_2(t, u_*) du_* \\ &\quad \times \int_{D_u} \xi_1(u^*, u) f_1(t, u^*) du^* - (\eta_{21} + \eta_{22}) f_2(t, u). \end{aligned} \quad (17)$$

This system (17) can be solved when the two functions of (14)₂ are given.

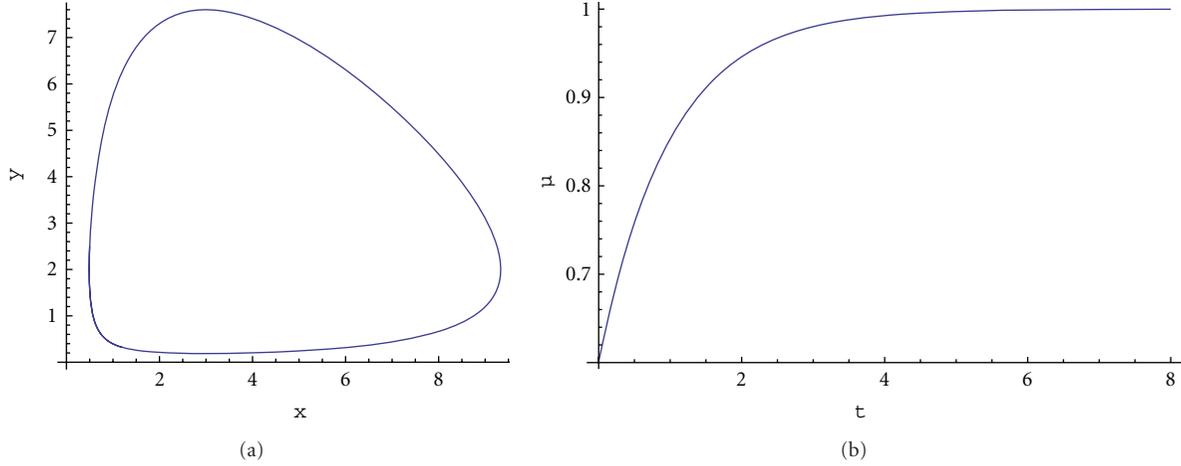


FIGURE 1: Time evolutions of the orbits of (30) with parameters $a = 2$, $b = c = 1$, and $d = 3$ (a); distribution function $\mu(t)$ for $\mu = 1/4$ (b).

As an example, let us solve this system under the following hypotheses:

$$\begin{aligned}\psi_1(u_*, u) &= \psi_2(u_*, u) = \delta(u - u_*), \\ \xi_1(u^*, u) &= \xi_2(u^*, u) = \delta(u - u^*),\end{aligned}\quad (18)$$

so that ψ_i and ξ_j ($i, j = 1, 2$) are a Dirac-delta which fulfill (14)₁

$$\int_{D_u} \delta(u - u_*) \delta(u - u^*) du = \delta(u_* - u^*). \quad (19)$$

The system (17), by using (18), becomes

$$\begin{aligned}\frac{\partial f_1}{\partial t}(t, u) &= \eta_{12} f_1(t, u) f_2(t, u) - (\eta_{11} + \eta_{12}) f_1(t, u), \\ \frac{\partial f_2}{\partial t}(t, u) &= \eta_{21} f_1(t, u) f_2(t, u) - (\eta_{21} + \eta_{22}) f_2(t, u).\end{aligned}\quad (20)$$

Moreover, by assuming that

$$\eta_{11} = \eta_{21} = \eta_{12} = \eta_{22} \stackrel{\text{def}}{=} \eta, \quad (21)$$

and putting

$$f(t, u) = f_1(t, u) - f_2(t, u), \quad (22)$$

from (20), one has

$$\boxed{\frac{\partial f(t, u)}{\partial t} = -2\eta f(t, u).} \quad (23)$$

The more general solution of this equation is

$$f(t, u) = f(0, u) e^{-2\eta t}. \quad (24)$$

Assuming that

$$f(0, u) = \frac{1}{\sqrt{\pi}} e^{-u^2}, \quad (25)$$

equation (24) becomes

$$\boxed{f(t, u) = \frac{1}{\sqrt{\pi}} e^{-(u^2 + 2\eta t)}.} \quad (26)$$

From (10), by virtue of (22) and (26), we have

$$\mu(t) = 1 - \int_{D_u} \frac{1}{\pi} e^{-2(u^2 + 2\eta t)} du. \quad (27)$$

Taking into account that

$$\int_{-\infty}^{+\infty} \frac{1}{\pi} e^{-2(u^2 + 2\eta t)} du = \frac{e^{-4\eta t}}{\sqrt{2\pi}}, \quad (28)$$

equation (27) gives

$$\boxed{\mu(t) = 1 - \frac{e^{-4\eta t}}{\sqrt{2\pi}}.} \quad (29)$$

5. A Simple Application

It is well known that the pioneering Lotka-Volterra's model of two interacting and competing populations ($x = \text{prey}$, $y = \text{predatory}$) is based on the following differential system:

$$\begin{aligned}\frac{dx}{dt} &= ax - bxy, \\ \frac{dy}{dt} &= cxy - dy,\end{aligned}\quad (30)$$

where a , b , c , and d are constants.

In this model, the hiding-learning processes are not considered and the interaction and competition of the two populations start immediately. The orbits of the solutions of (30) are circles around the equilibrium point: $x = d/c$, $y = a/b$ (see Figure 1).

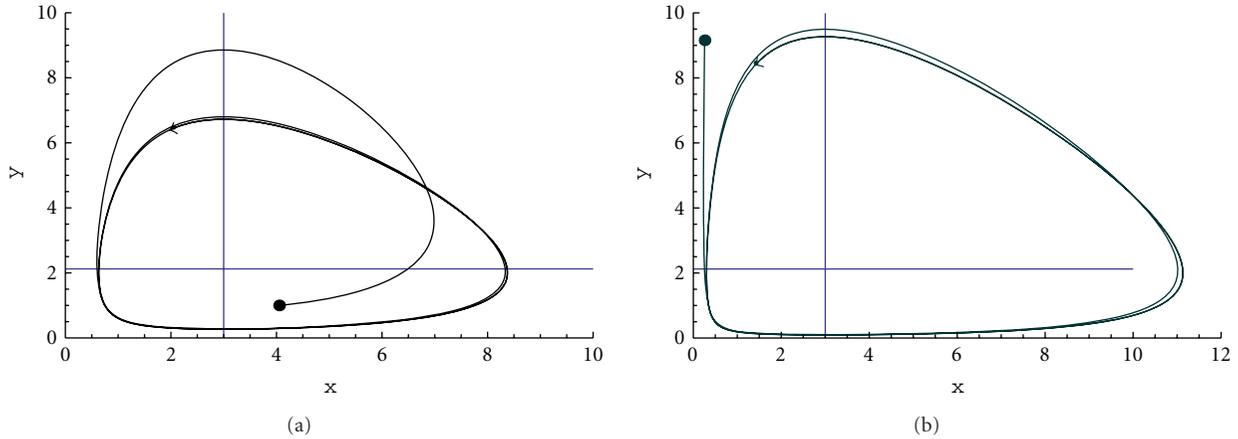


FIGURE 2: Time evolutions of the orbits of (32) with parameters $a = 2$, $b = c = 1$, and $d = 3$ and initial populations $x_0 = 4$, $y_0 = 1$ (a), $x_0 = 0.1$, $y_0 = 9$ (b).

If the hiding-learning processes occur, by using the results discussed in the previous sections, we propose the following system:

$$\begin{aligned} \frac{dx}{dt} &= ax - \mu bxy, \\ \frac{dy}{dt} &= cxy - dy, \end{aligned} \quad (31)$$

where μ , given by (29), is the functional (stochastic) parameter depending on the distribution of populations (see Figure 2).

The system (31) becomes

$$\begin{aligned} \frac{dx}{dt} &= ax - b \left(1 - \frac{e^{-4\eta t}}{\sqrt{2\pi}} \right) xy, \\ \frac{dy}{dt} &= cxy - dy. \end{aligned} \quad (32)$$

The nonzero equilibrium point is

$$x = \frac{d}{c}, \quad y = \frac{a}{b \left(1 - \left(\frac{e^{-4\eta t}}{\sqrt{2\pi}} \right) \right)}, \quad (33)$$

where

$$\lim_{t \rightarrow \infty} \frac{a}{b \left(1 - \left(\frac{e^{-4\eta t}}{\sqrt{2\pi}} \right) \right)} = \frac{a}{b}. \quad (34)$$

For $\eta = 1/4$, the solutions of the system (32) are shown in Figure 2.

From Figure 2 it can be noticed that $x_0 > y_0$ so that the hiding-learning process delay the achievement of the circle around the nonzero equilibrium point. If $x_0 \ll y_0$, then the cricle is reached more quickly.

6. Conclusion

In this paper, it has been studied a hybrid system of competition tumor cells versus immune system, within the kinetic model. A stochastic parameters is computed explicitly

in the case of special transition density functions. A simple application shows that due to this parameters we obtain some more realistic solutions of the Lotka-Volterra system, where the cicle around the nonzero equilibrium point is shifted in time, thus showing the importance of the stochastic parameters in a correct approach to the analysis of competition models.

References

- [1] N. Bellomo, A. Bellouquid, and M. Delitala, "Mathematical topics on the modelling complex multicellular systems and tumor immune cells competition," *Mathematical Models and Methods in Applied Sciences*, vol. 14, no. 11, pp. 1683–1733, 2004.
- [2] N. Bellomo and G. Forni, "Looking for new paradigms towards a biological-mathematical theory of complex multicellular systems," *Mathematical Models and Methods in Applied Sciences*, vol. 16, no. 7, pp. 1001–1029, 2006.
- [3] N. Bellomo and A. Bellouquid, "On the onset of non-linearity for diffusion models of binary mixtures of biological materials by asymptotic analysis," *International Journal of Non-Linear Mechanics*, vol. 41, no. 2, pp. 281–293, 2006.
- [4] N. Bellomo, A. Bellouquid, J. Nieto, and J. Soler, "Multicellular biological growing systems: hyperbolic limits towards macroscopic description," *Mathematical Models and Methods in Applied Sciences*, vol. 17, no. 1, pp. 1675–1692, 2007.
- [5] N. Bellomo and M. Delitala, "From the mathematical kinetic, and stochastic game theory to modelling mutations, onset, progression and immune competition of cancer cells," *Physics of Life Reviews*, vol. 5, no. 4, pp. 183–206, 2008.
- [6] N. Bellomo, *Modeling Complex Living Systems: A Kinetic Theory and Stochastic Game Approach*, Birkhäuser, Boston, Mass, USA, 2008.
- [7] C. Bianca and N. Bellomo, *Towards a Mathematical Theory of Multiscale Complex Biological Systems*, World Scientific, Singapore, 2010.
- [8] A. D'Onofrio, "A general framework for modeling tumor-immune system competition and immunotherapy: mathematical analysis and biomedical inferences," *Physica D*, vol. 208, no. 3-4, pp. 220–235, 2005.

- [9] A. d’Onofrio, “Tumor-immune system interaction: modeling the tumor-stimulated proliferation of effectors and immunotherapy,” *Mathematical Models and Methods in Applied Sciences*, vol. 16, no. 8, pp. 1375–1401, 2006.
- [10] A. d’Onofrio, “Tumor evasion from immune control: strategies of a MISS to become a MASS,” *Chaos, Solitons & Fractals*, vol. 31, no. 2, pp. 261–268, 2007.
- [11] A. d’Onofrio, “Metamodeling tumor-immune system interaction, tumor evasion and immunotherapy,” *Mathematical and Computer Modelling*, vol. 47, no. 5-6, pp. 614–637, 2008.
- [12] V. A. Kuznetsov, I. A. Makalkin, M. A. Taylor, and A. S. Perelson, “Nonlinear dynamics of immunogenic tumors: parameter estimation and global bifurcation analysis,” *Bulletin of Mathematical Biology*, vol. 56, no. 2, pp. 295–321, 1994.
- [13] N. V. Stepanova, “Course of the immune reaction during the development of a malignant tumour,” *Biophysics*, vol. 24, no. 5, pp. 917–923, 1979.
- [14] Y. Tao and H. Zhang, “A parabolic-hyperbolic free boundary problem modelling tumor treatment with virus,” *Mathematical Models and Methods in Applied Sciences*, vol. 17, no. 1, pp. 63–80, 2007.
- [15] V. A. Kuznetsov and G. D. Knott, “Modeling tumor regrowth and immunotherapy,” *Mathematical and Computer Modelling*, vol. 33, no. 12-13, pp. 1275–1287, 2001.
- [16] C. Cattani, A. Ciancio, and B. Lods, “On a mathematical model of immune competition,” *Applied Mathematics Letters*, vol. 19, no. 7, pp. 686–691, 2006.
- [17] C. Cattani and A. Ciancio, “Hybrid two scales mathematical tools for active particles modelling complex systems with learning hiding dynamics,” *Mathematical Models and Methods in Applied Sciences*, vol. 17, no. 2, pp. 171–187, 2007.
- [18] C. Cattani and A. Ciancio, “Third order model for tumor-immune system competition,” in *Proceedings of the 4th International Colloquium Mathematics in Engineering and Numerical Physics*, pp. 30–37, Bucharest, Romania, 2007.
- [19] C. Cattani and A. Ciancio, “Qualitative analysis of second-order models of tumor-immune system competition,” *Mathematical and Computer Modelling*, vol. 47, no. 11-12, pp. 1339–1355, 2008.
- [20] C. Cattani, A. Ciancio, and A. d’Onofrio, “Metamodeling the learning-hiding competition between tumours and the immune system: a kinematic approach,” *Mathematical and Computer Modelling*, vol. 52, no. 1-2, pp. 62–69, 2010.

Review Article

Recent Advances in Morphological Cell Image Analysis

Shengyong Chen,¹ Mingzhu Zhao,² Guang Wu,^{3,4} Chunyan Yao,^{1,5} and Jianwei Zhang⁵

¹ College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

² College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

³ Guangxi Academy of Sciences, 98 Daling Road, Nanning 530007, China

⁴ DreamSciTech Consulting, Shenzhen 518054, China

⁵ Department of Informatics, University of Hamburg, 22527 Hamburg, Germany

Correspondence should be addressed to Shengyong Chen, sy@ieee.org

Received 29 August 2011; Accepted 3 October 2011

Academic Editor: Carlo Cattani

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

This paper summarizes the recent advances in image processing methods for morphological cell analysis. The topic of morphological analysis has received much attention with the increasing demands in both bioinformatics and biomedical applications. Among many factors that affect the diagnosis of a disease, morphological cell analysis and statistics have made great contributions to results and effects for a doctor. Morphological cell analysis finds the cellular shape, cellular regularity, classification, statistics, diagnosis, and so forth. In the last 20 years, about 1000 publications have reported the use of morphological cell analysis in biomedical research. Relevant solutions encompass a rather wide application area, such as cell clumps segmentation, morphological characteristics extraction, 3D reconstruction, abnormal cells identification, and statistical analysis. These reports are summarized in this paper to enable easy referral to suitable methods for practical solutions. Representative contributions and future research trends are also addressed.

1. Introduction

Cell morphology has become a standard theory for computerized cell image processing and pattern recognition. The purpose of which is the quantitative characterization of cell morphology, including structure and inner-components analysis for better understanding functioning and pathogenesis associated with malignancy and behavior [1].

Morphological cell analysis is a key issue for abnormality identification and classification, early cancer detection, and dynamic changes analysis under specific environmental stress. The quantitative results and primary, objective, and reliable, which is beneficial to pathologists in making the final diagnosis and providing fast observation and automated analysis systems.

In the present study, advances in morphological cell analysis are briefly reviewed. Overall, significant progress has been made in several issues. Morphological cell analysis has been integrated in new methods for biomedical applications, such as automatic segmentation and analysis of histological

tumour sections [2–4], boundary detection of cervical cell nuclei considering overlapping and clustering [5, 6], the granules segmentation and spatial distribution analysis [7], morphological characteristics analysis of specific biomedical cells [8–10], understanding the chemotactic response and drug influences [11–14], or identifying cell morphogenesis in different cell cycle progression [15].

Morphological feature quantification for grading cancerous or precancerous cells is especially widely researched in the literature, such as nuclei segmentation based on marker-controlled watershed transform and snake model for hepatocellular carcinoma feature extraction and classification, which is important for prognosis and treatment planning [16], nuclei feature quantification for cancer cell cycle analysis [17], and using feature extraction including image morphological analysis, wavelet analysis, and texture analysis for automated classification of renal cell [18].

Computerized/automated early cancer or abnormalities detection provides a basis for reducing deaths and morbidity, especially for cervical cancer, which is reported to be

the most preventable disease through early detection [19], provision of prompt advice, and opportunities for follow-up treatments. As an example, [20] presents a prototype expert system for automated segmentation and effective cervical cancer detection, providing primary, objective, and reliable diagnostic results to gynaecologists in making the final diagnosis. These advances will contribute to realize computer-assisted, interactive, or automated processing, quantification, statistic analysis, and diagnosis systems for biomedical applications.

The scope of this paper is restricted to morphological cell analysis by image processing in the field of biomedical research. Although this topic has attracted researchers as since early as the 1980s [21–23], this survey concentrates on the contributions of the last 5 years. No review of this nature can possibly cite each and every paper that has been published. Therefore, we include only what we believe to be representative samples of important works and broad trends from recent years. In many cases, references were provided to better summarize and draw distinctions among key ideas and approaches.

The paper has five more sections. Section 2 briefly provides an overview of related contributions. Section 3 introduces the typical formulation of cell morphology. Section 4 lists the relevant tasks, problems, and applications of cell morphology. Section 5 concentrates typical solutions and methods. Section 6 is a discussion of our impressions on current and future trends. Section 7 is the conclusion.

2. Overview of Contributions

2.1. Summary. From 1980s to 2010, about 1000 research papers with topics on or closely related to morphological cell analysis for robot vision were published. Figure 1 shows the yearly distribution of these published papers. The plot shows that the topic of morphological cell analysis steadily developed in the past 20 years.

2.2. Representatives. Morphological cell analysis has many applications in biomedical engineering. Their most significant roles are summarized as follows.

- (1) Malignant cell identification and cancer detection [20, 24, 25].
- (2) Morphological changes during a cell cycle as division, proliferation, transition, and apoptosis [26–28] or to follow cell culture development [29].
- (3) Morphological differences to elucidate the physiological mechanisms [30] or classify a set of cell populations with different functions such as neurons [31, 32].
- (4) Dynamic characteristics investigation under specific environmental stress for personalized therapy [33–36] or for the selection of new drugs [37].
- (5) Morphometrical study such as subcellular structures (DNA, chromosome) analysis [38] for higher animals or plants based on 3D reconstruction [39, 40].

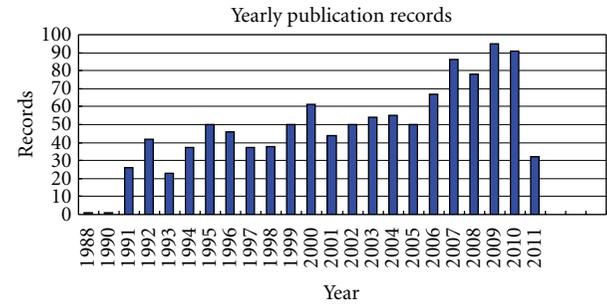


FIGURE 1: Yearly published records from 1990 to 2010.

TABLE 1: Representative contributions.

| Processing | Method | Representative |
|----------------|--|----------------|
| | Active contour model (ACM) | [5]—2011 |
| | Reconstruct the approximate location of cellular membranes | [51]—2011 |
| Segmentation | A marker-controlled watershed transform and a snake model | [16]—2010 |
| | Segmentation combing features | [51]—2011 |
| Classification | K -means and support vector machines (SVM) | [6]—2011 |
| | Bayesian classifier | [18]—2009 |

The commonly researched topics for solving morphological problems are listed below.

- (1) Mathematical morphology theory used in binary, gray, and color images for preprocessing or features analysis [41–48].
- (2) Location determination: objects located and analysis of distribution [7, 49, 50].
- (3) Meaningful areas segmentation: based on the features of pixel, edge, region, and model [2–4].
- (4) Characteristics quantification: based on cytopathology and the experience of physicians [51–58].
- (5) Recognition, classification automated analysis, and diagnosis [6, 16, 24, 51, 59].

Morphological analysis has become a powerful mathematical tool for analyzing and solving cell informatics. Automatic features quantification is undoubtedly the most widely used estimation technique in this topic. Among the variety of developed methods, the main differences and remarkable features can be summarized briefly: shape, geometrical, intensity, and texture. A few representative types of segmentation and classification are selected for easy appreciation of state-of-the-art as shown in Table 1.

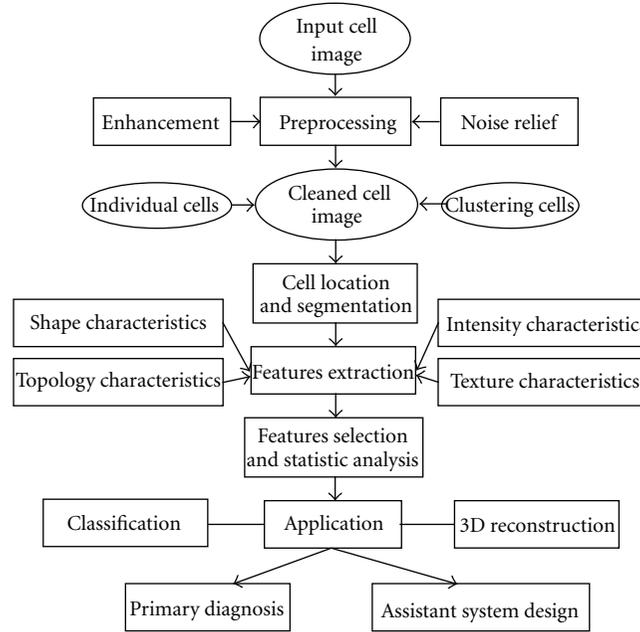


FIGURE 2: The general procedure of cell image analysis.

3. The Problem and Fundamental Principle

The fundamental principle of morphological cell analysis is dependent on cell biology, cytopathology, and the diagnostic experience of pathologists. To study cell characteristics, detect abnormalities, and determine the malignant degree, the pathologists examine biopsy material under a microscope, which is subjective, laborious, and time consuming. Therefore quantitative cell morphology is studied and computer-assisted systems are presented for diagnostic process at the same time. The general procedure of such applications can be described in Figure 2.

4. Tasks and Problems

4.1. Morphological Operation. Mathematical morphology is the basic theory for many image processing algorithms, which can also extract image shape features by operating with various shape-structuring elements [60]. This processing technique has proved to be a powerful tool for many computer-vision tasks in binary and gray scale images, such as edge detection, noise suppression, image enhancement, skeletonization, and pattern recognition, [45]. This technique is consisted of two parts: binary morphology and gray-scale morphology, and the commonly used operations as morphological dilation and erosion are defined as follows, respectively:

$$\begin{aligned} (f \oplus k)(x, y) &= \max\{f(x - m, y - n) + k(m, n)\}, \\ (f \ominus k)(x, y) &= \max\{f(x - m, y - n) - k(m, n)\}, \end{aligned} \quad (1)$$

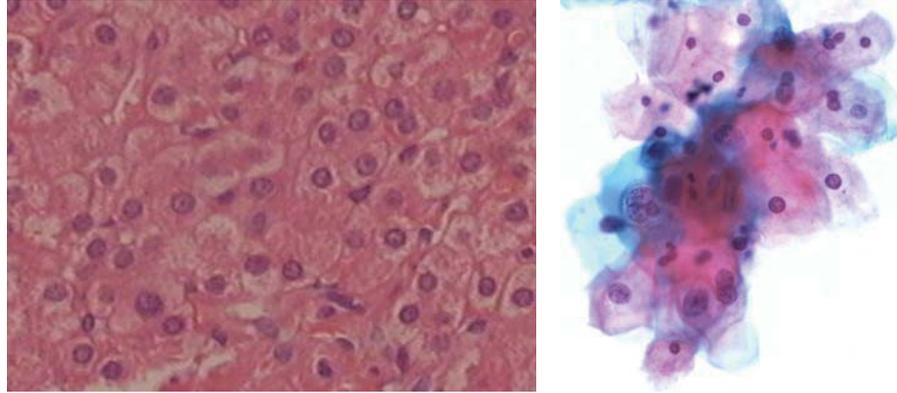
where f is the original image (gray scale or binary), which is operated by the corresponding structuring element k , and (x, y) is the pixel of image f , (m, n) is the size of element

k . After morphological operation, image shape features such as edges, fillets, holes, corners, wedges, and cracks can be extracted.

Mathematical morphology can also be used in color images avoiding the loss of information of traditional binary techniques [45]. The new operations are based on the order in multivariate data processing.

4.2. Cell Localization. Determination of the orientation of a cell, termed localization, is of paramount importance in achieving reliable and robust morphological analysis. Achieving high-level tasks such as segmentation and shape description is possible if the initial position is known. From the early literature, primary methods were used in sample images, such as [61] using a sequence of morphological image operations to identify the cell nuclei and [29] using conditional dilation techniques to estimate unbiasedly cell density and obtain precisely cell contours. The results were acceptable only in single images without any complex factors.

Even when membranes are partially or completely not visible in the image (Figure 3(a)), the approximate locations of cells can be detected by reconstructing cellular membranes [51]. This method is effective for lung cells location in immunohistochemistry tissue images. Cell nuclei that are in cell clusters detecting are the key point for eliminating the positions of cervical cells in conventional Pap smear images (Figure 3(b)). To deal with this problem, Plissiti et al. present a fully automated method [6]. It takes the advantage of color information to obtain the candidate nuclei centroids in the images and eliminate the undesirable artifacts by applying a distance-dependent rule on the resulted centroids and classification algorithms (fuzzy C-means and support vector machines). The experiments shows that even in the case of



(a) Lung cells

(b) Cervical cells

FIGURE 3: Biomedical cell images.

images with high degree of cell overlapping, the results are very promising.

For automatic detection of granules in different cell groups and statistical analysis of their spatial locations, the existing image analysis methods, such as single threshold, edge detection, and morphological operation, cannot be used. Thus, the empirical cumulative distribution function of the distances and the density of granules can be considered [7]. Jiang et al. propose a machine learning method [62], which is based on Haar feature (which is the combination of the intensity, shape, and scale information of the objects), to detect the particle's position.

4.3. Segmentation. Segmentation is one of the most important points for automated image analysis and better cell information understanding. The algorithms that have been presented can be divided into edge-based, region-based, and model-based modules. Region-based approaches attempt to segment an image into regions according to regional image data similarity (or dissimilarity), such as scale-space filtering, watershed clustering [63], gray-level threshold [26], and region growing [64]. For clear stained images, multilevel thresholds are the most simply and commonly applied methods for low-level segmentation to remove noise and obtain the interest region (nucleus, cytoplasm, or the whole cell), which are defined as follows:

$$g(x, y) = \begin{cases} I_i, & T_{i-1} \leq f(m, n) \leq T_i, \\ 0, & \text{others,} \end{cases} \quad (2)$$

where i is the number of regions need to be divided, T_i is the threshold and the extension ranges from T_{i-1} to T_i corresponding to the region i .

Nevertheless numerous algorithms have been developed, overlapping and connected cluster is still the key problem in cell image segmentation. The methods presented available to solve specific images with clear stained situation, semi-automated algorithms based on preknowledge for adequate segmentation of cell images under complex situation, are always more efficient than totally automated methods.

4.4. Quantitative Measurement of Meaningful Parameters.

The quantitative measurement of cell features is meaningful for both image segmentation and abnormalities detection. Fast, reproducible, accurate, and objective measurement of cell morphology is beneficial to avoid subjective and interobserver variations, which result in diagnostic shifts and consequently disagreement between different interpreters [20]. The quantitative characteristics of cell or nuclear structure alterations extracted after robust image processing algorithms and 3D reconstruction is also called morphological biosignatures, which learn about cellular level features and nuclear structure including inner-components analysis, such as the quantitative evaluation of the approximate number of mRNA varying during cell cycle, developing, aging, and in different pathologies and treatment with drugs by extracting morphological parameters (cytoplasm and nucleus areas) [28]. Accurate quantification of these parameters could be beneficial for developing robust biosignatures for early cancer detection [1]. Multivariate statistical analyses of morphological data to suggest that quantitative cytology may be a useful adjunct to conventional tests for the selection of new drugs with differentiating potential [37].

The extracting features as cell area, perimeter, centroid, and the length of major and minor axes for calculating more meaningful parameters such as displacement, protrusiveness, and ellipticity, are used to analyze the dynamic changes of human cancerous glioma cells [35], which can also be used to identify different classed of neurons and relate neural structure (such as total dendritic length and dendritic field area) to function [31].

The most meaningful parameters are obtained in discriminating different patterns, such as cell size, shape distribution, and nuclear-to-cytoplasmic ratio for normal and precancerous cervical squamous epithelium determination [44], and texture quantification as a measurement to interchromosome coarseness to study cell proliferation [38]. Local gray level differences and cell density combining with other morphological parameters are possible to follow cell culture development under various experimental conditions [29]. Hitherto, the relationship between malignancy-associated morphological features in single tumour cells and

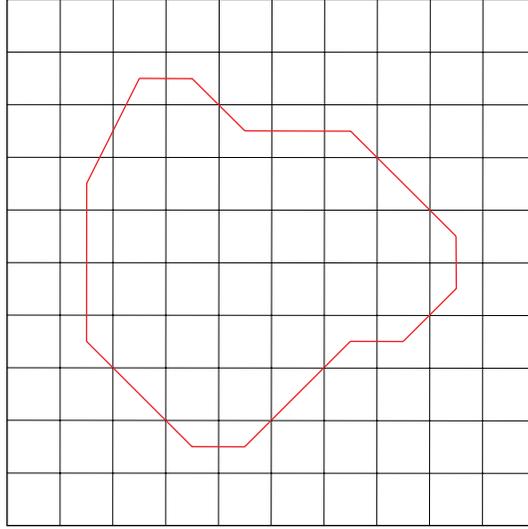


FIGURE 4: Geometrical features quantification.

the expression of markers indicating functional properties of these cells remained widely unknown [65].

4.5. Statistical Analysis. Multivariate statistic analysis is applied to compare multivariate data and establish the quantitative changes and differences between groups under investigation on their characteristics. The kernel approach is to find a high correlation feature set without redundancy. Principal components analysis (PCA) displays the original variables in a bidimensional space, thus reducing the dimensionality of the data and allowing the visualization of a large number of variables into a two-dimensional plot [11, 49, 66].

5. Methods and Solutions

5.1. Formulation in Morphological Analysis. Morphological analysis is often studied as the shape appearances of objects and the surfaces of the images, with intensity seen as height and texture appearing as relief. Formulation of morphological features is of benefit to computerized calculation and more efficient than manual morphological quantification, which is still laborious and subjective. The morphology characteristics can be described by shape, geometrical, intensity, and texture analysis.

The geometrical features of regions can be described by area, radii, perimeter, the major and the minor axis length, and so forth. The area of the object is calculated as the number of pixels of the region (Figure 4, the area defined by the closed curve). Radii are calculated based on projected cell area supposing that each cell is circular. The major and the minor axis length are the maximal and minimum numbers of pixels of the axis, respectively. Take Figure 4 as an example, the perimeter is calculated as follows:

$$P = N_1 + N_2 + \sqrt{2}N_3, \quad (3)$$

where, N_1 , N_2 , N_3 are the numbers of the horizontal, vertical bevel lines on the boundary, respectively.

Circularity, rectangle, eccentricity, and irregularity are used to describe the shape features. Circularity (C) and rectangle (R) represent the rotundity-like and rectangle-like degree, defined as follows:

$$C = \frac{P^2}{4\pi A}, \quad (4)$$

$$R = \frac{\text{Area}}{H * W}.$$

Eccentricity is defined as follows:

$$E = \frac{\text{The minor axis length}}{\text{The major axis length}}. \quad (5)$$

Texture is an important visual cue and widely exists in images. Texture feature extraction is the most basic problem for texture analysis including classification and segmentation. Dimension, discrimination, stability, and calculation are considered in practical application and studied for more than fifty years. Based on the statistical theory, structure, model, and signal processing, many effective methods were presented for different applications. Among which, gray level co-occurrence matrix (GLCM) has become one of the best known and most widely used statistic method for texture feature extraction [26], especially in cell image texture feature analyzing. The interrelationship of textural primitives which define morphological texture can be estimated by quite different descriptors, the discriminant value of which varies considerably [67]. The descriptors based on GLCM are summarized in Table 2.

The intensity feature is characterized by the average of the intensity value of all the pixels of the region. For RGB color images, it is calculated independently from the red, green, and blue component of the original image. Histogram is an efficient way to show intensity features. Kruk et al.

TABLE 2: Texture features.

| | |
|---|--|
| Energy: | $ASM = \sum_{i=1}^k (g_i - g)^{-2} p(g_i)$ |
| Uniformity: | $U = \sum_{i=1}^k p^2(g_i)$ |
| Entropy: | $ENT = -\sum_{i=1}^k p(g_i) \log_2 p(g_i)$ |
| Smoothness: IDM = $1 - 1/(1 + s^2)$, where $s = \sqrt{\sum_{i=1}^k (g_i - g)^{-2} p(g_i)}$ | |

Given that g_i is the gray value, k is the number of gray levels.

characterize the histograms of different color components by applying the following parameters: the mean, standard deviation, skewness, kurtosis, maximum value, and the span of the histogram [59].

5.2. Deformable Models. It is known that biomedical images are always under complex situation, which made segmentation a hard task for the extraction of the interest region. Because of the various challenges in medical image processing, deformable models were widely investigated and innovated, becoming a powerful tool for medical image segmentation. Active counter model is one of the most classical algorithms. Techniques based on active contour models have the potential to produce better estimates of cell morphologies.

The existing active contour models can be categorized into two classes: edge-based models [68], and region-based models [69]. On one hand edge-based model directly uses intensity gradient information to attract the contour toward the object boundaries. Therefore this kind of model has worse performance for weak object boundaries since cell image exhibits great fuzzy degree due to low contrast at the location of the cell membrane. On the other hand region-based model aims to identify each region of interest by using a certain region descriptor. It guides the motion of the contour, and is less sensitive to the location of initial contours in some extents. It is much more suitable for cell segmentation than the fore one.

Chan and Vese model [70] is one of the most popular region-based active contour models. This model has been successfully used for segmenting images. Chan and Vese model proposed an active contour model that segments an image into two sets of possibly disjoint regions, by minimizing a simplified Mumford-Shah functional. The basic idea is as follows. Assume that $\Omega \subset R^2$ is the image domain and $I: \Omega \rightarrow R$ is a given image. Mumford and Shah consider image segmentation as a problem of seeking an optimal contour C that divides the image domain into two approximately piecewise-constant regions with intensities u_i and u_0 . Let C denote its boundary. Thus the global data fitting term in the Chan and Vese model is defined as follows:

$$E^{cv}(c_1, c_2) = \int_{\bar{\Omega}} (I - c_1)^2 dx dy + \int_{\Omega} (I - c_2)^2 dx dy, \quad (6)$$

where Ω and $\bar{\Omega}$ represent the regions outside and inside the contour C , respectively, c_1 and c_2 are two constants that fit the image intensities outside C and inside C .

This model considers pixels within the same region having the most similarity, and makes up the shortcomings of edged etector. When the contour accurately captures the object boundary, the two fitting terms minimize the fitting energy value. In each segmented area, the clustered pixels' mean value approximately equals the c_1 and c_2 , respectively. Thus the fitting terms with respect to c_1 and c_2 are the driving forces that evolve the curve motion on the principle of inner-region homogeneity.

Since the regional difference is the guideline in image segmentation, the interregional differences should be considered as the model's driving force as follows:

$$E = -\frac{1}{2}(c_1 - c_2)^2. \quad (7)$$

This kind of region-based active contour model's energy is characterized by the maximum dissimilarity between regions. Minimizing the energy E in (7) is the same as maximizing the difference between different regions. Equation (7) formulates the global instructive guidance term.

5.3. Classification. The extracted features involved the input to classification procedure for better analysis, correct grading, and pattern recognition. From the literature, unsupervised (as K -means and spectral clustering) and supervised (as super vector machine, SVM) classification schemes and artificial neural network (ANN) architecture were applied. SVM clustering is a state-of-the-art method, which was originally proposed in [71]. The decision function of a two-class problem can be written as follows:

$$f(x) = \omega \cdot \phi(x) + b = \sum_{i=1}^N \alpha_i y_i K(x, x_i) + b, \quad (8)$$

where $x_i \in R^d$ is the sample and $y_i \in \{\pm 1\}$ is the class label of x_i . A transformation $\phi(\cdot)$ maps the data points x of the input space R^d into a higher-dimensional feature space R^D , ($D \geq d$). $K(\cdot, \cdot)$ is a kernel function, which defines an inner product in R^D . $K(\cdot, \cdot)$ is commonly defined as follows:

$$\begin{aligned} K(x, x_i) &= [(x \cdot x_i) + 1]^q, \\ K(x, x_i) &= \exp\left\{-\frac{|x - x_i|^2}{\sigma^2}\right\}, \\ K(x, x_i) &= \tanh(v(x \cdot x_i) + c). \end{aligned} \quad (9)$$

The parameters $\alpha_i \geq 0$ are optimized by finding the hyperplane in feature space with maximum distance to the closest image $\phi(x_i)$ from the training set. For multilevel classification based on SVM, a decision-tree classification scheme discriminated between different grades is showed in Figure 5.

Although SVM is one of the most famous methods for classification and has achieved a great success in pattern recognition, problems still exist, such as the neglect of different data distributions within classes. Recently, structural super vector machine (SSVM) is proposed accordingly, which firstly exploits the intrinsic structures of samples

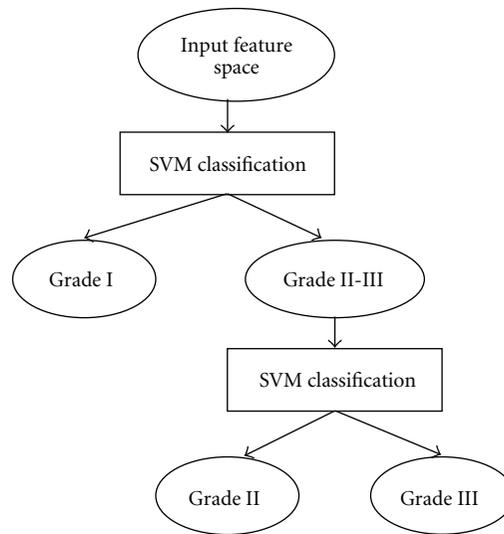


FIGURE 5: A decision-tree SVM classification scheme.

within classes by some unsupervised clustering methods and directly embedding the structural information into the SVM objective function [72]. SSVM is theoretically and empirically a better generalization than SVM algorithm.

5.4. D Morphology. Three-dimensional morphology using 3D reconstruction and image processing techniques is applied for quantitative morphometric analysis of cellular and subcellular structures, which is much more powerful than its 2D counterpart, but still largely based on the processing of separate 2D slices.

The approach to 3D morphological analysis consists of digital micrographs acquisition, reconstruction, and 3D-based feature extraction. The acquired images are serially taken by CT instrument at uniform angular intervals during a full 360° rotation [1], from the electron imaging film taken by photo products [73], or by electron microscopy [40]. Computer programs such as MATLAB or Visual Studio software can be used for automated 3D image reconstruction.

Based on the reconstructed models, features such as three-dimensional shape of the cells can be extracted, which are correlated with the assembly state of myofibrils in different stages [74] and ultrastructure such as the arrangement of compact chromatin of GO lymphocytes can be studied [23].

6. Existing Problems and Future Trends

Although morphological cell analysis has been developed in many applications as mature approaches for estimation and diagnosis, some problems still exist in its applications in biomedical engineering. Researchers are exerting efforts not only in simple localization and segmentation, but also in improving the methods mainly in the following aspects.

6.1. Real-Time Application and Computational Complexity. Morphological cell analysis has been applied in almost all hospitals, which are key means in automatic microscopic

analysis. However, because of its high computational complexity, it has strict limits on the number and stability of feature points. The traditional method selects a few features, which limits the application scope of morphological analysis. The computational complexity greatly affects real-time application systems [50, 75].

6.2. Reliability. Reliability is a great concern in practical applications [55, 76]. Morphological analysis relies on tuning of many parameters. Related techniques rely on existing noise statistics, initial positions, and sufficiently good approximation of measurement functions. Deviations from such assumptions usually lead to degraded estimations during automatic analysis. Stochastic stability is established in terms of the conditions of the initial errors, bound on observation noise covariance, observation nonlinearity, and modeling error. Features have to be effectively and efficiently treated by their removal from or addition to the system. New methods should be explored to discard outliers and improve the matching rate. These will help stabilize algorithms and allow more accurate localizations or parametric estimations.

6.3. With a Priori Knowledge. Constraints introduced in morphological cell parameters may help in some occasions. For example, morphological cell analysis is commonly used to estimate the cell shapes and activities, which incorporate *a priori* information in a consistent manner. However, the known model or information are often either ignored or heuristically dealt with [6].

6.4. Accuracy. Accuracy is always the most important factor in biomedical engineering. The accuracy of the calculated cells strongly depends on the computational potential and the statistical possibilities. For example, automated method provides accurate segmentation of the cellular membranes in the stained tracts and reconstructs the approximate location of the unstained tracts using nuclear membranes as a spatial

reference. Accurate cell-by-cell membrane segmentation allows per-cell morphological analysis and quantification of the target membrane [16, 51, 77].

6.5. Artificial Intelligence. The integration of the morphological cell analysis with some artificial intelligence methods may yield a better performance. Fuzzy logic, neural network, genetic algorithm, and so forth can be combined to wholly resolve the complex task.

7. Conclusion

This paper summarizes recent advances in morphological cell analysis for biomedical engineering applications. Typical contributions are addressed for initialization, localization, segmentation, estimation, modeling, shape analysis, cell parameters, and so forth. Representative works are listed for readers to have a general overview of state-of-the art. A number of methods for solving morphological problems are investigated. Many methods developed for morphological cell analysis, extended morphological cell segmentation, are introduced. In the 20-year history of morphological cell analysis, they gained entry into the field of biomedical engineering as a critical role. The largest volume of published reports in this literature belongs to the last ten years.

Acknowledgments

This paper was supported in part by the National Natural Science Foundation of China and Microsoft Research Asia (NSFC nos. 61173096, 51075367, and 60870002), Zhejiang Provincial S&T Department (2010R10006, 2010C33095), and Zhejiang Provincial Natural Science Foundation of China (R1110679).

References

- [1] V. Nandakumar, L. Kelbaskas, R. Johnson, and D. Meldrum, "Quantitative characterization of preneoplastic progression using single-cell computed tomography and three-dimensional karyometry," *Cytometry Part A*, vol. 79, no. 1, pp. 25–34, 2011.
- [2] C. C. Reyes-Aldasoro, L. J. Williams, S. Akerman, C. Kanthou, and G. M. Tozer, "An automatic algorithm for the segmentation and morphological analysis of microvessels in immunostained histological tumour sections," *Journal of Microscopy*, vol. 242, no. 3, pp. 262–278, 2011.
- [3] J. Z. Cheng, Y. H. Chou, C. S. Huang et al., "ACCOMP: augmented cell competition algorithm for breast lesion demarcation in sonography," *Medical Physics*, vol. 37, no. 12, pp. 6240–6252, 2010.
- [4] J. S. Schildkraut, N. Prosser, A. Savakis et al., "Level-set segmentation of pulmonary nodules in megavolt electronic portal images using a CT prior," *Medical Physics*, vol. 37, no. 11, pp. 5703–5710, 2010.
- [5] M. E. Plissiti, C. Nikou, and A. Charchanti, "Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images," *Pattern Recognition Letters*, vol. 32, no. 6, pp. 838–853, 2011.
- [6] M. E. Plissiti, C. Nikou, and A. Charchanti, "Automated detection of cell nuclei in Pap smear images using morphological reconstruction and clustering," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 233–241, 2011.
- [7] E. Díaz, G. Ayala, M. E. Díaz, L. W. Gong, and D. Toomre, "Automatic detection of large dense-core vesicles in secretory cells and statistical analysis of their intracellular distribution," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 7, no. 1, Article ID 4468698, pp. 2–11, 2010.
- [8] F. Brun, A. Accardo, M. Marchini, F. Ortolani, G. Turco, and S. Paoletti, "Texture analysis of TEM micrographs of alginate gels for cell microencapsulation," *Microscopy Research and Technique*, vol. 74, no. 1, pp. 58–66, 2011.
- [9] S. Amini, D. Veilleux, and I. Villemure, "Tissue and cellular morphological changes in growth plate explants under compression," *Journal of Biomechanics*, vol. 43, no. 13, pp. 2582–2588, 2010.
- [10] Y. Xiong, C. Kabacoff, J. Franca-Koh, P. N. Devreotes, D. N. Robinson, and P. A. Iglesias, "Automated characterization of cell shape changes during amoeboid motility by skeletonization," *BMC Systems Biology*, vol. 4, article 33, 2010.
- [11] Y. A. Xiong and P. A. Iglesias, "Tools for analyzing cell shape changes during chemotaxis," *Integrative Biology*, vol. 2, no. 11–12, pp. 561–567, 2010.
- [12] Z. Q. Hong, L. M. Tao, and L. Li, "Effect of stress on mRNA expression of H⁺-ATPase in osteoclasts," *Molecular and Cellular Biochemistry*, vol. 343, no. 1–2, pp. 183–190, 2010.
- [13] Y. S. Lin, C. C. Lin, Y. S. Tsai, T. C. Ku, Y. H. Huang, and C. N. Hsu, "A spectral graph theoretic approach to quantification and calibration of collective morphological differences in cell images," *Bioinformatics*, vol. 26, no. 12, Article ID btq194, pp. i29–i37, 2010.
- [14] P. Venkatesan, S. Das, M. M. R. Krishnan, C. Chakraborty, K. Chaudhury, and M. Mandal, "Effect of AEE788 and/or Celecoxib on colon cancer cell morphology using advanced microscopic techniques," *Micron*, vol. 41, no. 3, pp. 247–256, 2010.
- [15] S. Ohnuki, S. Nogami, and Y. Ohya, "A microfluidic device to acquire high-magnification microphotographs of yeast cells," *Cell Division*, vol. 4, article 5, 2009.
- [16] P. W. Huang and Y. H. Lai, "Effective segmentation and classification for HCC biopsy images," *Pattern Recognition*, vol. 43, no. 4, pp. 1550–1563, 2010.
- [17] F. H. Li, X. B. Zhou, J. W. Ma, and S. T. C. Wong, "Multiple nuclei tracking using integer programming for quantitative cancer cell cycle analysis," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, Article ID 5175475, pp. 96–105, 2010.
- [18] Q. Chaudry, S. H. Raza, A. N. Young, and M. D. Wang, "Automated renal cell carcinoma subtype classification using morphological, textural and wavelets based features," *Journal of Signal Processing Systems*, vol. 55, no. 1–3, pp. 15–23, 2009.
- [19] M. Arbyn, M. Rebolj, I. M. C. M. De Kok et al., "The challenges of organising cervical screening programmes in the 15 old member states of the European Union," *European Journal of Cancer*, vol. 45, no. 15, pp. 2671–2678, 2009.
- [20] S. W. K. Chan, K. S. Leung, and W. S. F. Wong, "An expert system for the detection of cervical cancer cells using knowledge-based image analyzer," *Artificial Intelligence in Medicine*, vol. 8, no. 1, pp. 67–90, 1996.
- [21] L. Ficsor, V. S. Varga, A. Tagscherer, Z. Tulassay, and B. Molnar, "Automated classification of inflammation in colon histological sections based on digital microscopy and advanced image analysis," *Cytometry Part A*, vol. 73, no. 3, pp. 230–237, 2008.

- [22] A. Tárnok, G. K. Valet, and F. Emmrich, "Systems biology and clinical cytomics: the 10th Leipziger Workshop and the 3rd International Workshop on Slide-Based Cytometry," *Cytometry Part A*, vol. 69, no. 1, pp. 36–40, 2006.
- [23] G. López-Velázquez, J. Márquez, E. Ubaldo, G. Corkidi, O. Echeverría, and G. H. Vázquez Nin, "Three-dimensional analysis of the arrangement of compact chromatin in the nucleus of G₀ rat lymphocytes," *Histochemistry and Cell Biology*, vol. 105, no. 2, pp. 153–161, 1996.
- [24] D. Glotsos, P. Spyridonos, D. Cavouras, P. Ravazoula, P. A. Dadioti, and G. Nikiforidis, "An image-analysis system based on support vector machines for automatic grade diagnosis of brain-tumour astrocytomas in clinical routine," *Medical Informatics and the Internet in Medicine*, vol. 30, no. 3, pp. 179–193, 2005.
- [25] C. O. De Solórzano, S. Costes, D. E. Callahan, B. Parvin, and M. H. Barcellos-Hoff, "Applications of quantitative digital image analysis to breast cancer research," *Microscopy Research and Technique*, vol. 59, no. 2, pp. 119–127, 2002.
- [26] G. A. Losa and C. Castelli, "Nuclear patterns of human breast cancer cells during apoptosis: characterisation by fractal dimension and co-occurrence matrix statistics," *Cell and Tissue Research*, vol. 322, no. 2, pp. 257–267, 2005.
- [27] T. Q. Xie, M. L. Zeidel, and Y. T. Pan, "Detection of tumorigenesis in urinary bladder with optical coherence tomography: optical characterization of morphological changes," *Optics Express*, vol. 10, no. 24, pp. 1431–1443, 2002.
- [28] M. Colombi, L. Moro, N. Zoppi, and S. Barlati, "Quantitative evaluation of mRNAs by in situ hybridization and image analysis: principles and applications," *DNA and Cell Biology*, vol. 12, no. 7, pp. 629–636, 1993.
- [29] W. Y. Xu-Van Opstal, C. Ranger, O. Lejeune et al., "Automated image analyzing system for the quantitative study of living cells in culture," *Microscopy Research and Technique*, vol. 28, no. 5, pp. 440–447, 1994.
- [30] P. A. Melrose, C. Pickel, H. S. Cheramie, W. G. Henk, M. A. Littlefield-Chabaud, and D. D. French, "Distribution and morphology of immunoreactive gonadotropin-releasing hormone (GnRH) neurons in the basal forebrain of ponies," *Journal of Comparative Neurology*, vol. 339, no. 2, pp. 269–287, 1994.
- [31] M. F. Villa and F. R. Amthor, "Automating the quantitative analysis of 2-D neural dendritic trees," *Journal of Neuroscience Methods*, vol. 56, no. 1, pp. 77–88, 1995.
- [32] M. Masseroli, A. Bollea, and G. Forloni, "Quantitative morphology and shape classification of neurons by computerized image analysis," *Computer Methods and Programs in Biomedicine*, vol. 41, no. 2, pp. 89–99, 1993.
- [33] M. C. Albertini, A. Accorsi, L. Teodori et al., "Use of multiparameter analysis for *Vibrio alginolyticus* viable but nonculturable state determination," *Cytometry Part A*, vol. 69, no. 4, pp. 260–265, 2006.
- [34] E. Vlodayvsky, E. Palzur, and J. F. Soustiel, "Hyperbaric oxygen therapy reduces neuroinflammation and expression of matrix metalloproteinase-9 in the rat model of traumatic brain injury," *Neuropathology and Applied Neurobiology*, vol. 32, no. 1, pp. 40–50, 2006.
- [35] K. A. Giuliano, "Dissecting the individuality of cancer cells: the morphological and molecular dynamics of single human glioma cells," *Cell Motility and the Cytoskeleton*, vol. 35, no. 3, pp. 237–253, 1996.
- [36] B. Rousset, Y. Munari-Silem, V. Gire, and P. Fonlupt, "Dynamic analysis of drug action on in vitro reconstituted thyroid follicle by microinjection of tracer molecules and videomicroscopy," *Cell Biology and Toxicology*, vol. 8, no. 3, pp. 1–7, 1992.
- [37] C. Broglio, J. Dufer, P. Joly, Y. Carpentier, and A. Desplaces, "Quantitative morphological assessment of erythroblastic differentiation induced, in vitro, in human K562 leukemic cells," *Analytical Cellular Pathology*, vol. 5, no. 3, pp. 135–146, 1993.
- [38] L. Vega-Alvarado, J. Márquez, and G. Corkidi, "Inter-chromosome texture as a feature for automatic identification of metaphase spreads," *Medical and Biological Engineering and Computing*, vol. 40, no. 4, pp. 479–484, 2002.
- [39] N. Kutsuna and S. Hasezawa, "Morphometrical study of plant vacuolar dynamics in single cells using three-dimensional reconstruction from optical sections," *Microscopy Research and Technique*, vol. 68, no. 5, pp. 296–306, 2005.
- [40] D. A. Winkelman, T. S. Baker, and I. Rayment, "Three-dimensional structure of myosin subfragment-1 from electron microscopy of sectioned crystals," *Journal of Cell Biology*, vol. 114, no. 4, pp. 701–713, 1991.
- [41] T. G. Li, S. P. Wang, and N. Zhao, "Gray-scale edge detection for gastric tumor pathologic cell images by morphological analysis," *Computers in Biology and Medicine*, vol. 39, no. 11, pp. 947–952, 2009.
- [42] J. Poikonen and A. Paasio, "An 8 × 8 cell analog order-statistic-filter array with asynchronous grayscale morphology in 0.13- μ m CMOS," *IEEE Transactions on Circuits and Systems I*, vol. 56, no. 8, pp. 1541–1553, 2009.
- [43] J. Angulo and S. Matou, "Application of mathematical morphology to the quantification of in vitro endothelial cell organization into tubular-like structures," *Cellular and Molecular Biology*, vol. 53, no. 2, pp. 22–35, 2007.
- [44] D. C. Walker, B. H. Brown, A. D. Blacket, J. Tidy, and R. H. Smallwood, "A study of the morphological parameters of cervical squamous epithelium," *Physiological Measurement*, vol. 24, no. 1, pp. 121–135, 2003.
- [45] F. Ortiz, F. Torres, E. De Juan, and N. Cuenca, "Colour mathematical morphology for neural image analysis," *Real-Time Imaging*, vol. 8, no. 6, pp. 455–465, 2002.
- [46] S. Li, K. H. Hu, N. Cai et al., "Automatic analysis of image of surface structure of cell wall-deficient EVC," *Bio-Medical Materials and Engineering*, vol. 11, no. 3, pp. 159–166, 2001.
- [47] V. Metzler, T. Lehmann, H. Bienert, K. Mottaghy, and K. Spitzer, "Scale-independent shape analysis for quantitative cytology using mathematical morphology," *Computers in Biology and Medicine*, vol. 30, no. 3, pp. 135–151, 2000.
- [48] G. Diaz, A. Zucca, M. D. Setzu, and C. Cappai, "Chromatin pattern by variogram analysis," *Microscopy Research and Technique*, vol. 39, no. 3, pp. 305–311, 1997.
- [49] S. Holmes, A. Kapelner, and P. P. Lee, "An interactive java statistical image segmentation system: gemIdent," *Journal of Statistical Software*, vol. 30, no. 10, pp. 1–20, 2009.
- [50] J. B. Hendricks, "Quantitative histology by laser scanning cytometry," *Journal of Histotechnology*, vol. 24, no. 1, pp. 59–62, 2001.
- [51] E. Ficarra, S. Di Cataldo, A. Acquaviva, and E. Macii, "Automated segmentation of cells with IHC membrane staining," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1421–1429, 2011.
- [52] K. S. Cheng, C. J. Chien, M. H. Hsu, and C. L. Li, "Development of a PC12 cell cultivation and monitoring system for neuronal-like study," *Biomedical Engineering*, vol. 22, no. 3, pp. 193–203, 2010.
- [53] O. Schmitt and M. Hasse, "Morphological multiscale decomposition of connected regions with emphasis on cell clusters,"

- Computer Vision and Image Understanding*, vol. 113, no. 2, pp. 188–201, 2009.
- [54] O. Schmitt and M. Hasse, “Radial symmetries based decomposition of cell clusters in binary and gray level images,” *Pattern Recognition*, vol. 41, no. 6, pp. 1905–1923, 2008.
- [55] C. López, M. Lejeune, M. T. Salvadó et al., “Automated quantification of nuclear immunohistochemical markers with different complexity,” *Histochemistry and Cell Biology*, vol. 129, no. 3, pp. 379–387, 2008.
- [56] P. Thurner, R. Müller, G. Raeber, U. Sennhauser, and J. A. Hubbell, “3D morphology of cell cultures: a quantitative approach using micrometer synchrotron light tomography,” *Microscopy Research and Technique*, vol. 66, no. 6, pp. 289–298, 2005.
- [57] G. Landini and I. E. Othman, “Architectural analysis of oral cancer, dysplastic, and normal epithelia,” *Cytometry Part A*, vol. 61, no. 1, pp. 45–55, 2004.
- [58] R. Nakajima, T. Nakamura, M. Ogawa, H. Miyakawa, and Y. Kudo, “Novel method for quantification of brain cell swelling in rat hippocampal slices,” *Journal of Neuroscience Research*, vol. 76, no. 5, pp. 723–733, 2004.
- [59] M. Kruk, S. Osowski, and R. Koktysz, “Recognition and classification of colon cells applying the ensemble of classifiers,” *Computers in Biology and Medicine*, vol. 39, no. 2, pp. 156–165, 2009.
- [60] F. Y. Shih, C. T. King, and C. C. Pu, “Pipeline architectures for recursive morphological operations,” *IEEE Transactions on Image Processing*, vol. 4, no. 1, pp. 11–18, 1995.
- [61] J. L. Humm, R. M. Macklis, Y. Yang, K. Bump, and L. M. Chin, “Image analysis for the study of radionuclide distribution in tissue sections,” *Journal of Nuclear Medicine*, vol. 35, no. 7, pp. 1217–1225, 1994.
- [62] S. Jiang, X. Zhou, T. Kirchhausen, and S. T. C. Wong, “Detection of molecular particles in live cells via machine learning,” *Cytometry Part A*, vol. 71, no. 8, pp. 563–575, 2007.
- [63] K. Jiang, Q. M. Liao, and Y. Xiong, “A novel white blood cell segmentation scheme based on feature space clustering,” *Soft Computing*, vol. 10, no. 1, pp. 12–19, 2006.
- [64] H. S. Wu, R. Xu, N. Harpaz, D. Burstein, and J. Gil, “Segmentation of intestinal gland images with iterative region growing,” *Journal of Microscopy*, vol. 220, no. 3, pp. 190–204, 2005.
- [65] G. Haroske, V. Dimmer, K. Friedrich et al., “Nuclear image analysis of immunohistochemically stained cells in breast carcinomas,” *Histochemistry and Cell Biology*, vol. 105, no. 6, pp. 479–485, 1996.
- [66] J. M. Higgins, D. T. Eddington, S. N. Bhatia, and L. Mahadevan, “Statistical dynamics of flowing red blood cells by morphological image processing,” *PLoS Computational Biology*, vol. 5, no. 2, Article ID e1000288, 2009.
- [67] M. Garcia-Bonafe and A. Moragas, “Differential diagnosis of malignant and reactive cells from serous effusions: image and texture analysis study,” *Analytical Cellular Pathology*, vol. 12, no. 2, pp. 85–98, 1996.
- [68] C. Li, C. Xu, C. Gui, and M. D. Fox, “Level set evolution without re-initialization: a new variational formulation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’05)*, pp. 430–436, June 2005.
- [69] A. Tsai, A. Yezzi, and A. S. Willsky, “Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification,” *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1169–1186, 2001.
- [70] T. Chan and L. Vese, “Active contours without edges,” *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [71] Asa Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, “Support vector machine clustering,” *Journal of Machine Learning Research*, vol. 2, pp. 125–137, 2001.
- [72] H. Xue, S. Chen, and Q. Yang, “Structural support vector machine,” *Lecture Notes in Computer Science*, vol. 5263, no. 1, pp. 501–511, 2008.
- [73] G. Zellnig, A. Perktold, and B. Zechmann, “Fine structural quantification of drought-stressed *Picea abies* (L.) organelles based on 3D reconstructions,” *Protoplasma*, vol. 243, no. 1–4, pp. 129–136, 2010.
- [74] J. M. Messerli, M. E. Eppenberger-Eberhardt, B. M. Rutishauser et al., “Remodelling of cardiomyocyte cytoarchitecture visualized by three-dimensional (3D) confocal microscopy,” *Histochemistry*, vol. 100, no. 3, pp. 193–202, 1993.
- [75] H. A. McNally and R. B. Borgens, “Three-dimensional imaging of living and dying neurons with atomic force microscopy,” *Journal of Neurocytology*, vol. 33, no. 2, pp. 251–258, 2004.
- [76] D. Comaniciu, P. Meer, and D. J. Foran, “Image-guided decision support system for pathology,” *Machine Vision and Applications*, vol. 11, no. 4, pp. 213–224, 1998.
- [77] D. Ristanović, N. T. Milošević, I. B. Stefanović, D. Marić, and I. Popov, “Cell image area as a tool for neuronal classification,” *Journal of Neuroscience Methods*, vol. 182, no. 2, pp. 272–278, 2009.

Research Article

Remotely Sensed Image Classification by Complex Network Eigenvalue and Connected Degree

Mengxi Xu¹ and Chenglin Wei²

¹ School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China

² Nanjing Rail Traffic Technology Company, Department of Electrical and Mechanical Control, NARI Technology Development Co., Ltd., Nanjing 210061, China

Correspondence should be addressed to Mengxi Xu, mengxi.xu@gmail.com

Received 29 July 2011; Accepted 22 September 2011

Academic Editor: Sheng-yong Chen

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

It is a well-known problem of remotely sensed images classification due to its complexity. This paper proposes a remotely sensed image classification method based on weighted complex network clustering using the traditional K -means clustering algorithm. First, the degree of complex network and clustering coefficient of weighted feature are used to extract the features of the remote sensing image. Then, the integrated features of remote sensing image are combined to be used as the basis of classification. Finally, K -means algorithm is used to classify the remotely sensed images. The advantage of the proposed classification method lies in obtaining better clustering centers. The experimental results show that the proposed method gives an increase of 8% in accuracy compared with the traditional K -means algorithm and the Iterative Self-Organizing Data Analysis Technique (ISODATA) algorithm.

1. Introduction

Remote sensing is an all-round detective technology rose and rapidly developed from 1960s, which shows its superiority in urban planning, resources exploration, environment protection, land monitoring, agriculture, forestry, military and so on, and still develops its applications in breadth and depth.

Remotely sensed image classification is an important issue in remote sensing technique applications, whose goal is to classify the pixels in the remotely sensed images by ground-object categories. For example, the images are divided into many districts which represent forest, grass, lake, town, and other ground-object categories. Remotely sensed image classification can be carried on according to the following steps. First, feature parameters are analyzed and chosen according to the spectral characteristics of each ground-object. Second, feature space is divided into non-overlapping subspaces. Then, each pixel vector in the images is assigned to each sub-space.

The classification of remotely sensed image is divided into supervised classification and unsupervised classification. The basic principle of supervised classification is to

determine discriminant functions and corresponding criterion according to prior knowledge of classification, and the progress is to determine undetermined parameters in discriminant functions by taking advantage of a certain amount of samples' observed values in known classifications, which is called learning, then, the samples' observed values of unknown classification are put into discriminant functions, and the sample's classification is determined according to the criterion. There are several kinds of commonly used supervised classification approaches. Minimum range classification is to use the distance in feature space to express element data and the similar degree of classified category characteristic. After each category characteristic parameter by the training data is obtained, the distance between the unknown element and each eigenvector or eigenvector represented is first calculated, and then the unknown element is assigned to the category with the least distance. Maximum likelihood classification is to calculate the likelihood of each pixel point by point and this pixel is assigned to the category corresponding to the maximum likelihood. This classification precision is high, but the assorting process is

complex and the computing time is long. Parallelepiped classification carries on the classification with a simple decision rule to the remotely sensed data. Decision boundary in the image data space forms an N -dimensional parallelepiped. Parallelepiped's dimension is determined by the threshold value of standard deviation of classification's mean value for each category. If the element is between the low threshold value and the high threshold value for N bands, then it belongs to this kind. If the element value falls in many kinds, then this element will turn over to the last match type. Decision tree classification follows the guiding principle of the hierarchical classification idea. The hierarchical classification idea aims at each kind of ground-object's different information characteristic and decomposes it according to certain principle layer by layer. In each decomposition process, according to the different subdomain characteristic and the prior knowledge, the researcher may choose the different bands or the band combination for classification. Based on images' different characteristics, the decision tree classification produces the rule and discovers the law by the tree structure expression classification or the decision-making set. First, discriminant function is produced using the training space entity set. Second, lower level points and branches in each branch subset are repeatedly established according to different values. Finally, the decision tree is built. This method is flexible, intuitive, clear, vigorous, and healthy and the operation efficiency is high. The decision tree classification approach has highlighted vegetation's spatial distribution, particularly the sparse vegetation's spatial distribution, and the vegetation and the nonvegetation region might use the threshold value to separate, which increases the classified precision.

The nonsupervised classification [1–3] refers to the assorting process without exerting any prior knowledge and carries on the classification naturally according to the remote sensing spectrum's characteristic. The commonly used nonsupervised classification approaches mainly include the K -means algorithm and the ISODATA algorithm.

The K -means algorithm's basic philosophy is through the iteration which moves each kind of classifications' centers until obtaining the best cluster result. The ISODATA sorting algorithm clusters the element data with the smallest spectrum formula. It starts from the random cluster mean value, computes the distance of the element and the mean value, and assigns the individual to the recent category. Every time the cluster repeats, the cluster mean value changes one time, and the new mean value is taken as the next-time cluster circulation. The ISODATA cluster is finished until either the greatest number is repeated or the largest percentage of invariable element between two repeated times is achieved.

The K -means algorithm and the ISODATA algorithm mainly have two different aspects: (1) each time the K -means algorithm adjusts a category of sample, each kind of sample's average value is recalculated, which is called one-by-one sample correction method. Instead, the ISODATA algorithm recalculates each kind of sample's average value after adjusting all samples, which is called batches of sample classifications; (2) the K -means algorithm may complete

the sample cluster analysis by adjusting sample attribute category, but the ISODATA algorithm can not only complete the sample cluster analysis through the adjustment of sample attribute category, but also automatically carry on the merge and the fission of the category, which has a quite reasonable cluster result of the class number.

The above classification approaches have been used for the mid- and low-resolution satellite remotely sensed image, which processes high spectrum in the universal applications, while the high-resolution remotely sensed image actually has many insufficiencies now.

This paper proposes one kind of new remotely sensed image classification method by the combination of the complex network architecture characteristic and the K -means cluster algorithm idea. The proposed algorithm constructs the weighted network synthesis characteristic value according to the similarity formula and then classifies by the K -means cluster algorithm's idea to find the most superior category division.

In this paper, the first section is the introduction. The main principle of the complex network is briefly outlined in the second section, the remotely sensed image classification method based on complex network is proposed in the third section, and the fourth section shows simulated experiments and results analysis. Finally, the fifth section gives the conclusion and prospect.

2. Complex Networks

Since the end of 20th century, complex network [4, 5] has emerged gradually and overlapped rapidly with other disciplines in the depth and the breadth. Specially, two groundbreaking international works have started not a small upsurge of researching complex network. First, Watts and Strogatz [6] published the article in the *Nature* magazine in 1998, which introduced the microcosm network model, and described the transformation from the completely regular network to the completely random network. The microcosm network both has cluster characteristic similar to the regular network and small average path length similar to the random network. Second, Barabasi and Albert [7] published the article in *Science* to point out that many actual complex network's connection distributions have the power law form in 1999. As the power law distribution has no obvious characteristic length, this kind of network is called the scale-free network [8]. Then the scientists have studied each characteristic of each kind of complex network. The domestic researchers also already noted this tendency and also started to launch the research. The scholars joining the complex network research mainly come from domains such as graph theory, statistical physics, computer network research, ecology, sociology as well as economics. The networks involved mainly include networks in life sciences domain (e.g., cellular network, protein-protein function network, protein fold network, neural network, ecology networks), the Internet/WWW networks, the social networks which include the dissemination network of the epidemic disease, the scientist cooperative network, human

sexual relationship network, and linguistics network. The main methods used in the networks are graph theory in mathematics, the statistical physics method in physics, and the society network analysis method. However, except the example of medical images applications [9], the complex network analysis is not applied in the remotely sensed image classification in domestic [10].

2.1. Complex Network and Key Characteristics. The statistical properties of the complex network architecture mainly have the average path length, the clustering coefficient, degree distribution, and betweenness.

2.1.1. Average Path Length. In the network, the distance d_{ij} between the two nodes i and j is defined as

$$d_{ij} = \sqrt{(i-j)^T(i-j)}. \quad (1)$$

The average path length L in the network is defined as the mean value of the distance between the random two nodes, namely,

$$L = \frac{1}{(1/2)N(N-1)} \sum_{i \geq j} d_{ij}, \quad (2)$$

where N is the number of the network nodes. The average path length of the network is also called the network characteristic path length [11].

2.1.2. Clustering Coefficients. Generally, assume the node i in the network has n edges to be connected with other n nodes, then these n nodes are called the neighbors of the node i . Obviously, there are $n(n-1)/2$ edges at most. But between these n nodes, the ratio of the actually existing m edges and the possibly existing edges is defined as the clustering coefficient of node i , namely,

$$C = \frac{2m}{n(n-1)}. \quad (3)$$

2.1.3. Distribution among Degrees. Degree is a simple and important concept in the independent node attribute. Node i is defined as another node number connecting this node. In network the mean value of all nodes' degrees is called the network average degree and is recorded as k . In network distributed situation of nodes' degrees can be described by available distribution function $P(k)$, which expresses the probability that the randomly designated node's degree is k .

2.1.4. Betweenness. In the complex network, there are some nodes which are not very large, but they are actually significant like a bridge in the entire structure [12]. The betweenness of the nodes has reflected the bridge ability size of the nodes.

Let n_{st} expresses the number of the most short path from apex s to t , the $n_{st}(i)$ expresses apex i 's number of the most short path from s to t , then the betweenness of the node i is

$$C_B(v) = \sum_{s \neq t \neq i \in V} \frac{n_{st}(i)}{n_{st}}. \quad (4)$$

Node accumulation coefficient [13] has reflected the network module nature. Namely, the interconnection of the interior node in the identical module is high, while the accumulation is strong, but the node accumulation coefficient between modules is weak. The node accumulation coefficient manifests the local interconnection density of this node, but for the complex network, the node's connection represents some similarity in the node attribute, therefore, the accumulation degree and the accumulation coefficient can be used as the characteristic to cluster the network node.

2.2. Weighted Complex Network and Its Key Characteristics. At present the research for complex network mainly aims at the unweighted complex network. But in the realistic network, the weights of the edges are often dissimilar and will affect the performance of the entire network. The weighted complex network can better express the structure of the complex network than unweighted complex network. Comparing to the definition of the degree and the accumulation coefficients above, the definition of the nodes' weighted degree, the weighted accumulation, and the weighted accumulation coefficient below in the weighted complex network are given [4].

2.2.1. Weighted Degree. A node's degree may be defined as the sum of the weights of this node and its all neighboring nodes, which is also called weighted degree, namely, the weighted degree WD_i of the node i is defined as follows:

$$WD_i = \sum w_{ij}, \quad (5)$$

where w_{ij} ($w_{ij} > 0$) expresses the weighted coefficients between node i and j .

The node's weighted degree reflects the joint strength between this node and other nodes. The larger the node's weighted degree, the more this node is suitable to be the cluster center.

2.2.2. Weighted Clustering Coefficient. The clustering coefficient WC_i of the node i is

$$WC_i = \frac{2 \sum w_{jk}}{D_i(D_i - 1)}, \quad (6)$$

where D_i express the degree of node i , which is defined as number of edges that is connected with node i , and j and k are neighboring nodes of node i .

The node's weighted clustering coefficient manifests local interconnection density and the intensity of this node. The larger the node's weighted clustering coefficient is, the more the node is suitable to be the cluster center.

2.2.3. Weighted Network Synthesis Characteristic Value. The weighted network synthesis characteristic value of the node i is

$$WCF_i = aWC_i + \frac{(1-a)WD_i}{N}, \quad (7)$$

where N is the node number in the network, a is the adjustable parameter, $0 < a < 1$. Regarding the different application background, a takes the different value, which indicates that weighted degree and the weighted clustering coefficient occupy different proportion in the weighted network synthesis characteristic value. The larger the a is, the bigger proportion the weighted clustering coefficient in the weighted network synthesis characteristic value occupies, and the smaller the a is, the bigger proportion the weighted degree in the weighted network synthesis characteristic value occupies.

2.2.4. Connected Degree. The connection r_{ij} between the node i and j in complex network is

$$r_{ij} = \frac{\langle \dot{i} \dot{j} \rangle - \langle \dot{i} \rangle \langle \dot{j} \rangle}{\left[(\langle \dot{i} \rangle - \langle \dot{i} \rangle^2) (\langle \dot{j} \rangle - \langle \dot{j} \rangle^2) \right]^{1/2}}, \quad (8)$$

where $\langle i \rangle$ expresses the mean value to corresponding element in node i . The size of r_{ij} can be used to weigh the connection between node i and j [10]; the vector of node i is \dot{i} and the vector of node j is \dot{j}

3. Remotely Sensed Image Classification Approach Based on Complex Networks

For better cluster realization to remotely sensed image classification, overcoming sensitive shortcoming of the K -means [14, 15] algorithm to the initial cluster center, we first choose initial cluster center according to the eye measurement, then cluster center selection of each iteration follows the sorting of the nodes' WCF size, and the high node of WCF is taken as initial cluster center (the connection between cluster center later chosen and present cluster center should be smaller than the given threshold value T). Such cluster center selected has strong joint strength and strong local accumulation nature which is similar to other nodes, moreover, the probability of the same kind is relatively small among the cluster centers, which reduces the iteration times of the algorithm.

In this paper, a vector which is composed of the pixels from the same location of each band is considered as a node. The similarity between nodes is taken as the weighted degree, which represents the connected degree of these two nodes. In addition, the weak connected edge that the value of weighted degree is smaller than the threshold is deleted.

The Kappa coefficient here is used to measure the agreement between two raters who each classify N items into C mutually exclusive categories. The equation for Kappa coefficient (κ) is

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \quad (9)$$

where $\Pr(a)$ is the relative observed agreement among raters, and $\Pr(e)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. If the raters are in complete agreement then $\kappa = 1$. If there is no agreement

among the raters other than what would be expected by chance (as defined by $\Pr(e)$), $\kappa = 0$ [16].

Input: remotely sensed image.

Output: image classification, classified precision, Kappa coefficient.

Step 1. The hyperspectral image file is read and pretreated. These include radiation adjustment processing, geometry correction processing, mosaic processing, and cutting out processing [17, 18].

Step 2. The band data is chosen to do the experiment according to the standard deviation and corresponding coefficients.

Step 3. The weighted degree WD based on the formula (5) is calculated, clustering coefficient WC based on the formulas (5) and (6) is calculated, and complex network synthesis characteristic value WCF again based on the formula (7) is computed.

Step 4. The k maximal complex network synthesis characteristic values are chosen from the results computed in Step 3 by the top- k algorithm which chooses k maximal numbers using one-dimensional array.

Step 5. The connected degree between pixel nodes is calculated based on the formula (8).

Step 6. The threshold T of connected degree between pixel complex network nodes is computed by maximum mean square error.

Step 7. The connected degree between k selected nodes in Step 4 and initial cluster center is calculated, if the connected degree is smaller than threshold value T , then this node is taken as the cluster center.

Step 8. For the new cluster center, the distance between the sample and each new cluster center is computed and compared, and the sample is assigned to the class with the smallest distance.

Step 9. For the new class, the cluster center is recalculated. If the results are completely the same with the previous results, then the assorting process ends. Otherwise, turn to Step 3.

The proposed algorithm can be depicted by the following flow chart (Figure 1).

4. Experiments and Results Analysis

4.1. Simulated Images Experiments. These experiments are used to validate the accuracy of remotely sensed image classification method based on the eigenvalue and connected degree of complex network and compare the classification accuracy with K -means classification, ISODATA classification.

The experiment designs simulated experimental images of three bands with Gaussian noise. Suppose the noise of each band is independent identically distributed Gaussian

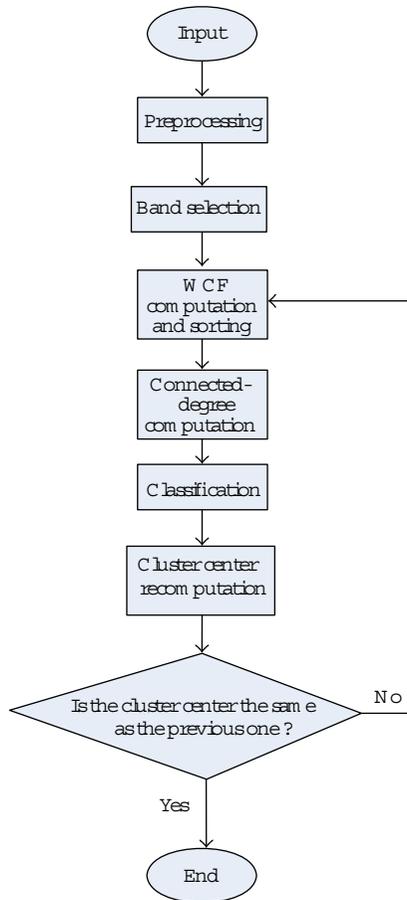


FIGURE 1: Flow chart of the proposed algorithm.

random noise, with zero mean and variance 0.01. The size of simulated experimental image is 128×128 , and the noise image is shown in Figure 2(a). Figures 2(b), 2(c), and 2(d) are the results of the K -means classification method, the ISODATA classification method, and the new method, respectively.

The classification accuracy percentage is firstly calculated, and the statistical results are shown in Table 1.

From Table 1, for the entire sample data, it can be seen that the classification accuracy is up to 99.5% by K -means classification method, 99.4% by ISODATA classification method, and 99.6% by the new method based on complex network. The experiments show that the classification accuracy by the new method based on complex network is better than that of both traditional K -means classification method and ISODATA classification method.

The Kappa coefficient of classified result by the new method is computed below.

Confusion matrix is firstly calculated, and the results are shown in Table 2. Then the Kappa coefficient, Kappa error, and maximum possible Kappa are calculated by confusion matrix. The Kappa coefficient is 0.994 and 0.993 for the K -means classification method and ISODATA classification method, respectively (given the limited space of the paper, tables are omitted here).

From the relationship between Kappa coefficient and classified accuracy in Table 2, proposed method based on complex network is excellent.

4.2. AVIRIS Images Experiments. The classification accuracy of algorithms is very important. In order to obtain more accurate classification precision, the parts of AVIRIS hyperspectral data are selected in the experiments, which is photographed in a remote sensing experimental plot of the northeast of the US Indiana in June 12, 1992 [19], and the band number is 220, and data comes from the website <http://engineering.purdue.edu/~biehl/MultiSpec>.

In the high spectrum remotely sensed image selected by the experiments, we should select the bands that are less polluted by the moisture noise because some bands are polluted seriously by the moisture noise. Band selection [20, 21] mainly depends upon two essential factors, the standard deviation and the correlation coefficient. The larger the band standard deviation is, the more information this band contains. The larger the correlation coefficient between two bands is, the more information similarity the two bands contain. The band selection steps are as follows.

Step 1. First the standard deviation of each band is calculated and arranged in descent order, as shown in Table 3. Due to limited space, here we only give the partial data (first 30 bands).

Step 2. Calculate the correlation coefficients (CCs) between two bands, as shown in Table 4. Due to limited space, we only give the partial data.

For the band having quite great similarity, we only need to choose one of the bands, between two essential factors the standard deviation and the correlation coefficient of the bands influenced, and we should choose bands with the larger standard deviation and smaller correlation coefficient between each other.

Step 3. According to the standard deviation in Step 1 and the correlation coefficients in Step 2, we choose band 42, band 29, and band 120 to do the experiments based on the principle that the standard deviation is as large as possible and correlation coefficient is as small as possible.

In order to compare the classification precision, the new algorithm, which is a nonsupervised classification approach, is compared with other two nonsupervised classification approaches, the K -means algorithm and the ISODATA algorithm. Finally the precision of the three kinds of classification results is compared. In order to count the uniformity of the classified precision, the classification number of the three classification approaches is all set to 3.

Figure 3(a) shows the real data of ground-object category, and Figure 3(b) shows the false color image synthesized by bands 42, 29, 120. The result with the K -means algorithm is shown in Figure 3(c). The result with the ISODATA algorithm is shown in Figure 3(d). The result with the new algorithm is shown in Figure 3(e), in which the value of a is set to 0.5.

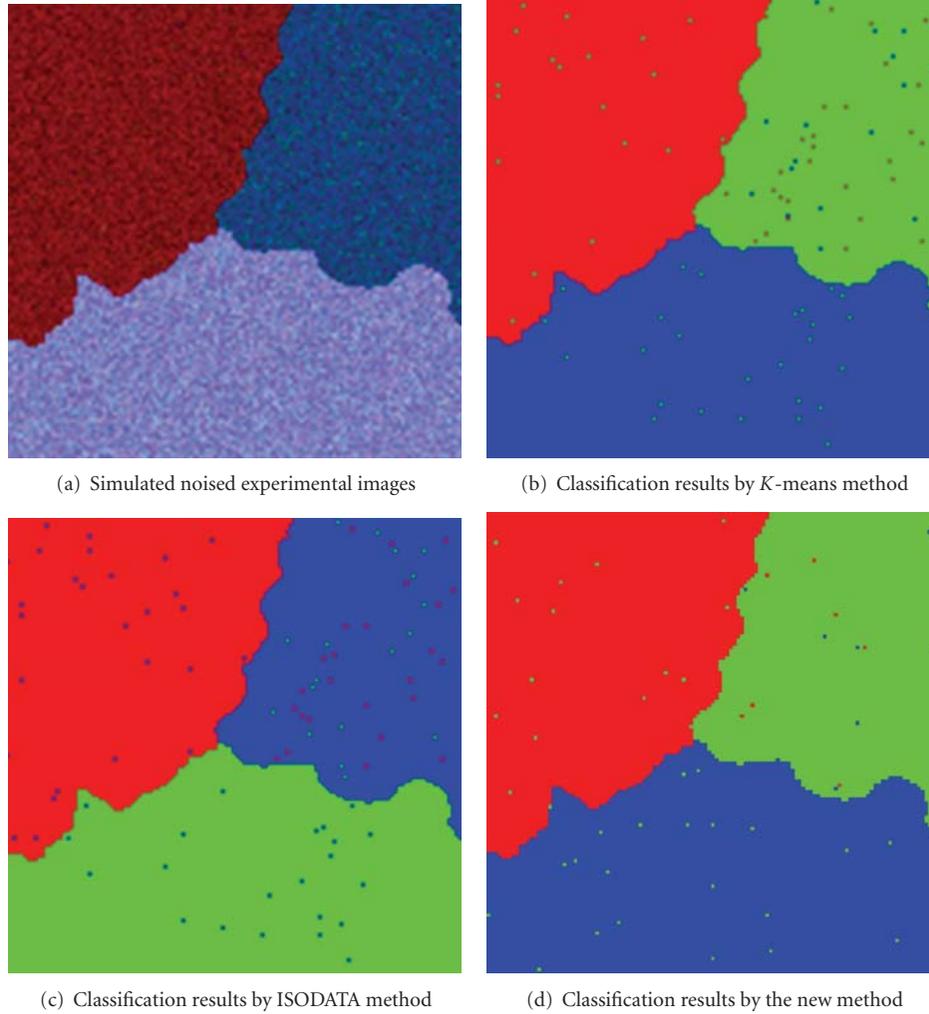


FIGURE 2: Classification results of simulated experimental images and different algorithms of three bands with added Gaussian noise.

TABLE 1: Classification accuracy percentage of simulated images by three classification methods.

| Project | Class 1 | Class 2 | Class 3 | Classification accuracy |
|-------------------------------------|---------|---------|---------|-------------------------|
| Ground truth | 4402 | 5482 | 6500 | 100% |
| K -means classification method | 4359 | 5461 | 6474 | 99.5% |
| ISODATA classification method | 4359 | 5455 | 6478 | 99.4% |
| New method based on complex network | 4387 | 5467 | 6477 | 99.6% |

The classification accuracy of simulated experimental image by each classified method is counted below. In this experiment the image pixels are chosen as size of 145×145 to count overall classification precision of the remotely sensed image. We extract the subset with size of 42×145 to compute the classification precision. The statistical results are shown in Table 5. From Table 5, the classification result

TABLE 2: Confusion matrix of classified results by proposed method based on complex network.

| Project | Class 1 | Class 2 | Class 3 |
|------------------------|---------|---------|---------|
| Class 1 | 4387 | 7 | 8 |
| Class 2 | 15 | 5467 | 0 |
| Class 3 | 23 | 0 | 6477 |
| Kappa coefficient | 0.9951 | | |
| Kappa error | 0.0007 | | |
| Maximum Possible Kappa | 0.9979 | | |

of the samples in category 1 by new algorithm is better than the K -means algorithm and the ISODATA algorithm; the result of the samples in category 2 by new algorithm is worse than the K -means algorithm and the ISODATA algorithm; the classification result of the samples in category 3 by new algorithm is much better than the K -means algorithm and the ISODATA algorithm. For all samples, the classified precision of the K -means algorithm achieves 82%, the classified precision of the ISODATA algorithm achieves 82%,

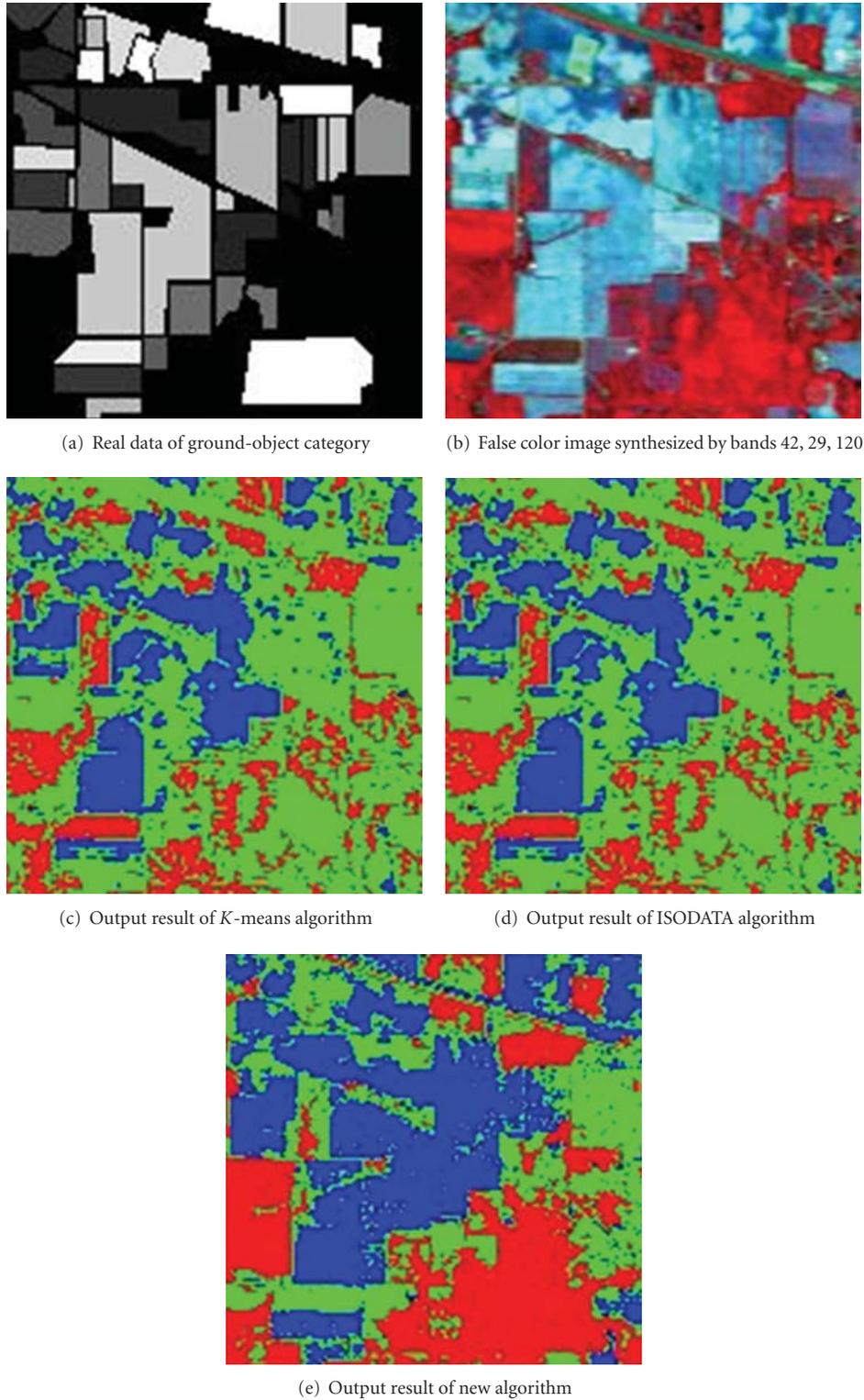


FIGURE 3: Raw remotely sensed data and classified results by different algorithms.

but the classified precision of the new algorithm achieves 90.4%. The experiment shows that the classified precision of the new algorithm increases by 8%, compared to the traditional *K*-means algorithm and the ISODATA algorithm.

The Kappa coefficient of the new algorithm based on complex network is calculated as follows: first the confusion matrix is calculated, which is shown as Table 6. In order to reduce the operand we only calculate the partial data.

TABLE 3: First 30 bands and standard deviation descending according to the standard deviation.

| Band | Standard deviation | Band | Standard deviation | Band | Standard deviation |
|---------|--------------------|---------|--------------------|---------|--------------------|
| Band 29 | 1012.186414 | Band 33 | 866.810154 | Band 50 | 771.726478 |
| Band 28 | 995.72999 | Band 23 | 864.978506 | Band 20 | 764.574815 |
| Band 27 | 932.214293 | Band 31 | 856.363198 | Band 45 | 757.351603 |
| Band 26 | 932.1544 | Band 43 | 853.999667 | Band 51 | 755.848178 |
| Band 25 | 910.697106 | Band 22 | 839.800843 | Band 52 | 754.248011 |
| Band 30 | 908.840288 | Band 44 | 837.885813 | Band 34 | 742.718947 |
| Band 42 | 907.544127 | Band 39 | 807.049573 | Band 19 | 731.861102 |
| Band 32 | 898.713883 | Band 21 | 797.873879 | Band 53 | 727.076556 |
| Band 41 | 884.337735 | Band 48 | 788.812468 | Band 38 | 720.686685 |
| Band 24 | 875.553739 | Band 49 | 775.506589 | Band 47 | 719.547723 |

TABLE 4: The correlation matrix of partial band.

| CCs | Band | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | Band 29 | Band 30 | Band 31 | Band 32 | Band 33 | Band 34 | Band 35 | Band 36 | Band 37 |
| Band 29 | 1 | 0.99 | 0.99 | 0.99 | 0.99 | 0.96 | 0.87 | 0.47 | -0.16 |
| Band 30 | 0.99 | 1 | 0.99 | 0.99 | 0.99 | 0.97 | 0.88 | 0.47 | -0.15 |
| Band 31 | 0.99 | 0.99 | 1 | 0.99 | 0.99 | 0.98 | 0.91 | 0.51 | -0.12 |
| Band 32 | 0.99 | 0.99 | 0.99 | 1 | 0.99 | 0.97 | 0.88 | 0.47 | -0.15 |
| Band 33 | 0.99 | 0.99 | 0.99 | 0.99 | 1 | 0.98 | 0.91 | 0.52 | -0.11 |
| Band 34 | 0.96 | 0.97 | 0.98 | 0.97 | 0.98 | 1 | 0.96 | 0.63 | 0.01 |
| Band 35 | 0.87 | 0.88 | 0.91 | 0.88 | 0.91 | 0.96 | 1 | 0.81 | 0.26 |
| Band 36 | 0.47 | 0.47 | 0.51 | 0.47 | 0.52 | 0.63 | 0.81 | 1 | 0.77 |
| Band 37 | -0.16 | -0.15 | -0.12 | -0.15 | -0.11 | 0.01 | 0.26 | 0.77 | 1 |

Note. In Table 3 the standard deviation of band 29 is the greatest. In order to express the characteristics of correlation coefficients, the band listed in Table 4 begins with band 29.

TABLE 5: Classification statistics of partial data.

| Item | Category 1 | Category 2 | Category 3 | Classified precision |
|--|------------|------------|------------|----------------------|
| Ground data | 966 | 3446 | 1678 | 100% |
| <i>K</i> -means | 423 | 3525 | 2142 | 82% |
| ISODATA | 418 | 3487 | 2185 | 82% |
| New algorithm based on complex network | 1251 | 3153 | 1684 | 90.4% |

The three digits of the second line in Table 4 present the numbers of the samples which should originally belong to category 1 but are classified to category 1, category 2, and category 3, respectively. The three digits of the third line in Table 4 present the numbers of the samples which should originally belong to category 2 but are classified to category 1, category 2, and category 3, respectively. The three digits of the fourth line in Table 6 present the numbers of the samples which should originally belong to category 3 but are classified to category 1, category 2, and category 3, respectively. Using the confusion matrix already obtained we calculate the Kappa coefficient, Kappa error, and Maximum possible Kappa which is shown in Table 6.

The merits of this algorithm are illustrated as follows. This paper chooses well the initial cluster center according to the connection among nodes and the nodes' weighted network synthesis characteristic value and overcomes sensitive

TABLE 6: Confusion matrix of classification results by new algorithm based on complex network.

| Item | Category 1 | Category 2 | Category 3 |
|------------------------|------------|------------|------------|
| Category 1 | 883 | 315 | 139 |
| Category 2 | 192 | 2377 | 121 |
| Category 3 | 176 | 461 | 1424 |
| Kappa Coefficient | 0.6353 | | |
| Kappa error | 0.0085 | | |
| Maximum possible Kappa | 0.8797 | | |

shortcoming of the *K*-means cluster algorithm to the initial value, thus, it enhances greatly the cluster quality. It is easy to find that the method in this paper is better than that of traditional methods either from the classified accuracy or the cluster stability.

- It overcomes the sensitive shortcoming of the *K*-means cluster algorithm to the initial value, which is easily fallen into the local region.
- According to the connection of the node, cluster center is selected, which reduces the probability of selecting different nodes of the same kind as the cluster center, reduced the iterative times of the algorithm, and raised the algorithm efficiency.

5. Conclusion

This paper proposes a remotely sensed image classification approach based on the complex network eigenvalue and connected-degree combining weighted complex network synthesis characteristic value with K -means cluster algorithm. Simulated experiments classify simulated experimental image of 3 bands with added Gaussian noise and AVIRIS hyperspectral remotely sensed data photographed in a remote sensing experimental plot of the northeast of the US Indiana in June 12, 1992, where the classified accuracy is improved compared to traditional K -means method and ISODATA method. Although the new method proposed carries on the classification well to the remotely sensed image, as a result of the own complexity of remotely sensed image, it needs deep research to choose the optimal characteristic and improve the accuracy of remotely sensed image classification.

References

- [1] Y. Bazi and F. Melgani, "Gaussian process approach to remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 186–197, 2010.
- [2] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, Academic Press, 3rd edition, 2007.
- [3] 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.
- [4] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [6] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [7] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [8] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, Article ID 026113, 15 pages, 2004.
- [9] F. D. Xie and F. X. Zhao, "Research of medical image classify based on weighted complex network cluster," *Computer Engineering and Design*, vol. 30, no. 17, pp. 40–57, 2009.
- [10] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416–1427, 2008.
- [11] B. Yang, D. Y. Liu, J. Liu, D. Jin, and H. B. Ma, "Complex network clustering algorithms," *Ruan Jian Xue Bao/Journal of Software*, vol. 20, no. 1, pp. 54–66, 2009.
- [12] F. X. Zhao, F. D. Xie, and M. Ji, "A new classification method based on complex network theory and genetic algorithm," *Computer Applications and Software*, vol. 27, no. 2, pp. 44–45, 2010.
- [13] H. Topan and H. S. Kutoglu, "Georeferencing accuracy assessment of high-resolution satellite images using figure condition method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, Article ID 4801656, pp. 1256–1261, 2009.
- [14] J. Zheng, Z. Cui, A. Liu, and Y. Jia, "A K -means remote sensing image classification method based on AdaBoost," in *the 4th International Conference on Natural Computation (ICNC '08)*, pp. 27–32, October 2008.
- [15] D. J. Liu and J. Yu, "Otsu method and K -means systems," in *the 9th International Conference on Hybrid Intelligent*, pp. 344–349, 2009.
- [16] N. C. Smeeton, "Early history of the kappa statistic," *Biometrics*, vol. 41, p. 795, 1985.
- [17] S. Y. Chen, Y. F. Li, Q. Guan, and G. Xiao, "Real-time three-dimensional surface measurement by color encoded light projection," *Applied Physics Letters*, vol. 89, no. 11, Article ID 111108, 2006.
- [18] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2011, Article ID 814356, 18 pages, 2011.
- [19] S. Das and T. T. Mirnalinee, "Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3906–3931, 2011.
- [20] M. Bai, H. Liu, W. Huang, Y. Qiao, and X. Mu, "Classification of remotely sensed imagery using adjacent features based approach," in *the 17th International Conference on Geoinformatics (Geoinformatics '09)*, August 2009.
- [21] M. J. Mendenhall and E. Merenyi, "Relevance-based feature extraction for hyperspectral images," *IEEE Transactions on Neural Networks*, vol. 19, no. 4, pp. 658–672, 2008.

Research Article

Influence of Expert-Dependent Variability over the Performance of Noninvasive Fibrosis Assessment in Patients with Chronic Hepatitis C by Means of Texture Analysis

Cristian Vicas,¹ Monica Lupsor,² Mihai Socaciu,² Sergiu Nedevschi,¹ and Radu Badea²

¹Computer Science Department, Faculty of Automation and Computer Science, Technical University of Cluj-Napoca, 28 Gheoghe Baritiu Street, 400027 Cluj-Napoca, Romania

²Regional Institute of Gastroenterology and Hepatology, Iuliu Hatieganu University of Medicine and Pharmacy, 19-21 Croitorilor Street, 400162 Cluj-Napoca, Romania

Correspondence should be addressed to Cristian Vicas, cristian.vicas@cs.utcluj.ro

Received 1 August 2011; Accepted 26 September 2011

Academic Editor: Carlo Cattani

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

Texture analysis is viewed as a method to enhance the diagnosis power of classical B-mode ultrasound image. The present paper aims to evaluate and eliminate the dependence between the human expert and the performance of such a texture analysis system in predicting the cirrhosis in chronic hepatitis C patients. 125 consecutive chronic hepatitis C patients were included in this study. Ultrasound images were acquired from each patient and four human experts established regions of interest. Textural analysis tool was evaluated. The performance of this approach depends highly on the human expert that establishes the regions of interest ($P < 0.05$). The novel algorithm that automatically establishes regions of interest can be compared with a trained radiologist. In classical form met in the literature, the noninvasive diagnosis through texture analysis has limited utility in clinical practice. The automatic ROI establishment tool is very useful in eliminating the expert-dependent variability.

1. Introduction

Noninvasive detection and staging of liver fibrosis have received more and more attention in scientific literature. One approach involves simple B-mode ultrasound in conjunction with textural analysis. The main assumption of the textural analysis approach is that fibrosis alterations at liver lobule level can induce significant changes in the speckle pattern of the ultrasound image [1]. Even if these alterations are not visible with the naked eye, a texture analysis system can detect and learn these alterations. Textural analysis is viewed as a method to enhance the diagnosis power of B-mode ultrasound by providing the physician with new information. This data can be otherwise inferred only by invasive methods.

The methodology presented in most of the papers [1–9] approaching textural analysis on B-mode ultrasound follows four general steps. First, a physician acquires a liver ultrasound image. Then, on the ultrasound image, another physician (or the same) establishes a rectangular region of interest (ROI). In the third step several textural algorithms

produce a feature vector. This vector is labeled according to biopsy findings. The fourth step implies the training of a classification schema. The resulting classifier can be used to predict fibrosis stages to unknown ultrasound images. In the first two steps there is a human expert that introduces an operator-dependent variability.

This paper addresses the user variability introduced by the second step, the establishment of the ROI. We also evaluate here a novel tool that automatically establishes the regions of interest. This tool was developed by our group and it was successfully applied in eliminating the expert-dependence in noninvasive steatosis quantification [10].

To our knowledge, the expert dependent variability in textural analysis for fibrosis detection was not addressed before. We included almost all the textural algorithms proposed in the literature as means of detecting liver fibrosis stages.

Present study aims to evaluate the dependence between the human expert and the performance of the texture analysis system in predicting cirrhosis in chronic hepatitis C patients.

2. Material and Methods

2.1. Patients. The local Ethical Committee of the University of Medicine and Pharmacy Cluj-Napoca approved this study. The patients provided written informed consent before the beginning of the study, in accordance to the principles of the Declaration of Helsinki (revision of Edinburgh, 2000). We prospectively included in this study 125 patients with hepatitis C infection having fibrosis stage 0 or 4 according to Metavir scoring system. Liver biopsy determined the fibrosis stages. This lot was selected from 1200 patients and was prospectively examined in Third Medical Clinic, Cluj-Napoca, Romania, between May 2007 and August 2009. All patients had positive HCV-RNA and underwent percutaneous liver biopsy (LB), in order to stage and grade their condition.

The exclusion criteria were presence of ascites at clinical or ultrasound examination, coinfection with HBV and/or HIV, other active infectious diseases, and pregnancy.

Alongside the epidemiological data, certain biological parameters were determined on a blood sample taken 12 hours after overnight fasting: alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transferase (GGT), total cholesterol, triglycerides, total bilirubin, and glycemia (Konelab 20i—Thermo Electron Corp., Finland).

2.2. Histopathological Analysis. A liver biopsy specimen was acquired using the TruCut technique with an 1.8 mm (14 G) diameter automatic needle device—Biopty Gun (Bard GMBH, Karlsruhe, Germany). The LB specimens were fixed in formalin and embedded in paraffin. The slides were evaluated by a single expert pathologist unaware of the clinical data. Only biopsy specimens with more than 6 intact portal tracts were eligible for evaluation [11]. The liver fibrosis and necroinflammatory activity were evaluated semiquantitatively according to the Metavir scoring system [12].

Fibrosis was staged on a 0–4 scale as follows: F0—no fibrosis; F1—portal fibrosis without septa; F2—portal fibrosis and few septa; F3—numerous septa without cirrhosis; F4—cirrhosis. The necroinflammatory activity was graded as A0—none; A1—mild; A2—moderate; A3—severe.

In present study, only patients having fibrosis stage 0 or 4 were included.

2.3. Ultrasound Examination. Each patient included in this study underwent an ultrasound examination using a GE Logiq 7 ultrasound machine (General Electric Company, Fairfield, England) with a 5.5 MHz convex phased array probe one day prior to liver biopsy. From each patient there were acquired right lobe ultrasound images with liver tissue without blood vessels or other artifacts with a depth setting of 16 cm using the same preestablished machine protocol. The acquisition protocol was established in such a way that we obtained a maximum amount of information from underlying tissue and in the same time keeping the noise level down. All postprocessing settings were set to minimum.

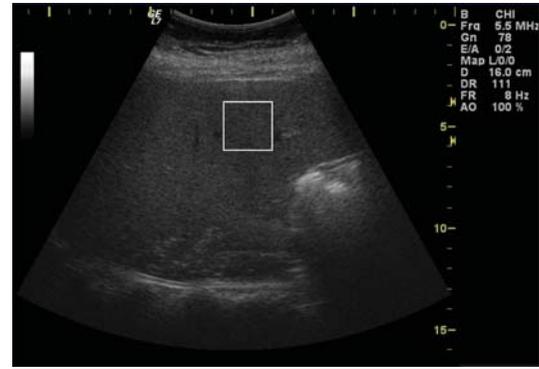


FIGURE 1: Right lobe ultrasound image. White square represents the region of interest.

The frame rate was kept as high as possible in order to avoid movement artifacts. The time gain compensation curve was set to neutral position. Once the device settings were established they were used to examine all the patients. Captured images were saved in DICOM format on the equipment's local hard drive. They were later transferred and processed on a personal computer.

2.4. Regions of Interest for Textural Analysis. The region of interest (ROI) establishment procedures followed the guidelines presented in the literature [1, 13]. The experts were instructed to choose one region of interest for each patient. The ROI had to be placed as close as possible to the vertical axis of the ultrasound image and at 1 cm below the liver capsule. The ROI had to avoid artifacts and anatomical features like blood vessels, liver capsule, shadowing, and so forth. The dimensions of the ROI were 64×64 pixels representing an area of 2.62×2.62 cm. Figure 1 shows an ultrasound image with an ROI. The physician acquired the image from the right live lobe.

In order to evaluate the user variability of the textural system, on the saved images, four experts with different skill level established the ROIs. The first two experts are trained radiologists with experience in gastrointestinal ultrasound investigation. First expert has more than 20 years in ultrasound investigation and the second more than 10 years. The third expert is a radiology intern with 2 years of experience. The fourth expert is a general practitioner trained in ultrasound examination. In addition to these experts, we employed an automated tool for establishing the region of interest. This tool establishes the ROI in a fixed position relative to the geometry of the image. Artifacts are detected using the method proposed in [10, 14]. After the artifacts are detected, we randomly choose a region of interest that has no artifacts. If such a region cannot be set in any of the patient's images, for the respective patient there will be no region of interest established.

The order of the patients and the order of the images for a patient were randomized. With this step we tried to avoid the influence of the image order over the performance detection. Algorithm 1 was used to ensure independent samples, it is graphically depicted in Figure 2.

Input: *patients*—a set with patients and ultrasound images
 Output: DS —a list with 25 sets DS_{ij} with regions of interest
 For $i = 1$ to 5
 D_i = Perform a randomization on *patients*
 For $j = 1$ to 5
 Sequentially present to expert j dataset D_i
 DS_{ij} = established ROIs

ALGORITHM 1: ROI establishment.

We computed the center of each region of interest in terms of Cartesian and polar coordinates. For the Cartesian system, the origin is the top left corner and for the polar system, the origin was considered the virtual source of ultrasound waves. Figure 3 sketches these coordinate systems.

2.5. Textural Analysis. In texture analysis there are two main steps [15]. The first step is the computation of several textural attributes that numerically describe the texture (using dedicated algorithms). The second step involves the training and evaluation of a classifier using the previously computed textural features.

Each texture description algorithm has a certain number of parameters that control the feature extraction process. For each algorithm implemented in the present study we used the same proposed set of parameters found in corresponding fibrosis detection papers. These algorithms are first-order statistics [4, 16], gray tone difference matrix [15], gray level co-occurrence matrix [1, 4, 16, 17], multiresolution fractal dimension [1], differential box counting [6, 18], morphological fractal dimension estimators [19], Fourier power spectrum [1, 13], Gabor filters [20], Law's energy measures [1], texture edge co-occurrence matrix [6], phase congruency-based edge detection [21], and texture feature coding matrix [22].

These 12 algorithms processed the entire ROI and computed 234 features per patient. Each feature vector was labeled with the corresponding histopathological finding as healthy or cirrhotic. From 25 sets of regions of interest we generated 25 sets of instances, each set containing one instance per patient.

The classification schema employed here was a logistic model [23–25]. The feature values were normalized in $[0, 1]$ interval prior to classification. Care was taken that the test subset was normalized with the same coefficients as the train set.

Before entering the classification schema, a feature selection process was applied. The relevant features were identified and selected using *correlation-based feature selection* (CFS) algorithm [26]. To avoid *overfitting* phenomena and to ensure that the feature selection step is independent of the underlying data, the following algorithm was applied.

- (1) From each of the 25 sets we selected k instances. These instances were randomly selected in such a way that each class has $k/2$ instances.
- (2) The selected instances were moved into another dataset.

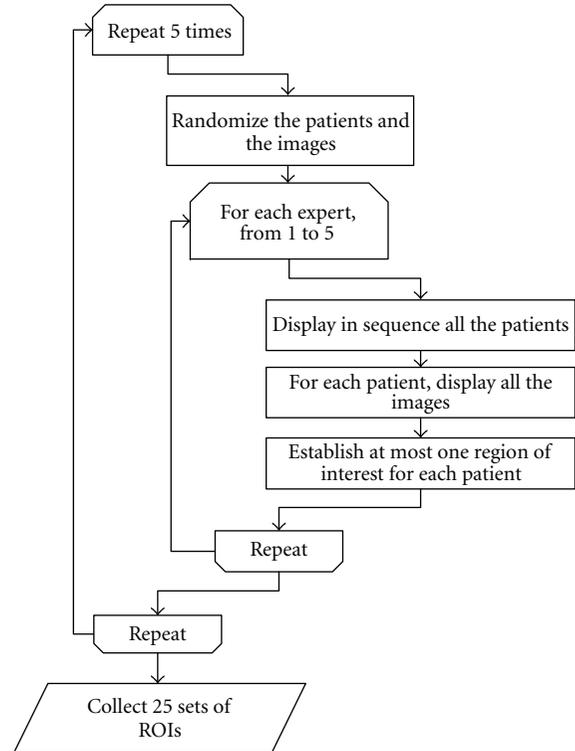


FIGURE 2: Algorithm for establishing the regions of interest. Each of 5 experts established 5 sets of ROIs. The automatic establishment algorithm was treated as a regular expert.

- (3) After 25 iterations we extracted $25 \times k$ instances.
- (4) On this $25 \times k$ dataset we applied the CFS algorithm. We noted the selected features and we processed the original datasets by keeping only relevant features.
- (5) The whole process was iterated 20 times.

For this paper, k was set to 10. The feature selection process is depicted in Figure 4.

The classifier performance estimation was determined using 10-fold stratified CV technique. The performance criterion was area under the curve (AUROC) computed on the collected predictions using Mann-Whitney-Wilcoxon U statistic [27]. In order to better estimate the average performance, the 10-fold CV procedure was iterated 10 times with random fold splitting [28].

The texture analysis system was validated using a set of known textures from Brodatz [29] library. Each image was divided into 100 nonoverlapping regions of interest. Each region has 64×64 pixels area. The textural analysis system was trained to predict the original image from where the region originated. The images were chosen following the guidelines in [15].

2.6. Statistical Analysis. Two-way ANOVA test was used to evaluate the performance variability. The dependent variable was set to be the average AUROC and the independent variables were the expert that established ROIs and the feature set obtained after the feature selection step. Tukey

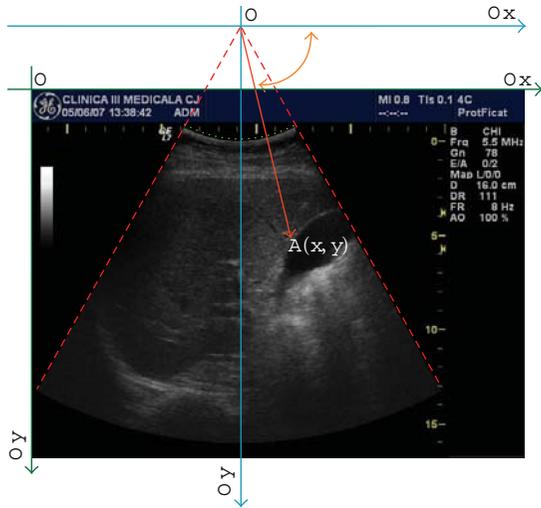


FIGURE 3: Cartesian system, Oxy (green lines) and polar system $O'x'y'$ (blue lines). An ROI center (A) has the $A(x, y)$ Cartesian coordinates and $A(\rho, \theta)$ polar coordinates. The green area represents the position of piezoelectric crystals in the probe and the red lines show the imaging aperture of the convex probe. The depth of the imaging system was set to 16 cm.

post hoc analysis was used to identify the source of variation when the ANOVA test was statistically relevant.

When the assumption of normal distribution with equal variances could not be met we used Kruskal-Wallis one way analysis of variance. The significance threshold was set to $P = 0.01$. In addition to the expert quality we investigated the impact of the ROI position relative to the geometry of the image. We computed the Pearson correlation coefficient between the ROI position and the detection performance for each expert and iteration.

Textural algorithms were implemented in a custom-made software system developed at Technical University of Cluj Napoca, Romania. Classification schema used the LibSVM implementation [30] (public domain, ver. 2.89) integrated in weka framework [23] (public domain, ver. 3.7). Statistical analysis was performed in R (public domain, ver. 2.10).

3. Results

The texture analysis system was validated using three sets of images. First dataset contained regions from D77, D84, D55, D53, and D24 Brodatz [29] textures. Second dataset consisted of D4 and D84 textures. The third set had regions from D5 and D92. The classification accuracy was 98.9 for the first set, 98.4 for the second set, and 97.9 for the third set.

Clinical and biochemical characteristics of the study patients are summarized in Table 1. The median length of the LB samples was 11.38 mm, and the mean number of portal spaces was 11.6. The fibrosis stage distribution in our patients was as follows: F0—51 (40.8%) and F4—74 (59.2%).

Each expert was instructed to select one region of interest for each patient. The process was iterated five times. Expert 1 established in average 121.6 regions (min = 121, max = 122), expert 2—120.8 (115–123), expert 3—122 (122–122), and

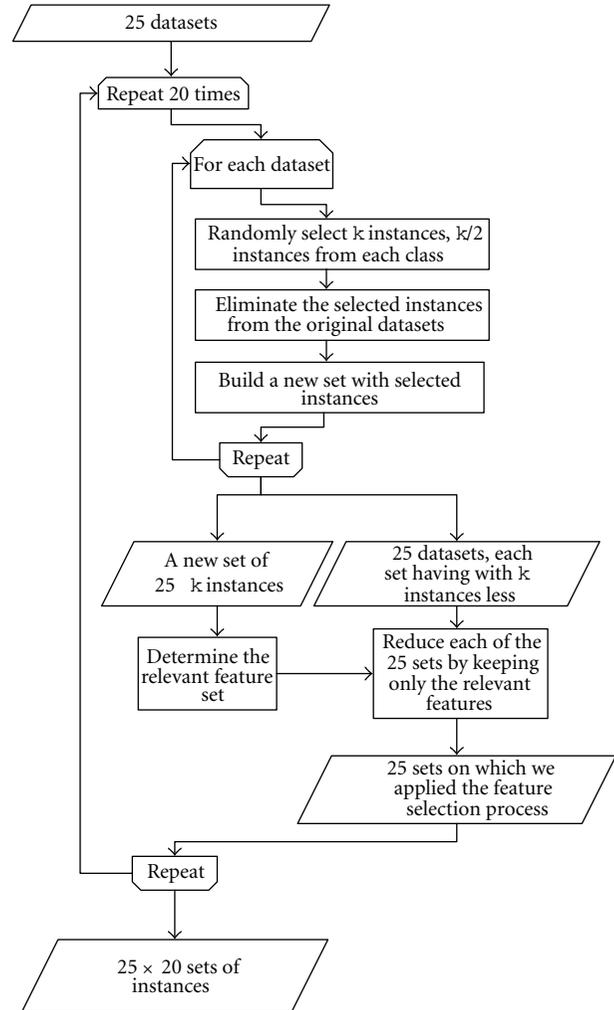


FIGURE 4: Relevant feature selection. To ensure that the selection process is not data dependent, a small number of instances were extracted from each dataset.

expert 4—113 (112–115). The automatic ROI establishment algorithm (expert 5) established 83 images (83–83). There were three patients that had poor quality images and no physician was able to establish an ROI. Two were healthy patients and one was cirrhotic.

We recorded the mean and standard deviation AUROC for each of the experts: expert 1— 0.618 ± 0.059 , expert 2— 0.611 ± 0.085 , expert 3— 0.537 ± 0.062 , expert 4— 0.528 ± 0.075 , and expert 5— 0.611 ± 0.074 .

We investigated the role of feature selection and the user expertise in the performance of the system using two way ANOVA. The only relevant factor was the human expert ($P < 0.0001$) as shown in Figure 5. The other factor, feature selection, was not relevant ($P = 0.8$). In Figure 6 are shown the corresponding box plots.

Post hoc analysis using Tukey method revealed that the differences between experts are significant ($P < 0.001$) with several exceptions, the difference between the expert 1 and 2 and the difference between expert 1 and 5. Note that the expert 5 is the automatic ROI establishment algorithm.

TABLE 1: Characteristics of the study group.

| Characteristics of patients | Entire lot | Patients with fibrosis stage 0 Mean \pm SD (interval or %) | Patients with cirrhosis |
|-------------------------------------|--------------------------------|---|-------------------------------|
| Number | 125 (100%) | 51 (40.8%) | 74 (59.2%) |
| Sex (male) | 50 (40%) | 16 (31.4%) | 34 (45.9%) |
| Age (years) | 47.45 \pm 12.13 (22–77) | 53.39 \pm 8.93 (33–77) | 38.82 \pm 10.97 (22–66) |
| BMI (kg/m ²) | 26.41 \pm 5.15 (18.56–46.48) | 28.29 \pm 5.33 (18.83–46.48) | 23.9 \pm 3.65 (18.56–33.87) |
| AST (U/I) | 58.54 \pm 47.67 (12–387) | 82 \pm 49.57 (23–387) | 25.79 \pm 13.47 (12–71) |
| ALT (U/I) | 75.68 \pm 55.66 (8–270) | 102.25 \pm 53.94 (21–270) | 38.58 \pm 31.87 (8–163) |
| GGT (U/I) | 77.83 \pm 107.77 (13–993) | 105.47 \pm 133.33 (27–993) | 39.83 \pm 28.13 (13–130) |
| Total bilirubin (mg/dL) | 0.88 \pm 0.64 (0.27–4.27) | 1.09 \pm 0.73 (0.4–4.27) | 0.59 \pm 0.28 (0.27–1.72) |
| Alkaline phosphatase (U/I) | 263.13 \pm 188.34 (127–1781) | 286.98 \pm 215.81 (127–1781) | 201.5 \pm 45.61 (142–307) |
| Glucose (mg/dL) | 106.73 \pm 27.75 (72–266) | 113.81 \pm 32.78 (72–266) | 96.86 \pm 13.72 (72–129) |
| Cholesterol (mg/dL) | 195.29 \pm 45.8 (97–331) | 174.22 \pm 36.31 (97–299) | 223.83 \pm 41.92 (149–331) |
| Triglycerides (mg/dL) | 124.11 \pm 57.67 (51–349) | 123.85 \pm 50.08 (53–316) | 124.46 \pm 67.16 (51–349) |
| Platelet count (10 ⁹ /L) | 166.06 \pm 70.32 (42–373) | 142.81 \pm 65.35 (42–373) | 226.52 \pm 40.94 (151–314) |
| INR | 1.12 \pm 0.2 (0.83–1.84) | 1.17 \pm 0.2 (0.89–1.84) | 0.99 \pm 0.12 (0.83–1.3) |
| Right lobe images per patient | 12.97 \pm 6.06 (2–33) | 13.02 \pm 5.03 (3–24) | 12.94 \pm 6.69 (2–33) |

Abbreviations: body mass index (BMI), aspartate aminotransferase (AST), alanine aminotransferase (ALT), gamma-glutamyl-transpeptidase (GGT), and international normalized ratio (INR).

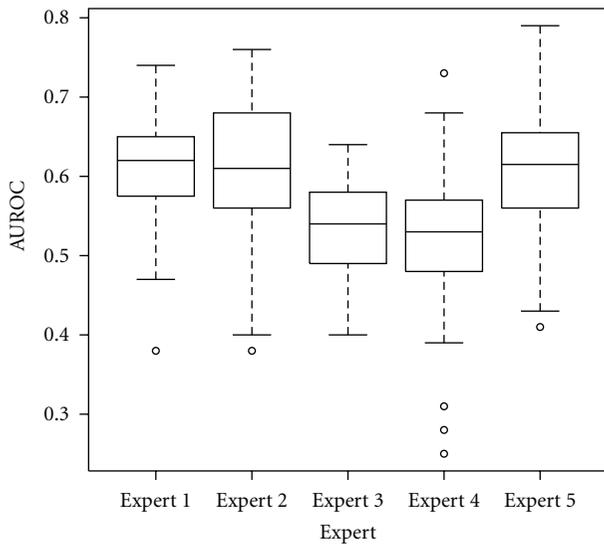


FIGURE 5: Box plot representing the dependency between the estimated performance and the expert that established the regions of interest. The top and the bottom of the boxes are the first and third quartiles, respectively. Thus, the length of the box represents the interquartile range within which 50% of the values were located. The line through the middle of each box represents the median.

In practice, a classifier is trained with data gathered from an expert but it can be used by other physicians. We identified two cases. First case, the expert that trained the classifier uses it in the current practice. In this scenario, the same expert that first established the ROIs establishes the ROIs for the new, unknown images. In the second scenario the expert that establishes the ROIs on the new images is different from the initial expert.

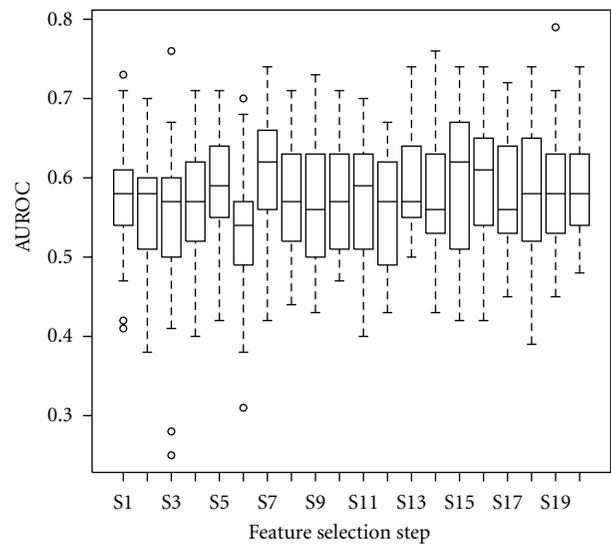


FIGURE 6: Box plot representing the dependency between the estimated performance and the feature selection process. Each label on the horizontal axis represents a separate feature selection step. The top and the bottom of the boxes are the first and third quartiles, respectively. Thus, the length of the box thus represents the interquartile range within which 50% of the values were located. The line through the middle of each box represents the median.

The first scenario was simulated here by training a classifier with each dataset from each expert. Resulting classifier was evaluated using the other datasets from the same expert obtained at different ROI establishment step. Kruskal-Wallis test revealed that there is a significant variation due to the human expert ($P < 0.001$), as seen in Figure 7.

Again, most experienced experts provided best performance. During this test we ignored the results from expert 5.

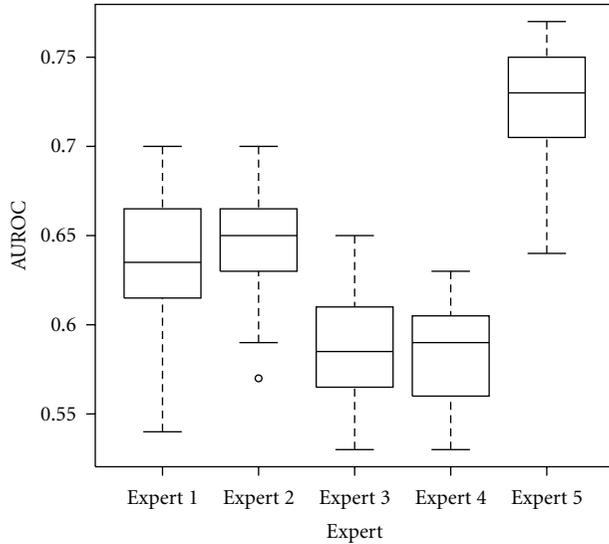


FIGURE 7: Box plot representing the estimated performance obtained when the same expert trains and uses the texture analysis tool in clinical practice. The top and the bottom of the boxes are the first and third quartiles, respectively. Thus, the length of the box thus represents the interquartile range within which 50% of the values were located. The line through the middle of each box represents the median.

Because this expert establishes the regions in the same position, deciding only to accept or reject an image, we noted that there is a significant subset of images that are always selected by this algorithm in all 5 iterations. This subset positively biases the performance evaluation in the case of expert 5, because one will find identical samples in the train and test set. The same analysis applied on the human experts revealed that few images were common between the ROI establishment iterations.

In the second scenario, the expert who uses the non-invasive tool is different from the expert that provided the training data for the system. We trained the classifier with the data collected from one expert and then test it with the data collected from the other experts. Kruskal-Wallis analysis revealed an interesting fact; there is no significant variance due to experts ($P = 0.0506$) as shown in Figure 8. In both scenarios the analysis did not revealed significant variance due to the feature selection step.

In the following we investigated the impact of the ROI position relative to the image geometry. The center coordinates of the ROIs were converted to polar space. The center of the polar space was set to be the virtual source of ultrasounds. For each ROI the angle θ and the vector length, ρ , were computed. For each expert and iteration we computed the mean angle and length. A linear regression was performed between these coordinates and the mean performance of the expert i during iteration j . We computed the Pearson correlation coefficient and its relevance. In Figures 9 and 10 are the shown the results.

The correlation coefficients were -0.44 (between ρ and AUROC) and -0.48 (between θ and AUROC). This correlation is not statistically significant for the chosen

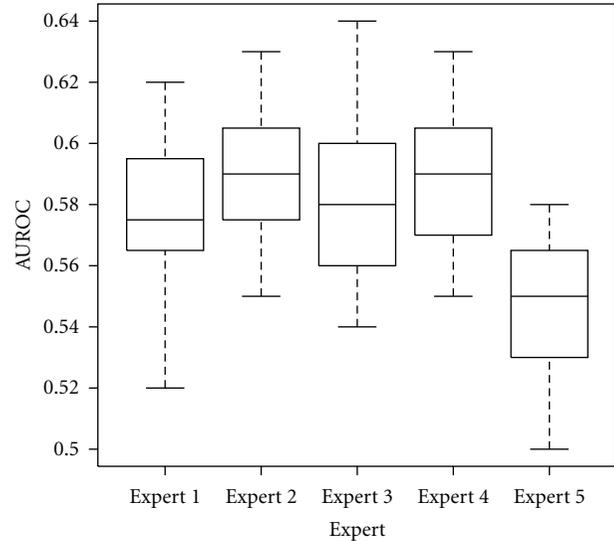


FIGURE 8: Box plot representing the estimated performance obtained when the texture analysis system is trained with datasets provided by one expert and used with ROIs established by a different expert. The top and the bottom of the boxes are the first and third quartiles, respectively. Thus, the length of the box thus represents the interquartile range within which 50% of the values were located. The line through the middle of each box represents the median.

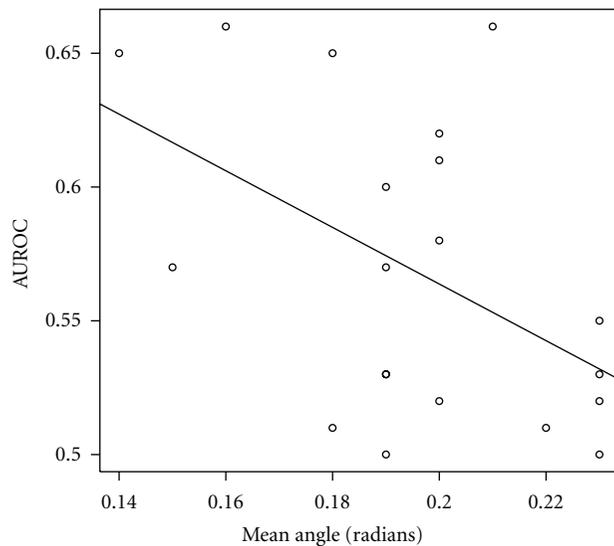


FIGURE 9: The dependence of the estimated performance in relation to the θ coordinate of the ROI's center.

threshold. However, it is possible that a link exists between the ROI position and the classification performance because the results became relevant for a higher threshold ($P < 0.05$).

We also compared the mean positions of the ROIs when expressed in simple Cartesian coordinates. For each ROI the center coordinates were computed relative to the top left corner of the image. One-way ANOVA showed that the O_x (horizontal) coordinate is not relevant but for the O_y (vertical) coordinate, higher performances were obtained for the regions that were established closer to the upper part of

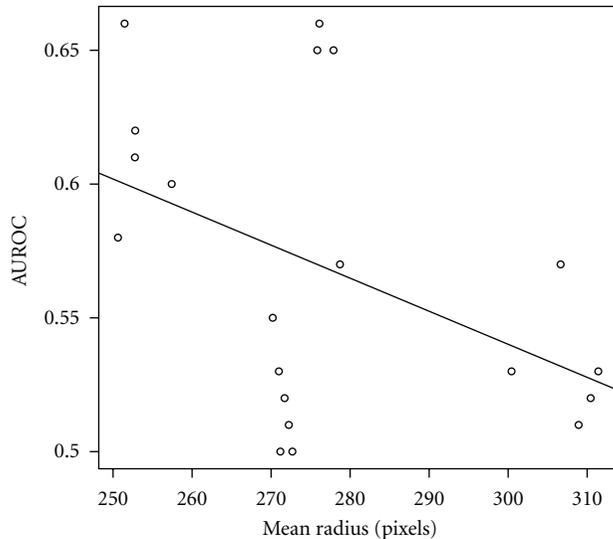


FIGURE 10: The dependence of the estimated performance in relation to the ρ coordinate of the ROI's center.

the image. Figure 11 shows the BOXPLOT graph. Again the expert 5 was ignored because the automatic algorithm always established the regions in the same position.

4. Discussions

Liver biopsy is an imperfect golden standard in fibrosis staging. It is an invasive procedure and even if the method allows direct examination of the liver tissue there is a certain variability due to the reduced tissue volume and due to the fact that a human expert qualitatively evaluates the biopsy [31–33].

There are numerous research directions involving non-invasive fibrosis staging and noninvasive diagnosis of liver diseases in general [34, 35]. Papers [8, 21, 22, 36, 37] studying texture analysis as a noninvasive staging tool reported high performances in cirrhosis detection [36] and even in fibrosis staging [8]. In these papers there are variations in terms of studied pathology and classification evaluation methodology. We believe that these factors might have positively biased the results reported by other authors.

Present study aims to evaluate the dependence between the human expert and the performance of such a texture analysis system in predicting the cirrhosis in chronic hepatitis C patients. In the same time the present paper brings the following contributions to the noninvasive fibrosis detection field: it includes only patients with chronic hepatitis C, excluding other pathologies; it integrates almost all textural algorithms met in fibrosis detection and it proposes a more rigorous performance evaluation methodology that gives results closer to the real performance of a classifier.

In present study we included only patients with chronic hepatitis C etiology. Other papers that address the noninvasive detection of cirrhosis include patients having different pathologies like fatty infiltration [16]. Another important highlight of this paper is the volume of patients. There are few papers that study more than 100 patients but not all the

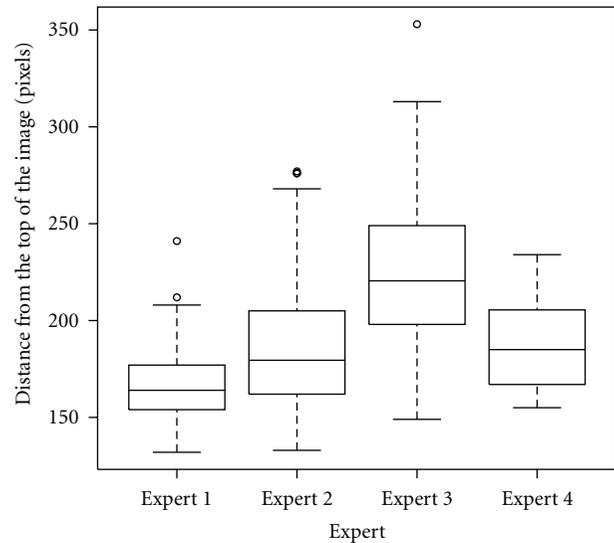


FIGURE 11: Box plot representing the O_y coordinate distribution. The top and the bottom of the boxes are the first and third quartiles, respectively. Thus, the length of the box thus represents the interquartile range within which 50% of the values were located. The line through the middle of each box represents the median.

patients included in these studies have chronic hepatitis C, or the etiology for cirrhosis is not specified [4, 16, 38].

Performance estimation algorithm proposed in this paper ensures that each time the classifier is tested the test data are new and unseen at the training or feature selection phase. The metaparameter sets are evaluated on unseen data to ensure that we do not select a classifier instance that overfits the training data. The cross-validation loop ensures that even this search procedure does not overfit the data. The 10-time repetition of the evaluation phase ensures a better estimation of the mean performance. No other papers employed repeated performance estimation on their classification schemas. When performing one iteration the data might get partitioned in such a way that by accident the performance estimation is very high. For example, in some iterations the performance reached levels as high as 0.79. Of course, the mean performance estimated over 10 runs is smaller. The same phenomenon of increased variance can be noted when the performance measure is computed on each test fold and not on the entire prediction vector. In 2-fold CV a “lucky” splitting might give a very high performance reading.

In present paper, the CV predictions are collected and the performance is measured on a vector that has the same dimensions as the initial dataset.

Textural feature selection is performed on an independent dataset. This dataset is obtained by randomly sampling the original datasets. It is important to note that each instance that is included in the feature selection dataset is excluded from the original dataset. As a result, the feature selection process has less chances of overfitting.

The particular set of features does not influence the detection rates. The subset of features selected at each step has a high variability. High ranking features cover large

spectra of algorithms, from statistical algorithms to multiresolution analysis. This indicates that the specific algorithm used to numerically describe the texture has its importance but there are fewer chances that new textural algorithms will make a great impact over cirrhosis detection and fibrosis staging.

The design of the experiment, where each expert establishes 5 sets of ROIs ensures that the samples are independent and normally distributed. Each set of patients have different randomizations in order to minimize the effect of patient/image succession over the experiment. Moreover, for each patient the order of the images is altered. It is important to note that the order of images is the same for all the experts. Expert x viewed the patients and images in the same order as the expert y when establishing ROI for the same dataset z .

The main finding of this paper is that the performance of the studied software diagnosis tool depends on the expert that employs this tool. In the results section we have shown that there is a significant performance variation between experts. The results presented here showed that more experienced experts tend to capture the same aspects of the ultrasound image, aspects that are consistent with the histological findings. If this tool is trained and employed by an experienced physician it might give some extra information about the underlying pathology.

The results from the second scenario, when the expert that uses the texture analysis tool is different from the expert that provided the data for training, revealed the fact that there is little use for texture analysis tool in screening processes.

The classical methodology has a severe drawback. It requires a human expert to establish a representative area where the texture will be analyzed. Replacing the human expert with a computerized solution improves the usefulness of such a software analysis tool. The results shows that such a tool can have a performance similar to a highly trained expert. This result is another important contribution of this paper to the noninvasive diagnosis field.

5. Conclusions

Texture analysis can enhance the diagnosis power of the B-mode ultrasound image. The performance of this approach depends highly on the human expert that establishes the regions of interest. In classical form met in the literature noninvasive diagnosis through texture analysis has limited utility in clinical practice. Further work in this domain has to be focused in finding another noninvasive descriptors for fibrosis.

Acknowledgments

The authors would like to thank Vlad Popovici from Swiss Institute of Bioinformatics for the priceless help in designing the experiments presented here. All authors contributed equal parts to this paper.

References

- [1] C. M. Wu, Y. C. Chen, and K. S. Hsieh, "Texture features for classification of ultrasonic liver images," *IEEE Transactions on Medical Imaging*, vol. 11, no. 2, pp. 141–152, 1992.
- [2] K. J. W. Taylor, C. A. Riely, and L. Hammers, "Quantitative US attenuation in normal liver and in patients with diffuse liver disease: importance of fat," *Radiology*, vol. 160, no. 1, pp. 65–71, 1986.
- [3] B. J. Oosterveld, J. M. Thijssen, P. C. Hartman, R. L. Romijn, and G. J. E. Rosenbusch, "Ultrasound attenuation and texture analysis of diffuse liver disease: methods and preliminary results," *Physics in Medicine and Biology*, vol. 36, no. 8, pp. 1039–1064, 1991.
- [4] Y. M. Kadah, A. A. Farag, J. M. Zurada, A. M. Badawi, and A. B. M. Youssef, "Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 466–478, 1996.
- [5] D. Gaitini, Y. Baruch, E. Ghersin et al., "Feasibility study of ultrasonic fatty liver biopsy: texture vs. attenuation and backscatter," *Ultrasound in Medicine and Biology*, vol. 30, no. 10, pp. 1321–1327, 2004.
- [6] G. T. Cao, P. F. Shi, and B. Hu, "Liver fibrosis identification based on ultrasound images captured under varied imaging protocols," *Journal of Zhejiang University*, vol. 6, no. 11, pp. 1107–1114, 2005.
- [7] D. Gaitini, M. Lederman, Y. Baruch et al., "Computerised analysis of liver texture with correlation to needle biopsy," *Ultraschall in der Medizin*, vol. 26, no. 3, pp. 197–202, 2005.
- [8] H. Yamada, M. Ebara, T. Yamaguchi et al., "A pilot approach for quantitative assessment of liver fibrosis using ultrasound: preliminary results in 79 cases," *Journal of Hepatology*, vol. 44, no. 1, pp. 68–75, 2006.
- [9] J. W. Jeong, S. Lee, J. W. Lee, D. S. Yoo, and S. Kim, "The echotextural characteristics for the diagnosis of the liver cirrhosis using the sonographic images," in *Proceedings of the 29th Annual International Conference of Engineering in Medicine and Biology Society (EMBS '07)*, pp. 1343–1345, August 2007.
- [10] C. Vicas, S. Nedevschi, M. Lupsor, and R. Badea, "Automatic detection of liver capsule using Gabor filters. Applications in steatosis quantification," in *Proceedings of the IEEE 5th International Conference on Intelligent Computer Communication and Processing (ICCP '09)*, pp. 133–140, August 2009.
- [11] P. Bedossa and T. Poynard, "An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group," *Hepatology*, vol. 24, no. 2, pp. 289–293, 1996.
- [12] P. Bedossa, P. Bioulac-Sage, P. Callard et al., "Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C," *Hepatology*, vol. 20, no. 1 I, pp. 15–20, 1994.
- [13] C. Abe, C. E. Kahn Jr., K. Doi, and S. Katsuragawa, "Computer-aided detection of diffuse liver disease in ultrasound images," *Investigative Radiology*, vol. 27, no. 1, pp. 71–77, 1992.
- [14] C. Vicas, M. Lupsor, R. Badea, and S. Nedevschi, "Detection of anatomical structures on ultrasound liver images using Gabor filters," in *Proceedings of the 17th IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR '10)*, pp. 249–253, Romania, Cluj Napoca, May 2010.
- [15] A. Materka and M. Strzelecki, *Texture Analysis Methods: A Review*, Technical University of Lodz, Institute of Electronics, 1998.

- [16] A. M. Badawi, A. S. Derbala, and A. B. M. Youssef, "Fuzzy logic algorithm for quantitative tissue characterization of diffuse liver diseases from ultrasound images," *International Journal of Medical Informatics*, vol. 55, no. 2, pp. 135–147, 1999.
- [17] W. C. Yeh, S. W. Huang, and P. C. Li, "Liver fibrosis grade classification with B-mode ultrasound," *Ultrasound in Medicine and Biology*, vol. 29, no. 9, pp. 1229–1235, 2003.
- [18] W. L. Lee, K. S. Hsieh, Y. C. Chen, and Y. C. Chen, "A study of ultrasonic liver images classification with artificial neural networks based on fractal geometry and multiresolution analysis," *Biomedical Engineering*, vol. 16, no. 2, pp. 59–67, 2004.
- [19] Y. Xia, D. G. Feng, and R. Zhao, "Morphology-based multifractal estimation for texture segmentation," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 614–623, 2006.
- [20] A. Ahmadian, A. Mostafa, M. D. Abolhassani, and Y. Salimpour, "A texture classification method for diffused liver diseases using Gabor wavelets," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS '05)*, pp. 1567–1570, September 2005.
- [21] G. Cao, P. Shi, and B. Hu, "Ultrasonic liver characterization using phase congruency," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS '05)*, pp. 6356–6359, September 2005.
- [22] M. H. Horng, Y. N. Sun, and X. Z. Lin, "Texture feature coding method for classification of liver sonography," *Computerized Medical Imaging and Graphics*, vol. 26, no. 1, pp. 33–42, 2002.
- [23] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2005.
- [24] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer, 2001.
- [25] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Machine Learning*, vol. 59, no. 1-2, pp. 161–205, 2005.
- [26] M. Hall, *Correlation-Based Feature Selection of Discrete and Numeric Class Machine Learning*, University of Waikato, 1998.
- [27] M. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, Oxford, UK, 2003.
- [28] V. Popovici, "Assessment of classification models for medical applications," in *Proceedings of the Workshop on Computers in Medical Diagnoses*, Cluj-Napoca, Romania, 2009.
- [29] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover, New York, NY, USA, 1966.
- [30] LIBSVM: a library for support vector machines, 2001, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [31] A. Regev, M. Berho, L. J. Jeffers et al., "Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection," *American Journal of Gastroenterology*, vol. 97, no. 10, pp. 2614–2618, 2002.
- [32] P. Bedossa, D. Dargere, and V. Paradis, "Sampling variability of liver fibrosis in chronic hepatitis C," *Hepatology*, vol. 38, no. 6, pp. 1449–1457, 2003.
- [33] M. Guido and M. Ruge, "Liver biopsy sampling in chronic viral hepatitis," *Seminars in Liver Disease*, vol. 24, no. 1, pp. 89–97, 2004.
- [34] I. N. Guha and W. M. Rosenberg, "Noninvasive assessment of liver fibrosis: serum markers, imaging, and other modalities," *Clinics in Liver Disease*, vol. 12, no. 4, pp. 883–900, 2008.
- [35] S. Bonekamp, I. Kamel, S. Solga, and J. Clark, "Can imaging modalities diagnose and stage hepatic fibrosis and cirrhosis accurately?" *Journal of Hepatology*, vol. 50, no. 1, pp. 17–35, 2009.
- [36] G. T. Cao, P. F. Shi, and B. Hu, "Liver fibrosis identification based on ultrasound images," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS '05)*, vol. 1–7, pp. 6317–6320, September 2005.
- [37] A. Mojsilovic, M. Popovic, S. Markovic, and M. Krstic, "Characterization of visually similar diffuse diseases from b-scan liver images using nonseparable wavelet transform," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 541–549, 1998.
- [38] A. Szebeni, G. Tolvaj, and A. Zalatnai, "Correlation of ultrasound attenuation and histopathological parameters of the liver in chronic diffuse liver diseases," *European Journal of Gastroenterology and Hepatology*, vol. 18, no. 1, pp. 37–42, 2006.

Research Article

Nested Quantization Index Modulation for Reversible Watermarking and Its Application to Healthcare Information Management Systems

Lu-Ting Ko,¹ Jwu-E. Chen,¹ Yaw-Shih Shieh,² Hsi-Chin Hsin,³ and Tze-Yun Sung²

¹ Department of Electrical Engineering, National Central University, Chungli 320-01, Taiwan

² Department of Electronics Engineering, Chung Hua University, Hsinchu 300-12, Taiwan

³ 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 25 August 2011; Accepted 15 September 2011

Academic Editor: Shengyong Chen

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

Digital watermarking has attracted lots of researches to healthcare information management systems for access control, patients' data protection, and information retrieval. The well-known quantization index modulation-(QIM-) based watermarking has its limitations as the host image will be destroyed; however, the recovery of medical images is essential to avoid misdiagnosis. In this paper, we propose the nested QIM-based watermarking, which is preferable to the QIM-based watermarking for the medical image applications. As the host image can be exactly reconstructed by the nested QIM-based watermarking. The capacity of the embedded watermark can be increased by taking advantage of the proposed nest structure. The algorithm and mathematical model of the nested QIM-based watermarking including forward and inverse model is presented. Due to algorithms and architectures of forward and inverse nested QIM, the concurrent programs and special processors for the nested QIM-based watermarking are easily implemented.

1. Introduction

Digital watermarking is a scheme of embedding data in an image called host image for the purpose of copyright protection, integrity check, and/or access control [1–6]. Some of the requirements of digital watermarking are transparency, robustness, and capacity. Specifically, transparency means that the watermark embedded in the host image is imperceptible to human eyes, robustness means the resistance of the watermark to malicious attacks, and capacity denotes the amount of data that can be hidden in the host image. Digital watermarking has been applied to many applications [7–10].

For medical images, such as radiography, magnetic resonance imaging (MRI), nuclear medicine imaging, photo acoustic imaging, tomography, and ultrasound, the conventional watermarking schemes are not suitable due to the distortion problem, which can lead to misdiagnosis [11–14]. In order to provide the requirements of confidential data

protection and intact information retrieval [15–24], watermarking with legal and ethical functionalities is desirable especially for the medical images applications. For diagnostic information retrieving and host image reconstruction, the reversible watermarking can be achieved by using many common used techniques, which are based on the characteristics of nonlinear time series [25–30]. More specifically, confidential data such as patients' diagnosis reports can be taken as watermark data and then embedded in the host image by using digital watermarking with authorized utilization. Thus, digital watermarking can be used to facilitate healthcare information management systems.

In this paper, we propose a novel scheme called the nested quantization index modulation-(QIM-) based watermarking for the healthcare information management applications. The remainder of the paper proceeds as follows. In Section 2, the conventional QIM-based watermarking is reviewed briefly. Section 3 describes the nested QIM-based watermarking.

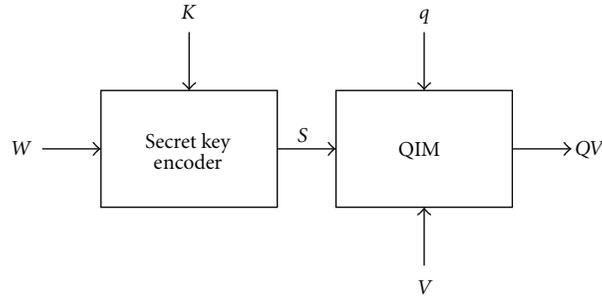


FIGURE 1: The conventional QIM-based watermarking (W : the watermark, K : the secret key, S : the coded watermark, q : the quantization step, V : the host image, QV : the watermarked image).

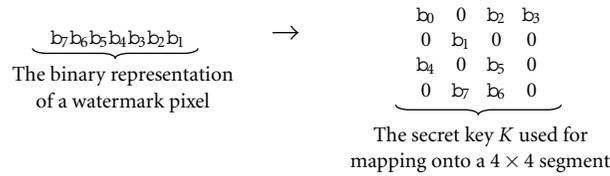


FIGURE 2: The secret key K used for mapping the watermark onto the host image.

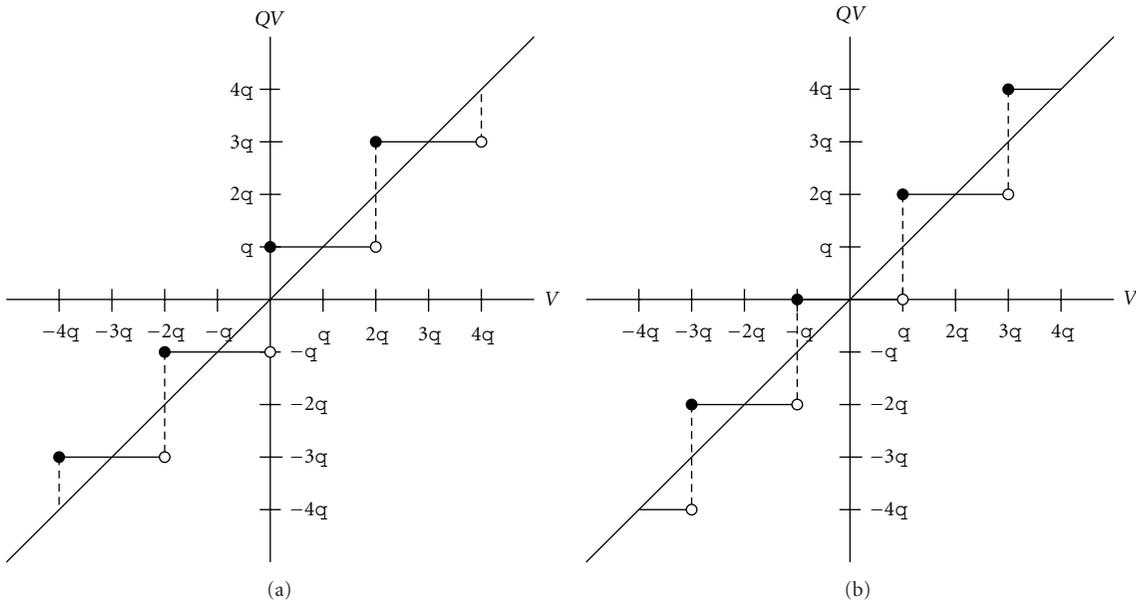


FIGURE 3: Operations of the QIM scheme for the coded watermark pixels being (a) bit 1 and (b) bit 0, respectively.

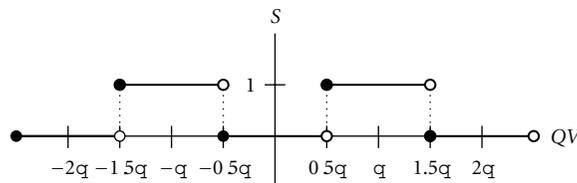


FIGURE 4: Operations of the inverse QIM scheme for the coded watermark pixels.

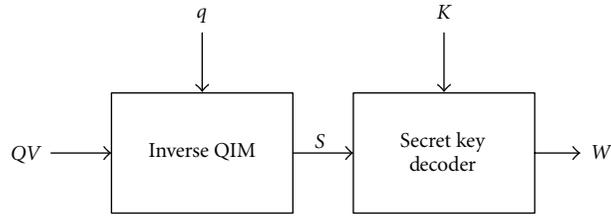


FIGURE 5: Extraction of the watermark, W , from the watermarked image, QV , based on the conventional QIM scheme.

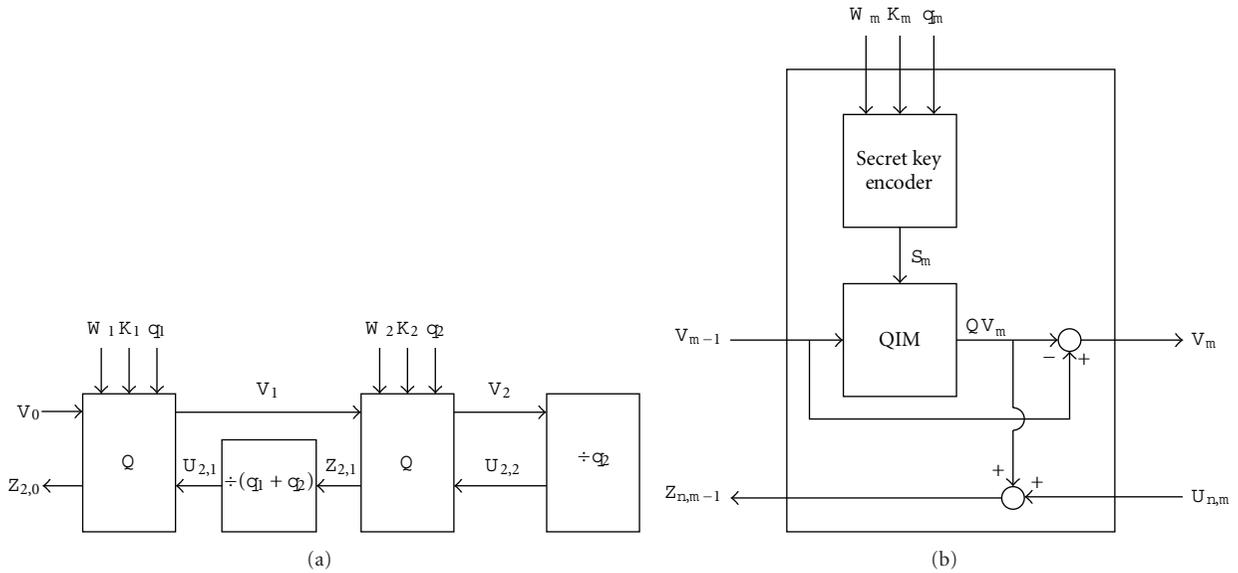


FIGURE 6: (a) The simplest nested QIM with two Q blocks (V_0 : the host image, $Z_{2,1}$: the intermediate watermarked image, $Z_{2,0}$: the final watermarked image, W_1 and W_2 : the watermarks, K_1 and K_2 : the secret keys, q_1 and q_2 : the quantization steps, V_1 and V_2 : the quantization errors, $U_{2,1}$ and $U_{2,2}$: the normalized quantization errors). (b) The Q block used in the nested QIM.

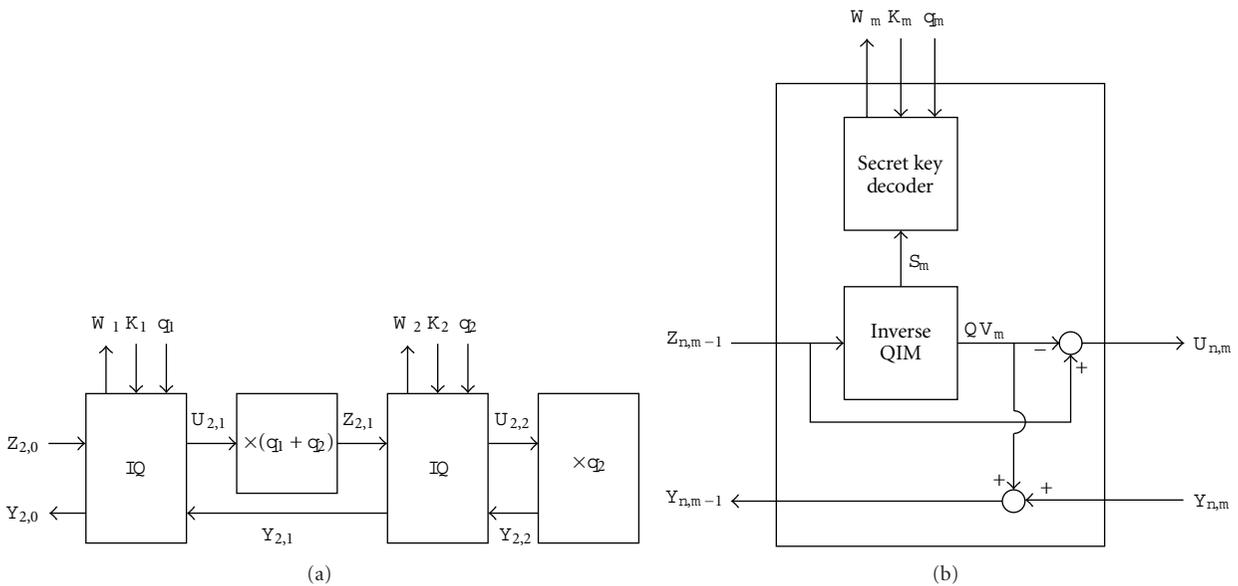
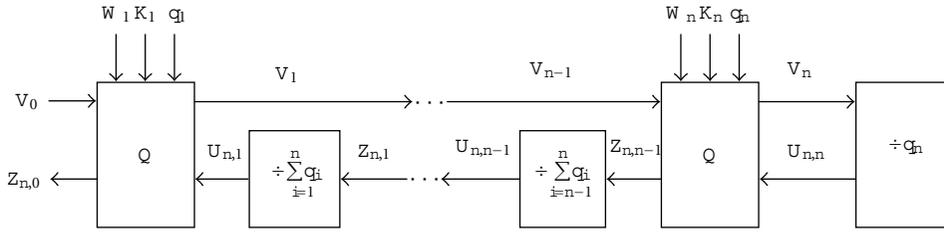
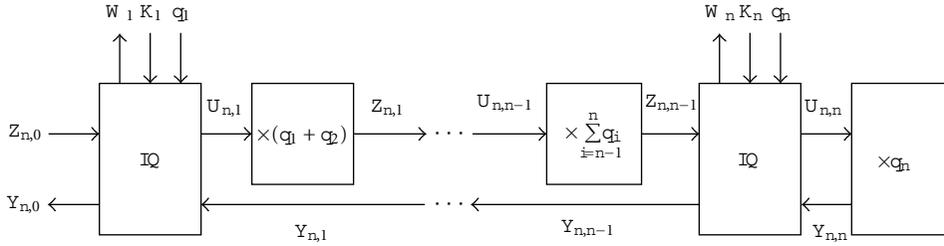


FIGURE 7: (a) The inverse nested QIM with two inverse Q (IQ) blocks. (b) The IQ block used in the inverse nested QIM for watermark extraction.

FIGURE 8: The nested QIM with n QIM operations for watermarking.FIGURE 9: Extraction of the original host and watermark images from the watermarked image based on the nested QIM with n inverse QIM operations.

Experimental results on medical images are presented in Section 4. Conclusion is given in Section 5.

2. Review of Quantization Index Modulation

Figure 1 depicts the conventional quantization index modulation-(QIM-) based watermarking scheme [31], where W , K , S , V , and QV denote the watermark, the secret key, the coded watermark, the host image, and the watermarked image, respectively. For the sake of simplicity, let us consider monochromatic images with 256 grey levels, and the size of the watermark is one-fourth of that of the host image. The secret key is used to map the binary representation of the watermark onto the host image, for example, Figure 2 depicts the binary representation of a watermark pixel that is mapped onto a 4×4 segment using a given secret key.

Figure 3 shows the operation of the QIM block, in which the grey levels of the host image, V , ranging between $2c \cdot q$ and $(2c + 1) \cdot q$ will be quantized into $(2c + 1) \cdot q$ if the corresponding pixels of the coded watermark, S , are bit 1; otherwise they are quantized into $2c \cdot q$ if the corresponding pixels are bit 0. For the grey levels of V that are between $(2c + 1) \cdot q$ and $(2c + 2) \cdot q$, they will be quantized into $(2c + 1) \cdot q$ or $(2c + 2) \cdot q$ depending on the corresponding pixels of S being bit 1 or 0, respectively. Note that q denotes the quantization step, $0 \leq c < (255/2 \cdot q)$, and c is an integer number.

It is noted that the watermarked image, QV , can be written as

$$QV(i, j) = \begin{cases} (2c+1)q, & \text{if } V(i, j) \\ & \in ((2c+0.5)q, (2c+1.5)q], \\ & S(i, j) = 1, \\ (2c)q, & \text{if } V(i, j) \\ & \in ((2c-0.5)q, (2c+0.5)q], \\ & S(i, j) = 0, \end{cases} \quad (1)$$

where (i, j) denotes the position index of pixels, and the coded watermark, S , can be obtained by

$$S(i, j) = \begin{cases} 1, & \text{if } QV(i, j) \in ((2d+0.5)q, (2d+1.5)q], \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

as shown in Figure 4. Together with the secret key, K , the watermark, W , can be exactly extracted from the watermarked image, QV , as shown in Figure 5.

3. The Proposed Nested QIM for Reversible Watermarking

One of the fundamental requirements for the medical applications is recovery of the host image. As the conventional

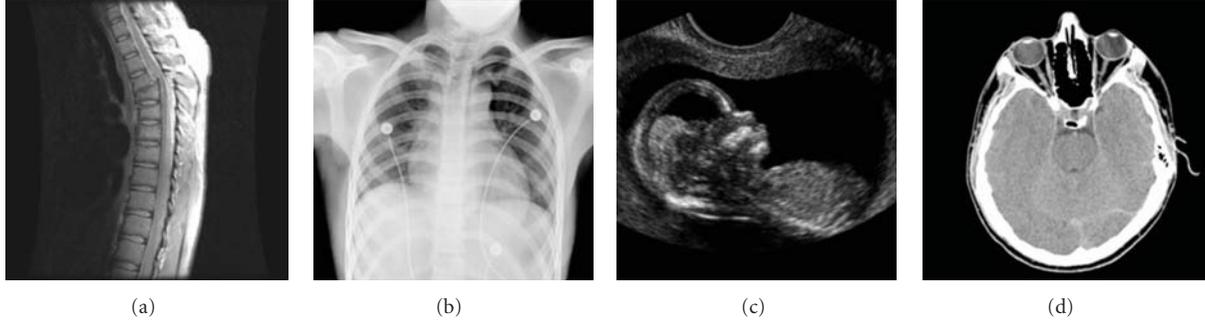


FIGURE 10: The 256×256 host images with 256 grey levels; (a) spine (MRI), (b) chest (X-ray), (c) fetus (ultrasonic), and (d) head (CT).

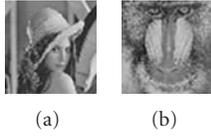


FIGURE 11: The 64×64 (a) Lena image and (b) Baboon image with 256 grey levels used as watermarks.

QIM-based watermarking is irreversible and the host image can not be exactly reconstructed, we propose a novel algorithm called the nested QIM algorithm for reversible watermarking. Figure 6(a) depicts the simplest nested QIM consisting of only two QIM operations, each of which is performed in the Q block shown in Figure 6(b), where W_m , K_m , S_m , V_{m-1} , $Z_{n,m-1}$, QV_m , V_m , and q_m are the watermark, the secret key, the coded watermark, the input host image, the output watermarked image, the quantization, the quantization error, and the quantization step at the m th stage of a n -level nested QIM, respectively. The original host image and the final watermarked image denoted by V_0 and $Z_{2,0}$ are taken as the input and the output of the first stage.

As one can see, QIM is a nonlinear function, and we have

$$\begin{aligned} QV_1 &= \text{QIM}(V_0), \\ V_1 &= V_0 - QV_1, \\ V_2 &= V_1 - QV_2 = V_1 - \text{QIM}(V_1), \\ Z_{2,1} &= QV_2 + U_{2,2} = QV_2 + \frac{V_2}{q_2}. \end{aligned} \quad (3)$$

The final watermarked image, $Z_{2,0}$ is thus derived as

$$\begin{aligned} Z_{2,0} &= QV_1 + U_{2,1} \\ &= QV_1 + \frac{Z_{2,1}}{q_1 + q_2} \\ &= QV_1 + \frac{QV_2 + U_{2,2}}{q_1 + q_2} \\ &= QV_1 + \frac{QV_2 + V_2/q_2}{q_1 + q_2}. \end{aligned} \quad (4)$$

The original host image and the embedded watermark images can be exactly reconstructed from the watermarked

image, $Z_{2,0}$, by using the inverse nested QIM shown in Figure 7(a), where the inverse Q (IQ) block is shown in Figure 7(b).

Based on the data flow of Figures 7(a) and 7(b), we have

$$\begin{aligned} QV_1 &= \text{IQIM}(Z_{2,0}), \\ Z_{2,1} &= (Z_{2,0} - QV_1) \times (q_1 + q_2), \\ QV_2 &= \text{IQIM}(Z_{2,1}), \\ Y_{2,2} &= (Z_{2,1} - QV_2) \times q_2, \\ Y_{2,1} &= QV_2 + Y_{2,2}. \end{aligned} \quad (5)$$

Together with (3) and (4), the original host image, V_0 , is thus obtained by

$$\begin{aligned} Y_{2,0} &= QV_1 + Y_{2,1} \\ &= QV_1 + QV_2 + Y_{2,2} \\ &= V_0. \end{aligned} \quad (6)$$

The above-mentioned equations can be generalized as follows:

$$\begin{aligned} QV_m &= \text{QIM}(V_{m-1}), \\ V_m &= V_{m-1} - QV_m, \\ Z_{n,m-1} &= QV_m + U_{n,m}, \\ Z_{n,0} &= QV_1 + \frac{V_n - QV_n}{\prod_{h=1}^n (\sum_{k=h}^n q_k)} + \sum_{j=2}^n \frac{QV_j}{\prod_{h=1}^{j-1} (\sum_{k=h}^n q_k)}, \end{aligned} \quad (7)$$

$$\begin{aligned} QV_m &= \text{IQIM}(Z_{n,m-1}), \\ U_{n,m} &= Z_{n,m-1} - QV_m, \\ Y_{n,m-1} &= Y_{n,m} + QV_m, \\ Y_{n,0} &= QV_1 + Y_{n,1} \\ &= QV_1 + QV_2 + \cdots + QV_m + Y_{n,n} \\ &= \left(\sum_{k=1}^n QV_k \right) + Y_{n,n} \\ &= V_0. \end{aligned} \quad (8)$$

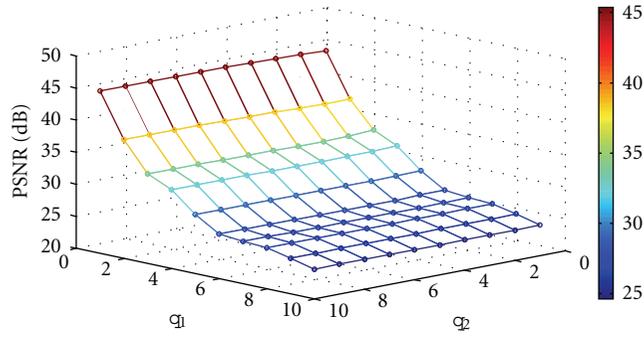


FIGURE 12: The PSNR of the watermarked image of spine (MRI) at various quantization steps q_1 and q_2 .

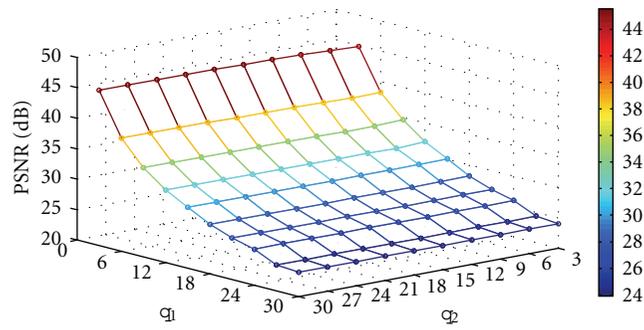


FIGURE 13: The PSNR of the watermarked image of chest (X-ray) at various quantization steps q_1 and q_2 .

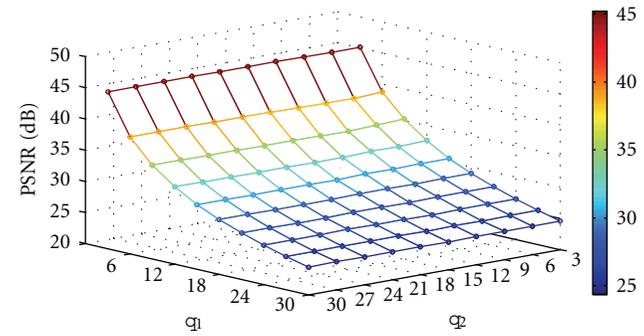


FIGURE 14: The PSNR of the watermarked image of fetus (ultrasonic) at various quantization steps q_1 and q_2 .

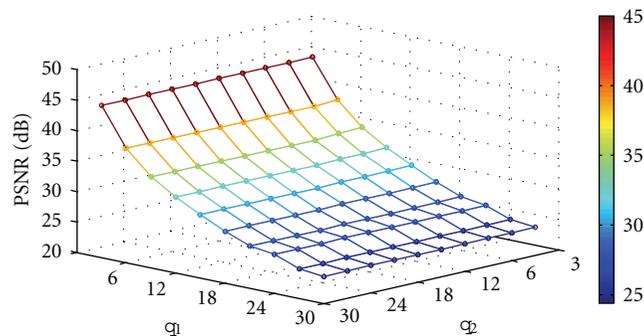


FIGURE 15: The PSNR of the watermarked image of head (CT) at various quantization steps q_1 and q_2 .

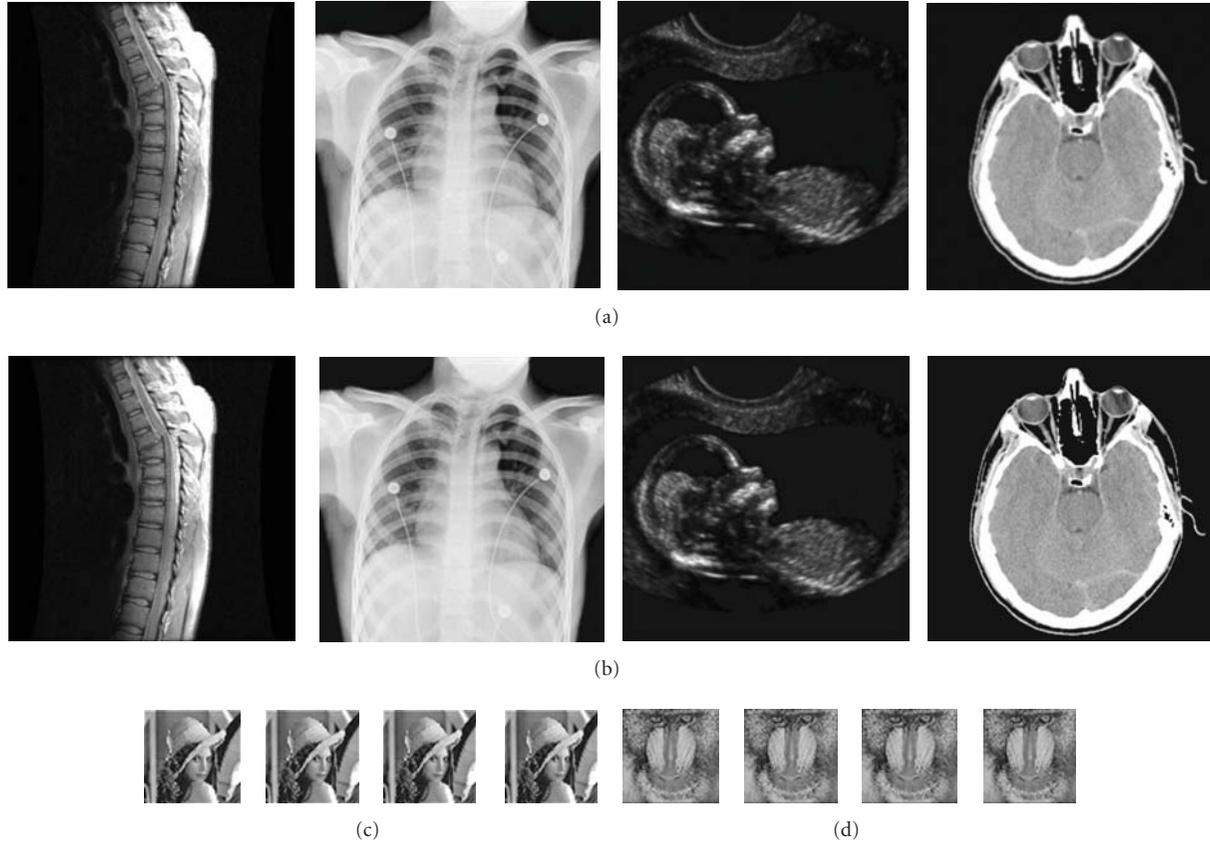


FIGURE 16: The watermarked images (a), the reconstructed images (b), and the extracted watermarks ((c) and (d)) with the quantization steps $q_1 = 30$ and $q_2 = 30$.

where (7) are used for the nested QIM, and (8) for the inverse nested QIM. The corresponding block diagrams for embedding and extracting watermarks with the host image are shown in Figures 8 and 9, respectively.

4. Experimental Results on Medical Images

The nested QIM-based watermarking algorithm has been evaluated on various medical images. Figure 10 shows the test 256×256 images with 256 grey levels, namely, spine, chest, fetus and head obtained by magnetic resonance image (MRI), X-ray, ultrasound, and computed tomography (CT), respectively, which are used as host images. Figure 11 shows the 64×64 Lena image and Baboon image with 256 grey levels, which are used as watermarks.

The peak signal to noise ratio (PSNR) is used to evaluate the image quality [18, 31], which is defined as

$$\text{PSNR} = 20 \log \left(\frac{255}{\sqrt{\text{MSE}}} \right), \quad (9)$$

where MSE denotes the mean square error. Figures 12, 13, 14, and 15 show the PSNR of the watermarked images of spine (MRI), chest (X-ray), fetus (ultrasonic), and head (CT) at various quantization steps, q_1 and q_2 . Figure 16 shows

the watermarked images (first row), the reconstructed images, (second row) and the extracted watermarks (third and fourth rows) with $q_1 = 30$ and $q_2 = 30$. It is noted that the watermarked images even with large quantization steps are almost indistinguishable from the exactly reconstructed host images.

5. Conclusion

In this paper, a novel algorithm called the nested QIM has been proposed for medical image watermarking. The capacity of the embedded watermark can be increased by taking advantage of the proposed nest structure. As the host image can be exactly reconstructed, it is suitable especially for the medical image applications. In addition, the healthcare information such as patients' data, digital signatures, and identification codes can be well embedded in medical images. Thus, the nested QIM-based medical image watermarking is preferable to facilitate data management in healthcare information management systems.

Acknowledgment

The National Science Council of Taiwan, under Grants NSC100-2628-E-239-002-MY2, supported this work.

References

- [1] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.
- [2] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1064–1087, 1998.
- [3] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 783–791, 2001.
- [4] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data," *IEEE Signal Processing Magazine*, vol. 17, no. 5, pp. 20–46, 2000.
- [5] C. I. Podilchuk and E. J. Delp, "Digital watermarking: algorithm and application," *IEEE Signal Processing Magazine*, vol. 18, no. 4, pp. 33–46, 2001.
- [6] C. S. Lu and H. Y. M. Liao, "Multipurpose watermarking for image authentication and protection," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1579–1592, 2001.
- [7] S. Erküçük, S. Krishnan, and M. Zeytinoglu, "A robust audio watermark representation based on linear chirps," *IEEE Transactions on Multimedia*, vol. 8, no. 5, Article ID 1703507, pp. 925–936, 2006.
- [8] F. Chuhong, D. Kundur, and R. H. Kwong, "Analysis and design of secure watermark-based authentication systems," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 43–55, 2006.
- [9] Y. Feng, N. H. Ling, G. Lin, W. Zhe, and W. J. Zhang, "Transmitter identification with watermark signal in DVB-H signal frequency network," *IEEE Transactions on Broadcasting*, vol. 55, no. 3, Article ID 5170023, pp. 663–667, 2009.
- [10] A. Wakatani, "Digital watermarking for ROI medical images by using compressed signature image," in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences, (HICSS '02)*, pp. 2043–2048, Big Island, Hawaii, USA, January 2002.
- [11] L. T. Ko, J. E. Chen, H. C. Hsin, Y. S. Shieh, and T. Y. Sung, "Haar wavelet based just noticeable distortion model for transparent watermark," *Mathematical Problems in Engineering*, vol. 2012, Article ID 635738, 14 pages, 2012.
- [12] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 480–487, 2011.
- [13] G. Coatrieux, H. Maître, B. Sankur, Y. Rolland, and R. Collorec, "Relevance of watermarking in medical imaging," in *Proceedings of the IEEE/EMBS Region 8th International Conference on Information Technology Applications in Biomedicine, (ITAB-ITIS '00)*, pp. 250–255, Arlington, Va, USA, November 2000.
- [14] A. Bruckmann and A. Uhl, "Selective medical image compression using wavelet techniques," *Journal of Computing and Information Technology*, vol. 6, no. 2, pp. 203–213, 1998.
- [15] H. M. Chao, C. M. Hsu, and S. G. Miaou, "A data-hiding technique with authentication, integration, and confidentiality for electronic patient records," *IEEE Transactions on Information Technology in Biomedicine*, vol. 6, no. 1, pp. 46–53, 2002.
- [16] U. R. Acharya, D. Anand, P. S. Bhat, and U. C. Niranjana, "Compact storage of medical images with patient information," *IEEE Transactions on Information Technology in Biomedicine*, vol. 5, no. 4, pp. 320–323, 2001.
- [17] X. Kong and R. Feng, "Watermarking medical signals for telemedicine," *IEEE Transactions on Information Technology in Biomedicine*, vol. 5, no. 3, pp. 195–201, 2001.
- [18] A. Giakoumaki, S. Pavlopoulos, and D. Koutsouris, "Multiple image watermarking applied to health information management," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 4, pp. 722–732, 2006.
- [19] H. M. Chao, C. M. Hsu, and S. G. Miaou, "A data-hiding technique with authentication, integration, and confidentiality for electronic patient records," *IEEE Transactions on Information Technology in Biomedicine*, vol. 6, no. 1, pp. 46–53, 2002.
- [20] S. G. Miaou, C. M. Hsu, Y. S. Tsai, and H. M. Chao, "A secure data hiding technique with heterogeneous data-combining capability for electronic patient records," in *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 280–283, Chicago, Ill, USA, July 2000.
- [21] L. T. Ko, J. E. Chen, Y. S. Shieh, M. Scalia, and T. Y. Sung, "A novel fractional discrete cosine transform based reversible watermarking for healthcare information management systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 757018, 2012.
- [22] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079–1107, 1999.
- [23] R. J. Anderson and F. A. P. Petitcolas, "On the limits of steganography," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 474–481, 1998.
- [24] E. T. Lin and E. J. Delp, "A review of fragile image watermarks," in *Proceedings of the ACM Multimedia Security Workshop*, pp. 47–51, Orlando, Fla, USA, 1999.
- [25] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
- [26] 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.
- [27] E. G. Bakhoun and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.
- [28] E. G. Bakhoun and C. Toma, "Dynamical aspects of macroscopic and quantum transitions due to coherence function and time series events," *Mathematical Problems in Engineering*, vol. 2010, Article ID 428903, 13 pages, 2010.
- [29] M. Li, C. Cattani, and S. Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 13 pages, 2011.
- [30] M. Li, "Four transformation patterns of two-dimensional function on wavelet basis," *International Journal Engineering and Interdisciplinary Mathematics*, vol. 1, no. 2, pp. 107–110, 2009.
- [31] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

Review Article

Modeling of Biological Intelligence for SCM System Optimization

Shengyong Chen,¹ Yujun Zheng,¹ Carlo Cattani,² and Wanliang Wang¹

¹ College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China

² Department of Mathematics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano, Italy

Correspondence should be addressed to Shengyong Chen, sy@ieee.org

Received 15 September 2011; Accepted 15 October 2011

Academic Editor: Maria Crisan

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

This article summarizes some methods from biological intelligence for modeling and optimization of supply chain management (SCM) systems, including genetic algorithms, evolutionary programming, differential evolution, swarm intelligence, artificial immune, and other biological intelligence related methods. An SCM system is adaptive, dynamic, open self-organizing, which is maintained by flows of information, materials, goods, funds, and energy. Traditional methods for modeling and optimizing complex SCM systems require huge amounts of computing resources, and biological intelligence-based solutions can often provide valuable alternatives for efficiently solving problems. The paper summarizes the recent related methods for the design and optimization of SCM systems, which covers the most widely used genetic algorithms and other evolutionary algorithms.

1. Introduction

Supply chains are a kind of network with facilities and distribution entities (suppliers, manufacturers, distributors, retailers). The supply chain performs the functions of procurement of raw materials, transformation of raw materials into intermediate and finished products, and distribution of finished products to customers [1]. Due to rising product and market complexity, expanding competition, shorter product lifecycles, and changing customer demands, today's supply chain management (SCM) systems increasingly involve complex sets of objectives and constraints, and variations of uncertainty and randomization [2], and thus static and/or centralized models are insufficient to effectively plan, coordinate, and optimize activities in a supply chain.

Modeling and optimization of an SCM system provide a critical support for decision making in a competitive market. According to [3], the basic approaches to supply chain modeling and optimization can be divided into five classes:

- (1) fundamental formulation of supply chains,
- (2) integer-mixed programming optimization,
- (3) stochastic programming,

- (4) heuristic methods,
- (5) simulation-based methods.

Traditional simulation methods for large-scale complex systems require huge amounts of computing resources [4]. In recent years, bioinspired methods have gained increasing interest in the research of modeling and optimization for SCM systems which are typically dynamic, open self-organizing systems maintained by flows of information, materials, goods, funds, and energy. Bioinspired and living system mechanisms, such as learning, growth, evolution, collaboration, and competition, bring an innovative solution for the analysis and improvement of emergent complex behaviors in virtually computing modules. In this paper, we review the recent major accomplishments in bioinspired solution methods and tools for SCM systems. In particular, we concentrate on the modeling and optimization methods based on a class of metaheuristics inspired by biologically living beings, including genetic algorithms (GAs) [5], evolutionary programming (EP) [6], evolution strategies (ESs) [7, 8] and differential evolution (DE) [9], swarm intelligence [10, 11], and artificial immune [12]. These heuristic methods usually do not require deep mathematical knowledge, and have been demonstrated to be quite useful and efficient

in optimization search for large-scale problems. We believe that this work will help researchers and practitioners to gain knowledge about the major developments emerged throughout the years and find valuable approaches that can be referred in the research or applied in the practice of SCM modeling and optimization.

The rest of the paper is synthesized as follows: Section 2 describes modeling and optimization methods based on GA, Section 3 depicts three other evolution-related methods including EP, ES, and DE, Section 4 describes methods based on swarm intelligence, and Section 5 introduces some other biological methods. Finally we discuss some future trends in Section 6 and conclude in Section 7.

2. Genetic Algorithms

2.1. Design of Supply Chain Network. Network design plays a key role in achieving efficient and effective management of SCM systems. Typically, a supply chain can be represented as a form of multistage-based structure, the optimal design of which has been recognized as *NP*-hard problems that combine the multiple choice knapsack problem with the capacitated location-allocation problem [13]. The first attempt to use the GA approach to solve the SCM network design problems has been proposed by Zhou et al. [14]. They developed a balanced star-spanning forest formulation for encoding the solutions and then used uniform crossover and exchange mutation operators in the algorithm. The experiment showed that for a maximum of 10 distributors and 100 customers, the algorithm can balanced all the distributors. Gen et al. [15] proposed a set of spanning tree-based for a class of network design problems such as degree-constrained minimum spanning tree problems, capacitated minimum spanning tree problems, fixed charge transportation problems, and network topological design problems, which were applied to some real-world SCM systems.

A multistage distribution problem is a standard one with supply chain network design. Many works focus on the two-stage supply chain distribution problem [16–20], which can be represented in Figure 1. That is, each of the m plants can ship to any of the n distribution centers, and each of the n distribution centers can ship to any of the p customers. Typically, each plant i has s_i units of supply, each customer k has d_k units of demand, and each distribution centre j has t_j units of stocking capacity, and the main purpose is to minimize the total cost which may include transportation cost, fixed cost, and transportation time.

Aiming at the demand allocation optimization of a two-stage and single product supply chain design problem, Chan et al. [18] implemented a multicriteria optimization algorithm which combines GA with the decision-making technique of the Analytic Hierarchy Process (AHP). Jawahar and Balaji [19] considered a two-stage distribution problem of a supply chain that is associated with a fixed charge and presented a GA that evolves the solution for best fitness of total cost of distribution. By fine-tuning its parameters, the algorithm can work out optimal solutions for small-size problem instances, but its performance on large-size

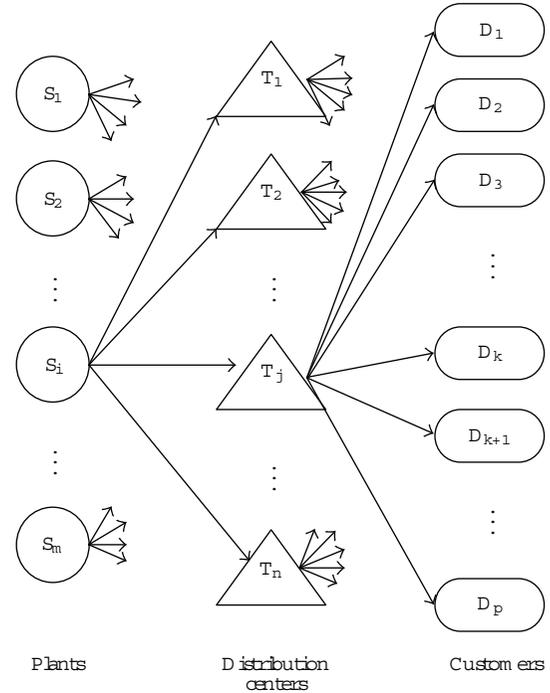


Figure 1: Illustration of the two-stage supply chain network design problems.

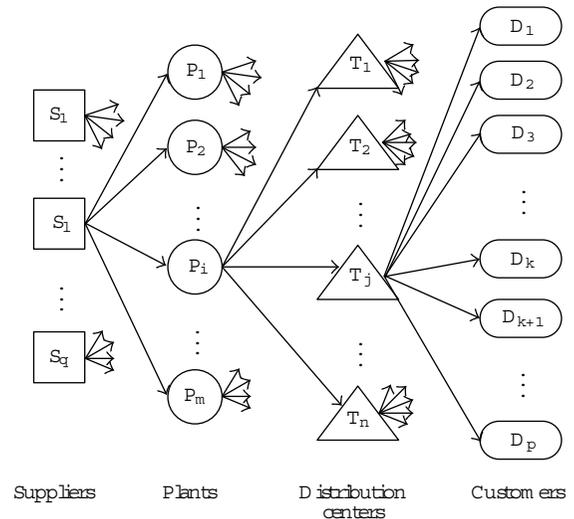


Figure 2: Design problem in three-stage supply chain networks.

instances has not been demonstrated. Feng and Zhang [20] extended the problem by involving multiple transportation modes and proposed a GA that can deal with middle-size problem instances.

Since that today's market environment becomes more and more complex, supply chains with multiple (three or more) stages are common. To model a single product three-stage supply chain network (as illustrated in Figure 2), Altıparmak et al. [21] proposed a GA that uses a three-segment chromosome string which is decoded through

a backward procedure from the third stage up to the first stage connecting suppliers and plants. They also extended their priority-based encoding in a new GA for the multi-product case in [22]. However, their decoding structure can require some repair procedures when the upper limits are exceeded; otherwise the capacities will be not enough to meet the customers' demand. Costa et al. [23] presented a new chromosome encoding and a complementary decoding procedure able to overcome the drawbacks and thus improve the efficiency and effectiveness of three-stage supply chains.

Considering that in a multistage supply chain network the flow can be only transferred between two consecutive stages, Yeh [24] presented a memetic algorithm for the design optimization of distribution network, which combined GA, linear programming, a greedy-heuristic based method, and three local search methods. Yun et al. [25] indicated that, when applying GA to multistage-based supply chains, conventional GA can do global search but there is no way for local search around the convergence area generate, and they proposed a new hybrid GA with an adaptive local search scheme which can automatically control whether local search technique is used in GA loop or not, and thus can reliably solve various optimization problems without trial-and-error experimentation.

2.2. Supply Chain Planning and Scheduling. The success of SCM highly depends on the timely and efficient production/distribution in the supply chain network, which can be typically regarded as a combination of planning and scheduling problems, each seriously affected by nearly prohibitive combinatorial complexity. In general, the development of an effective approach for SCM planning and scheduling can be divided into two steps:

(1) developing a precise mathematical model of the problem considered, explaining the objective function(s), decision variables, and problem constraints;

(2) selecting one or more effective metaheuristics suited to the problem, and developing a problem-solving algorithm-based on the heuristics.

Şerifoğlu and Ulusoy [26] considered the problem of scheduling independent jobs on identical parallel machines to minimize earliness/tardiness, which incorporates distinct due dates and arrival times for jobs, different processing rates for machines, and sequence-dependent set-up times into the problem formulation. They developed a GA with a new multicomponent uniform order-based crossover (MCUOX) operator and demonstrated that the GA outperforms other simple heuristic methods especially in larger-sized problems. Min and Cheng [27] also proposed a GA for solving the problem of scheduling identical parallel machines, where the objective is formulated as the minimization of the makespan.

García et al. [28] considered the problem of scheduling a single production plant in order to satisfy delivery time constraints. They proposed two approaches, an exact method for simple cases and a GA for instances of more realistic size. Feng et al. [29] considered the problem of scheduling a single depot equipped with a fleet of vehicles with identical capacity and fixed loading/unloading times. They proposed a GA for searching a production sequence maximizing a predefined

performance index, taking truck waiting times and a penalty for violating the unloading continuity of multitruck orders into consideration.

In [30], Lee et al. studied an operation-level advance planning and scheduling problem in supply chain, the model of which is to determine the best schedule using alternative operation sequences and machines, considering scheduling with outsourcing and strong constraints of the due dates of customer orders. They developed a genetic algorithm-based heuristic to solve it. Karabuk [31] studied the assignment of jobs to suppliers and aims to determine optimal production scheduling that minimizes the makespan, where each supplier requires a different length of time to process each job. The research proposed an adaptive GA with a new dominated gene crossover operator to solve this problem. Zegordi et al. [32] considered the scheduling of products and vehicles in a two-stage supply chain environment, assuming that the various output products occupy different percentages of each vehicle's capacity and proposed a gendered GA with two different chromosomes with nonequivalent structures that performs better than standard GA with a unique chromosomal structure.

There are also a number of reports on the applications of genetic algorithm in different trades, for example, [33–40]. Considering a semiconductor supply chain for reflecting the nonlinear throughput time of manufacturing, Chidambaram and Armbruster [33] proposed a hybrid LP and GA framework to solve the non-linear programming problem in order to avoid significant differences between the planned and realized output. Naso et al. [34] considered the problem of finding an optimized schedule for the just-in-time production and delivery of ready-mixed concrete on a set of distributed and coordinated production centers. They proposed a hybrid evolutionary algorithm in which the GA constitutes the core of the search strategy, while multiple heuristic rules called in specific circumstances contribute to reconstruct a feasible solution that satisfies all the constraints and objectives. The algorithm can guarantee the determination of a feasible schedule for any given set of requests and can address the highly complex scheduling problem of an entire supply-chain for just-in-time production. Dong and Ding [35] introduced a dynamic berth allocation model for container terminal and proposed a GA-based heuristic method that improves the existing research on static berth allocation models. The GA-based approach explores detailed capabilities of a complex problem solution using these two encoding methods. Recently, Delavar et al. [36] considered coordinated scheduling of production and air transportation and proposed two GA approaches to optimize customer service at minimum total cost: the first uses a portion of the search time to seek for the best transportation allocation and dedicates the remaining time to search for the best production-transportation solution; the second allocates some definite number of generations to search for the best production-transportation solution only after finding a better solution in each generation of transportation. In both algorithms, the Taguchi parameter design method was employed to adjust the parameters.

2.3. Multiobjective Optimization. In most cases, the design, planning, and scheduling of complex supply chains will involve trade-offs among different goals. Since 2000, multiobjective optimization of SCM systems has gained a lot of interests of the researchers. Generally, for a multiobjective optimization problem we need to search for a set of Pareto optimal solutions rather than a single optimal one. Using a minimization problem for illustration, let f_1, f_2, \dots, f_m be objective functions; a solution \mathbf{x} is said to dominate \mathbf{y} if and only if

- (i) $f_i(\mathbf{x}) \leq f_i(\mathbf{y})$ for all $i \in \{1, 2, \dots, m\}$,
- (ii) $f_i(\mathbf{x}) < f_i(\mathbf{y})$ for existing $i \in \{1, 2, \dots, m\}$.

Regarding this, a Pareto-optimal front of the problem consists of all solutions for which the corresponding objective vectors cannot be improved in a given dimension without worsening another [41].

A typical multiobjective genetic optimization algorithm was proposed by Chan and Chung [42] for simultaneously minimizing the total cost of the system, total delivery days, and the equity of the capacity utilization ratio for manufacturers. In [43] Chan et al. proposed a hybrid GA for production and distribution problem, which utilized AHP to construct these criteria and calculate the fitness value of chromosome, and considered operating cost, service level, and resources utilization as objectives. Considering that all organizational units that participate on a single SC network are distributed by nature, constrained, and self-interested, Al-Mutawah et al. [44] used a distributed multiobjective GA to solve a three-subchain optimization problem, and their test results showed that the distributed GA provided an improved computational performance, because real-world supply chain applications are distributed in nature-distributed approach.

Altıparmak et al. [45] considered SCM optimization problem with three objectives: minimizing total cost, maximizing of customer services, and maximizing capacity utilization balance for distribution centers. To deal with the objectives and enable the decision maker for evaluating a number of alternative solutions, they proposed a GA which was designed to generate Pareto-optimal solutions considering two different weight approaches. Farahani and Elahipanah [46] adopted a hybrid non-dominated sorting GA to optimize total cost and service level for JIT distribution in a supply chain, whose results were compared with Lingo software to evaluate the performance of proposed algorithm. Che and Chiang [47] established a multiobjective optimization mathematical model for the build-to-order supply chain model which are defined as “the system that produces goods and services based on individual customer requirements in a timely and cost competitive manner by leveraging global outsourcing, the application of information technology and through the standardization of components and delayed product differentiation strategies” [48]. Considering three evaluation criteria including costs, delivery time, and quality, they proposed a modified Pareto GA to improve efficiency of the crossover and mutation operators of basic Pareto GA.

3. EP, ES, and DE

Besides GA, other evolutionary algorithmic methods have also been applied to many SCM modeling and optimization problems. EP was devised in order to evolve finite state machines for the prediction of events on the basis of former observations and has been demonstrated useful for searching the optimum of nonlinear functions [49]. Huang and Lu [50] proposed an interactive EP approach based on the relativistic error and selection of some other parameters to improve initial value determination, mutation, and variance parameter selective operation. The simulation results of supply chains showed that the improved EP is much more appropriate of nonlinear model with a great volume of data. Based on EP approach, Li et al. [51] proposed a heuristic strategic safety stock optimization algorithm for reverse logistics SCM considering the modeling complexity of external as well as internal product returns and reuses of supply chains.

Original ES uses a mutation operator that produces a single descendent from a given ancestor [7, 8], denominated ES-(1 + 1), and was progressively generalized to ES-($\mu + \lambda$), that is, several ancestors ($\mu > 1$) and descendents ($\lambda > 1$) in each generation. Homberger [52] proposed an approach that combines the (1 + λ)-selection procedure with the Borda maximin voting rule, to coordinate decentral planning of a group of independent and self-interested decision makers, who are searching for an agreeable contract regarding multiple interdependent issues, in the case of asymmetric information presented. In [53] Dalkilic et al. developed an ES algorithm to solve a multiple-supplier multiple-item problem with stochastic lead times, which was successfully applied to some real-world healthcare SCM cases.

DE approach combines simple arithmetic operators with the classical operators of crossover, mutation, and selection to evolve a randomly generated starting population to a final solution. It is similar to a ($\mu + \lambda$) ES, but in DE the mutation is not done via some separately defined probability density function [54]. Routroy and Kodali [55] developed a DE algorithm for minimizing the total systemwide cost, which consists of supply chain inventory capital, supply chain ordering/set-up cost, and supply chain stock-out cost. The result showed that the algorithm helps in determining ordering/production quantity and inventory/service level that should be maintained by each member of the supply chain. The algorithm was further extended for multiechelon supply chain inventory problems [56] and the problems with demand and leadtime uncertainty [57]. Prasertwattana and Shimizu also [58] applied a similar DE algorithm to optimize material ordering and inventory control of SCM systems.

DE was used for multiobjective optimization of SCM systems first by Babu and Gujarathi [59]. In their algorithm, crossover is carried out between the target vector and the noisy random vector to generate a trial vector, the cost of trial and target vectors is compared, and the variables corresponding to best cost are passed into next generation. The algorithm was successfully applied to a three-stage SCM problem. Dos Santos Coelho and Lopes [60] firstly developed a chaotic DE algorithm for the optimization of

a supply chain, which was based on different DE approaches combined with chaotic sequences and led to better results than basic DE approaches. In [61] Xu et al. proposed a migration DE algorithm by imitating nomadic migration for this supply chain problem and presented an ensemble method based on different DE methods for not only avoiding the premature convergence but also improving the global search capability.

Falcone et al. [54] compared the performance of GA, EP, ES, and DE based on a case of integrated production-inventory-distribution SCM system. Their results showed that the robustness of the evolutionary methods is in general, and the efficiency of DE, in particular, suggests their great utility for the supply chain optimization problem.

4. Swarm Intelligence

4.1. Ant Colony Optimization. Ant colony optimization (ACO) algorithm mimics the behavior of real ants living in colonies that communicate with each other using pheromones in order to accomplish complex tasks such as establishing a shortest path from the nest to food sources [62]. Silva et al. [63] proposed an ACO algorithm for distributed optimization of a logistic system, but the work only considered only the allocation of suppliers in the system. In a successive work [64], the authors modeled a distributed optimization problem for a generic supply chain with suppliers, logistics, and distributors and developed an ACO algorithm that allows the exchange of information between different optimization problems by means of a pheromone matrix. The experimental results showed that the approach can significantly improve global supply chain performance with respect to other simple decentralized approaches.

Wang [65] studied the partner selection and production-distribution planning problem in a supply chain with the losses of production, which is called the defective supply chain. He developed for this problem a two-phase ACO algorithm, which finds out the combination of the maximum yield rate and the minimum number of partners in the first phase and implements the distribution with the partners and seeks out the minimum value of T-score in the second phase. Comparative numerical experiment showed that his algorithm achieves better performance than the common single-phase ACO algorithms.

In a very recent work, Moncayo-Martínez and Zhang [66] studied the multiobjective ACO for supply chain optimization. They considered the problem for minimizing the total cost while keeping the total lead-times within required delivery due dates. They formulated the design problem into an ACO optimization form and implemented a number of ant colonies in a sequence to explore the solution space and search for successively better nondominated set of supply chain designs.

4.2. Particle Swarm Optimization. Particle swarm optimization (PSO) [10] is another population-based global optimization technique that enables a number of individual solutions, called particles, to move through a hyperdimensional search space to search for the optimum. Each par-

ticle has a position vector and a velocity vector, which are adjusted at iterations by learning from a local best found by the particle itself and a current global best found by the whole swarm. Modeling a system where multiple-candidate solution coexists and collaborates simultaneously, PSO approaches embed problem-solving attempts in a social network and are suitable in nature for the optimization of very complex systems [67] and thus have been successfully applied in the research of SCM, for example, [68–74].

Izquierdo et al. [68] applied a PSO algorithm to a supply chain for searching optimal biomass flows from sources to energy production plants. Kadavevaramath et al. [70] proposed a PSO algorithm for the modeling and optimization of a four-stage supply chain and gained satisfying results. Bachlaus et al. [71] considered the design of a multiechelon supply chain network that integrates production, distribution and logistics activities and developed a hybrid PSO algorithm based on Taguchi robust design optimization tool [75]. The algorithm incorporates the characteristics of statistical design of experiments and random search techniques, which is an attractive way for determining flexible location and distribution strategies.

In [73] Sinha et al. considered the optimization of resource allocation for agents in a petroleum supply chain. They developed a coevolutionary PSO algorithm with two populations, and the decision vectors and the Lagrangian multipliers are taken to be constant in the first population and to be variables in the second population. The algorithm also uses a Cauchy random number distribution which is proved to be much better than a Gaussian distribution.

Soares [76] et al. utilized four kinds algorithms (including EP, GA, PSO, and EDA—estimation of distribution algorithm) to solve a multiple-retailer SCM problem, which is for finding an optimal balance of quantities ordered from suppliers and acceptable lead time costs while taking into account limiting factors such as the time each retailer will wait for a backorder. According to the results on the test-suite, three PSO algorithms of the 32 attempted algorithms demonstrate great flexibility and high performance.

4.3. Artificial Bee Colony Optimization. There are several optimization algorithms [77–80] that simulate the intelligent foraging behavior of a honeybee swarm. A more recent and popular approach is the artificial bee colony (ABC) algorithm that divides a bee colony into three groups, namely, employed bees exploiting on current food sources, onlookers waiting in the hive for choosing existing food sources, and scouts bees exploring new food sources. In [81] Kumar et al. analyzed the complexities of a remanufacturing problem in which the return rate is a function of environmental factor and proposed an ABC algorithm for the problem model. In their test results, the ABC algorithm significantly outperformed a PSO algorithm for comparison. In [82], Pal et al. considered a problem of integrated procurement, production, and shipment planning for a three-echelon supply chain. They developed two ABC-based algorithms for optimizing the order scheduling and production-shipment planning to achieve a minimum cost.

Considering the problem of a milk production, scheduling, and supply network design with extensively multiple conflicting objectives, Banerjee et al. [83] proposed a Pareto BCO approach, which was demonstrated to be better than some other bioinspired algorithms by simulation and comparison. For improving machine utilization and reducing cycle-time in manufacturing industry, Li et al. [84] applied a Pareto bee colony optimization (BCO) algorithm for a multiobjective flexible job shop scheduling problem and gained good computational result.

5. Other Methods from Biological Intelligence

Provoked by the theoretical immunology, observed immune functions, principles and models, artificial immune system (AIS) stimulates the adaptive immune system of a living creature to unravel the various complexities in real-world engineering optimization problems [85]. Shukla et al. [86] employed an AIS approach to a batch sequencing problem in a multistage supply chain, which considers three objectives including minimization of lead time, blocking time, and due date violation, and the experimental results showed that the AIS outperforms GA and simulated annealing (SA).

Prakash and Deshmukh [85] considered a multiple vendor transportation problem with time and cost criteria and proposed an AIS algorithm strengthened by a fuzzy logic controller (FLC) to solve the multicriteria problem. AIS works as an evolutionary search algorithm to find out the Pareto optimal front, whereas FLC is implemented to change the hypermutation rate adaptively on the basis of the fitness values at each iteration. They also employed a web-based supply chain to facilitate the SCM enterprise by e-learning.

In [87] Hajiaghayi-Keshteli considered a two-stage supply chain network of distribution centers and customers. To solve the problem for selecting some potential places as distribution centers in order to supply demands of all customers with minimum opening cost plus shipping cost, he, respectively, developed a GA and an AIS algorithm, and the results showed that the AIS algorithm exhibits robust performance improvements in large size problems versus GA.

6. Discussion

We have summarized the main bioinspired methods for SCM system design and optimization. It is deserved to note that swarm-based methods and artificial immune systems are not yet mature and thus are expected to gain more research interests. With the increasing importance and complexity of SCM systems, researchers are facing the challenges to promote the performance, reliability, and scalability of SCM problem-solving methods, and here we highlight the following future trends in bioinspired computation in SCM systems.

6.1. Hybrid. Different bioinspired methods have different design principles and application areas. As we mentioned in previous sections, there are a number of studies that

exploit the strengths of several individual methods to obtain a more powerful approach to dealing with complex SCM problems, and to a great extent, these hybrids methods are shown to be more competitive than individual methods. It can be anticipated that future research will continuously put great emphasis on the hybridization of bioinspired methods, for example, swarm-based evolutionary algorithms [88–90], and the hybridization of bioinspired methods with other approaches such as local search [91], tabu search [92], and simulated annealing [93].

6.2. Extension with New Computing Paradigms. We are seeing that innovative informational/computational paradigms, such as chaotic systems, quantum informatics, and DNA computing, provide valuable inspiration to create new heuristics for complex optimization problems including a host of *NP*-hard problems. Thus, the extensions of current bioinspired methods based on these new paradigms are expected to achieve dramatic improvement on computational performance. For example, chaotic sequencing and local search operations have been successfully applied for helping evolutionary algorithms avoiding premature convergence effectively [94, 95]. Also, quantum-inspired evolutionary algorithms are regarded as one of the three main research areas related to the complex interaction between quantum computing and evolutionary algorithms [96–98] and have been applied in some SCM optimization problems in very recent research [99]. These approaches are expected to show great promises for the future.

7. Conclusion

Today's SCM systems have to deal with ever-changing markets and intrinsic structural complexity emerging from virtually infinite number of interacting entities. Therefore, the community requires effective artificial intelligence methods and tools for modeling and optimizing large-scale complex supply chains. The paper has reviewed the recent development of bioinspired methods in SCM applications. Typical illustrations are addressed for evolutionary algorithms including GA, EP, ES, and DE, swarm-based intelligent algorithms including ant colony, particle swarm and artificial bee colony, and other bioinspired methods like AIS. Representative works are summarized for helping readers to have a general overview of the state of the art and to easily refer to suitable methods in practical solutions. Over the last decade, bioinspired methods have experienced a rapid growth and have successfully applied to the design and optimization of highly complex systems such as SCM systems. The fruits of these researches are continuously becoming new technological solutions to new open problems, and the full potential is far from being reached.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61105073, 61173096, 60874074,

60870002), Zhejiang Provincial S&T Department (2010-R10006, 2010C33095), and Zhejiang Provincial Natural Science Foundation (R1110679).

References

- [1] H. Sarimveis, P. Patrinos, C. D. Tarantilis, and C. T. Kiranoudis, "Dynamic modeling and control of supply chain systems: a review," *Computers and Operations Research*, vol. 35, no. 11, pp. 3530–3561, 2008.
- [2] Y. Zheng, J. Wang, and J. Xue, "A-team based supply chain management agent architecture," *International Journal on Artificial Intelligence Tools*, vol. 18, no. 6, pp. 801–823, 2009.
- [3] M. Dong, *Process modeling, performance analysis and configuration simulation in integrated supply chain network design*, Ph.D. thesis, Virginia Polytechnic Institute and State University, Blacksburg, Va, USA, 2001.
- [4] E. Sanchez, A. Perez-Urbe, A. Upegui et al., "PERPLEXUS: pervasive computing framework for modeling complex virtually-unbounded systems," in *Proceedings of the 2nd NASA/ESA Conference on Adaptive Hardware and Systems (AHS '07)*, pp. 587–591, August 2007.
- [5] J. H. Holland, *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, Mich, USA, 1975.
- [6] D. B. Fogel, "Introduction to simulated evolutionary optimization," *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 3–14, 1994.
- [7] I. Rechenberg, *Evolutionstrategie: Optimierung Technischer Systeme Nach Prinzipien der Biologischen Evolution*, Frommberg-Holzboog, Stuttgart, Germany, 1973.
- [8] H. P. Schwefel, *Kybernetischer Evolution als Strategie der experimentellen Forschung in der Strömungstechnik*, Diploma thesis, Technical University of Berlin, Berlin, Germany, 1975.
- [9] R. Storn and K. Price, "Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces," Tech. Rep. TR-95-012, International Computer Science Institute, Berkeley, Calif, USA, 1995.
- [10] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, Perth WA, Australia, December 1995.
- [11] M. Clerc, *Particle Swarm Optimization*, ISTE, London, UK, 2006.
- [12] J. D. Farmer, N. H. Packard, and A. S. Perelson, "The immune system, adaptation, and machine learning," *Physica D: Nonlinear Phenomena*, vol. 22, no. 1–3, pp. 187–204, 1986.
- [13] M. Gen and R. Cheng, *Genetic algorithms and engineering design*, Wiley-Interscience, New York, NY, USA, 1997.
- [14] G. Zhou, H. Min, and M. Gen, "The balanced allocation of customers to multiple distribution centers in the supply chain network: a genetic algorithm approach," *Computers and Industrial Engineering*, vol. 43, no. 1-2, pp. 251–261, 2002.
- [15] M. Gen, A. Kumar, and J. R. Kim, "Recent network design techniques using evolutionary algorithms," *International Journal of Production Economics*, vol. 98, no. 2, pp. 251–261, 2005.
- [16] A. Cakravastia, I. S. Toha, and N. Nakamura, "A two-stage model for the design of supply chain networks," *International Journal of Production Economics*, vol. 80, no. 3, pp. 231–248, 2002.
- [17] G. Wang, Y. Liu, and M. Zheng, "Fuzzy two-stage supply chain problem and its intelligent algorithm," in *Proceedings of the 6th International Symposium on Neural Networks*, vol. 5552, part 2 of *Lecture Notes in Computer Science*, pp. 15–24, Wuhan, China, 2009.
- [18] F. T. S. Chan, S. H. Chung, and S. Wadhwa, "A heuristic methodology for order distribution in a demand driven collaborative supply chain," *International Journal of Production Research*, vol. 42, no. 1, pp. 1–19, 2004.
- [19] N. Jawahar and A. N. Balaji, "A genetic algorithm for the two-stage supply chain distribution problem associated with a fixed charge," *European Journal of Operational Research*, vol. 194, no. 2, pp. 496–537, 2009.
- [20] C. Feng and Y. Zhang, "A genetic algorithm of two-stage supply chain distribution problem associated with fixed charge and multiple transportation modes," in *Proceedings of the 5th International Conference on Natural Computation (ICNC '09)*, pp. 76–80, Tianjin, China, August 2009.
- [21] F. Altıparmak, M. Gen, L. Lin, and T. Paksoy, "A genetic algorithm approach for multi-objective optimization of supply chain networks," *Computers and Industrial Engineering*, vol. 51, no. 1, pp. 196–215, 2006.
- [22] F. Altıparmak, M. Gen, L. Lin, and I. Karaoglan, "A steady-state genetic algorithm for multi-product supply chain network design," *Computers and Industrial Engineering*, vol. 56, no. 2, pp. 521–537, 2009.
- [23] A. Costa, G. Celano, S. Fichera, and E. Trovato, "A new efficient encoding/decoding procedure for the design of a supply chain network with genetic algorithms," *Computers and Industrial Engineering*, vol. 59, no. 4, pp. 986–999, 2010.
- [24] W. C. Yeh, "An efficient memetic algorithm for the multi-stage supply chain network problem," *International Journal of Advanced Manufacturing Technology*, vol. 29, no. 7-8, pp. 803–813, 2006.
- [25] Y. Yun, C. Moon, and D. Kim, "Hybrid genetic algorithm with adaptive local search scheme for solving multistage-based supply chain problems," *Computers and Industrial Engineering*, vol. 56, no. 3, pp. 821–838, 2009.
- [26] F. Sivrikaya-Şrifioğlu and G. Ulusoy, "Parallel machine scheduling with earliness and tardiness penalties," *Computers and Operations Research*, vol. 26, no. 8, pp. 773–787, 1999.
- [27] L. Min and W. Cheng, "A genetic algorithm for minimizing the makespan in the case of scheduling identical parallel machines," *Artificial Intelligence in Engineering*, vol. 13, no. 4, pp. 399–403, 1999.
- [28] J. M. Garcia, S. Lozano, K. Smith, T. Kwok, and G. Villa, "Coordinated scheduling of production and delivery from multiple plants and with time windows using genetic algorithms," in *Proceedings of the 9th International Conference on Neural Information Processing*, vol. 3, pp. 1153–1158, Singapore, 2002.
- [29] C. W. Feng, T. M. Cheng, and H. T. Wu, "Optimizing the schedule of dispatching RMC trucks through genetic algorithms," *Automation in Construction*, vol. 13, no. 3, pp. 327–340, 2004.
- [30] Y. H. Lee, C. S. Jeong, and C. Moon, "Advanced planning and scheduling with outsourcing in manufacturing supply chain," *Computers and Industrial Engineering*, vol. 43, no. 1-2, pp. 351–374, 2002.
- [31] S. Karabuk, "Modeling and optimizing transportation decisions in a manufacturing supply chain," *Transportation Research Part E: Logistics and Transportation Review*, vol. 43, no. 4, pp. 321–337, 2007.
- [32] S. H. Zegordi, I. N. K. Abadi, and M. A. B. Nia, "A novel genetic algorithm for solving production and transportation

- scheduling in a two-stage supply chain,” *Computers and Industrial Engineering*, vol. 58, no. 3, pp. 373–381, 2010.
- [33] R. Chidambaram and D. Armbruster, “Application of genetic algorithms to semiconductor supply chain planning,” in *Proceedings of the International Conference on Algorithmic Mathematics and Computer Science (AMCS '05)*, pp. 77–83, Las Vegas, Nev, USA, June 2005.
- [34] D. Naso, M. Surico, B. Turchiano, and U. Kaymak, “Genetic algorithms for supply-chain scheduling: a case study in the distribution of ready-mixed concrete,” *European Journal of Operational Research*, vol. 177, no. 3, pp. 2069–2099, 2007.
- [35] L. Dong and H. Ding, “A genetic algorithm based dynamic berth allocation strategy for container terminals,” in *Proceedings of the 8th International Conference of Chinese Logistics and Transportation Professionals*, Chengdu, China, August 2008.
- [36] M. R. Delavar, M. Hajiaghahi-Keshteli, and S. Molla-Alizadeh-Zavardehi, “Genetic algorithms for coordinated scheduling of production and air transportation,” *Expert Systems with Applications*, vol. 37, no. 12, pp. 8255–8266, 2010.
- [37] C. L. Yang and Y. K. Chen, “Using genetic algorithm for better route arrangement a case of Taiwan pelican express company,” in *Proceedings of the IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI '06)*, pp. 669–673, Shanghai, China, June 2006.
- [38] S. J. Sadjadi, M. Jafari, and T. Amini, “A new mathematical modeling and a genetic algorithm search for milk run problem (an auto industry supply chain case study),” *International Journal of Advanced Manufacturing Technology*, vol. 44, no. 1–2, pp. 194–200, 2009.
- [39] M. Azarmi, A. Jahanmiri, and S. Vakili, “An optimization model for refinery production scheduling (design a software using genetic algorithms & MILP),” in *Proceedings of the 8th World Congress of Chemical Engineering*, August 2009.
- [40] H. C. W. Lau, T. M. Chan, W. T. Tsui, and W. K. Pang, “Application of genetic algorithms to solve the multidepot vehicle routing problem,” *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 2, Article ID 4840417, pp. 383–392, 2010.
- [41] V. Chankong and Y. Y. Haimes, *Multiobjective Decision Making Theory and Methodology*, Elsevier, New York, NY, USA, 1983.
- [42] F. T. S. Chan and S. H. Chung, “A multi-criterion genetic algorithm for order distribution in a demand driven supply chain,” *International Journal of Computer Integrated Manufacturing*, vol. 17, no. 4, pp. 339–351, 2004.
- [43] F. T. S. Chan, S. H. Chung, and S. Wadhwa, “A hybrid genetic algorithm for production and distribution,” *Omega*, vol. 33, no. 4, pp. 345–355, 2005.
- [44] K. Al-Mutawah, V. Lee, and Y. Cheung, “Modeling supply chain complexity using a distributed multi-objective genetic algorithm,” in *Proceedings of International Conference on Computational Science and Its Applications*, vol. 3980 of *Lecture Notes in Computer Science*, pp. 586–595, Melbourne, Australia, 2006.
- [45] F. Altıparmak, M. Gen, L. Lin, and T. Paksoy, “A genetic algorithm approach for multi-objective optimization of supply chain networks,” *Computers and Industrial Engineering*, vol. 51, no. 1, pp. 196–215, 2006.
- [46] R. Z. Farahani and M. Elahipanah, “A genetic algorithm to optimize the total cost and service level for just-in-time distribution in a supply chain,” *International Journal of Production Economics*, vol. 111, no. 2, pp. 229–243, 2008.
- [47] Z. H. Che and C. J. Chiang, “A modified Pareto genetic algorithm for multi-objective build-to-order supply chain planning with product assembly,” *Advances in Engineering Software*, vol. 41, no. 7–8, pp. 1011–1022, 2010.
- [48] A. Gunasekaran and E. W. T. Ngai, “Modeling and analysis of build-to-order supply chains,” *European Journal of Operational Research*, vol. 195, no. 2, pp. 319–334, 2009.
- [49] D. Fogel and K. Chellapilla, “Revisiting evolutionary programming,” in *Applications and Science of Computational Intelligence*, vol. 3390, pp. 2–11, Orlando, Fla, USA, March 1998.
- [50] X. Y. Huang and Z. Lu, “Interactive evolutionary programming and its application in supply chain,” *Dongbei Daxue Xuebao/Journal of Northeastern University*, vol. 21, no. 5, pp. 569–572, 2000.
- [51] J. Li, J. R. Wang, Z. W. Hu, and J. H. Zhang, “EP-based optimization of strategic safety stocks in reverse logistics systems,” *Journal of Dong Hua University*, vol. 21, no. 5, pp. 65–68, 2004.
- [52] J. Homberger, “A generic coordination mechanism for lot-sizing in supply chains,” *Electronic Commerce Research*, vol. 11, no. 2, pp. 1123–149, 2011.
- [53] V. Dalkilic, E. J. Lodree Jr., A. Ramesh, and G. Sarpkaya, “Order splitting for multiple-supplier multiple-item health-care supply chains,” in *Proceedings of the IIE Annual Conference and Exhibition*, Orlando, Fla, USA, 2006.
- [54] M. A. Falcone, H. S. Lopes, and L. Dos Santos Coelho, “Supply chain optimisation using evolutionary algorithms,” *International Journal of Computer Applications in Technology*, vol. 31, no. 3–4, pp. 158–167, 2008.
- [55] S. Routroy and R. Kodali, “Differential evolution algorithm for supply chain inventory planning,” *Journal of Manufacturing Technology Management*, vol. 16, no. 1, pp. 7–17, 2005.
- [56] S. Routroy and P. Sanisetty, “Inventory planning for a multi-echelon supply chain,” *International Journal of Operational Research*, vol. 2, no. 3, pp. 269–283, 2007.
- [57] S. Routroy and K. C. Maddala, “Multi-echelon supply chain inventory planning with demand and leadtime uncertainty,” *International Journal of Operational Research*, vol. 5, no. 3, pp. 251–264, 2009.
- [58] K. Prasertwattana and Y. Shimizu, “Optimization of material ordering and inventory control of supply chain through an incentive scheme using differential evolution,” *Journal of Japan Industrial Management Association*, vol. 59, no. 4, pp. 283–289, 2008.
- [59] B. V. Babu and A. M. Gujarathi, “Multi-objective differential evolution (MODE) for optimization of supply chain planning and management,” in *Proceedings of the IEEE Congress on Evolutionary Computation (CEC '07)*, pp. 2732–2739, Singapore, September 2007.
- [60] L. Dos Santos Coelho and H. S. Lopes, “Supply chain optimization using chaotic differential evolution method,” in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 3114–3119, Montréal, Canada, October 2006.
- [61] J.-M. Xu, J.-Z. Xiong, Y. Chen, and G.-W. Hu, “Supply chain optimization using migration differential evolution ensemble,” in *Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS '10)*, vol. 3, pp. 755–759, Xiamen, China, 2010.
- [62] A. Colorni, M. Dorigo, and V. Maniezzo, “Distributed optimization by ant colonies,” in *Proceedings of the European Conference on Artificial Life*, pp. 134–142, Paris, France, 1991.
- [63] C. A. Silva, J. M. C. Sousa, T. A. Runkler, and J. M. G. S. Da Costa, “Distributed optimisation of a logistic system and its suppliers using ant colonies,” *International Journal of Systems Science*, vol. 37, no. 8, pp. 503–512, 2006.

- [64] C. A. Silva, J. M. C. Sousa, T. A. Runkler, and J. M. G. Sá da Costa, "Distributed supply chain management using ant colony optimization," *European Journal of Operational Research*, vol. 199, no. 2, pp. 349–358, 2009.
- [65] H. S. Wang, "A two-phase ant colony algorithm for multi-echelon defective supply chain network design," *European Journal of Operational Research*, vol. 192, no. 1, pp. 243–252, 2009.
- [66] L. A. Moncayo-Martínez and D. Z. Zhang, "Multi-objective ant colony optimisation: a meta-heuristic approach to supply chain design," *International Journal of Production Economics*, vol. 131, no. 1, pp. 407–420, 2011.
- [67] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.
- [68] J. Izquierdo, R. Minciardi, I. Montalvo, M. Robba, and M. Tavera, "Particle Swarm Optimization for the biomass supply chain strategic planning," in *Proceedings of the International Congress on Environmental Modelling and Software*, pp. 1272–1280, 2008.
- [69] G. W. Hu, "Generalized genetic particle swarm optimization for supply chain optimization," *Journal of Computer Applications*, vol. 28, no. 11, pp. 2840–2843, 2008.
- [70] R. Kadavevaramath, K. Mohanasundaram, K. Ramesh Kumar, and B. Chandrashekhar, "Production and distribution scheduling of Supply Chain structure using intelligent Particle Swarm Optimisation algorithm," *International Journal of Intelligent Systems Technologies and Applications*, vol. 6, no. 3–4, pp. 249–268, 2009.
- [71] M. Bachlaus, M. K. Pandey, C. M. Shankar, and M. K. Tiwari, "Designing an integrated multi-echelon agile supply chain network: a Hybrid Taguchi-Particle swarm optimization approach," *International Journal of Flexible Manufacturing System*, vol. 19, no. 6, pp. 486–515, 2007.
- [72] A. A. Taleizadeh, S. T. A. Niaki, N. Shafii, R. G. Meibodi, and A. Jabbarzadeh, "A particle swarm optimization approach for constraint joint single buyer-single vendor inventory problem with changeable lead time and (r,Q) policy in supply chain," *International Journal of Advanced Manufacturing Technology*, vol. 51, no. 9–12, pp. 1209–1223, 2010.
- [73] A. K. Sinha, H. K. Aditya, M. K. Tiwari, and F. T. S. Chan, "Agent oriented petroleum supply chain coordination: co-evolutionary particle swarm optimization based approach," *Expert Systems with Applications*, vol. 38, no. 4, pp. 6132–6145, 2011.
- [74] X. J. Wu, H. Zhou, and C. H. Liang, "Collaborative optimization of multi-echelon supply chain based on co-evolutionary particle swarm optimization," *Computer Integrated Manufacturing Systems*, vol. 16, no. 1, pp. 127–132, 2010.
- [75] G. Taguchi, S. Chowdhary, and S. Taguchi, *Robust engineering*, McGraw-Hill, New York, NY, USA, 2000.
- [76] C. Soares, G. Dozier, E. Lodree, J. Phillips, K. Nobles, and W. P. Yong, "Optimization of the multiple retailer supply chain management problem," in *Proceedings of the 46th Annual Southeast Regional Conference on XX ACM-SE*, pp. 490–495, Auburn, Ala, USA, March 2008.
- [77] X. S. Yang, "Engineering optimizations via nature-inspired virtual bee algorithms," in *Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach*, vol. 3562 of *Lecture Notes in Computer Science*, pp. 317–323, Springer, Berlin, Germany, 2005.
- [78] D. T. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim, and M. Zaidi, "The bees algorithm," Tech. Rep., Manufacturing Engineering Centre, Cardiff University, 2005.
- [79] D. Karaboga, "An idea based on honeybee swarm for numerical optimization," Tech. Rep. TR06, Erciyes University, 2005.
- [80] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *Journal of Global Optimization*, vol. 39, no. 3, pp. 459–471, 2007.
- [81] V. V. Kumar, F. T. S. Chan, N. Mishra, and V. Kumar, "Environmental integrated closed loop logistics model: an artificial bee colony approach," in *Proceedings of the 8th International Conference on Supply Chain Management and Information Systems: Logistics Systems and Engineering (SCMIS '10)*, Hong Kong, China, 2010.
- [82] A. Pal, F. T. S. Chan, B. Mahanty, and M. K. Tiwari, "Aggregate procurement, production, and shipment planning decision problem for a three-echelon supply chain using swarm-based heuristics," *International Journal of Production Research*, vol. 49, no. 10, pp. 2873–2905, 2011.
- [83] S. Banerjee, G. S. Dangayach, S. K. Mukherjee, and P. K. Mohanti, "Modelling process and supply chain scheduling using hybrid meta-heuristics," *Studies in Computational Intelligence*, vol. 128, pp. 277–300, 2008.
- [84] J.-Q. Li, Q.-K. Pan, and K.-Z. Gao, "Pareto-based discrete artificial bee colony algorithm for multi-objective flexible job shop scheduling problems," *International Journal of Advanced Manufacturing Technology*, vol. 55, no. 9–12, pp. 1159–1169, 2011.
- [85] A. Prakash and S. G. Deshmukh, "A multi-criteria customer allocation problem in supply chain environment: an artificial immune system with fuzzy logic controller based approach," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3199–3208, 2011.
- [86] M. Shukla, N. Shukla, M. K. Tiwari, and F. T. S. Chan, "Integrated model for the batch sequencing problem in a multi-stage supply chain: an artificial immune system based approach," *International Journal of Production Research*, vol. 47, no. 4, pp. 1015–1037, 2009.
- [87] M. Hajiaghahi-Keshteli, "The allocation of customers to potential distribution centers in supply chain networks: GA and AIA approaches," *Applied Soft Computing Journal*, vol. 11, no. 2, pp. 2069–2078, 2011.
- [88] D. Srinivasan and T. H. Seow, "Particle swarm inspired evolutionary algorithm (PS-EA) for multi-objective optimization problem," in *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 2292–2297, Canberra, Australia, 2003.
- [89] R. Poli, C. Di Chio, and W. B. Langdon, "Exploring extended particle swarms: a genetic programming approach," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '05)*, pp. 169–176, Washington, DC, USA, June 2005.
- [90] W. Fu, M. Johnston, and M. Zhang, "Hybrid particle swarm optimisation algorithms based on differential evolution and local search," in *Proceedings of the Australasian Conference on Artificial Intelligence*, vol. 6464 of *Lecture Notes in Artificial Intelligence*, pp. 313–322, 2010.
- [91] Y. J. Wang, "Improving particle swarm optimization performance with local search for high-dimensional function optimization," *Optimization Methods and Software*, vol. 25, no. 5, pp. 781–795, 2010.
- [92] S. Meeran and M. S. Morshed, "A hybrid genetic tabu search algorithm for solving job shop scheduling problems: a case study," *Journal of Intelligent Manufacturing*. In press.

- [93] X. Wang, X. Z. Gao, and S. J. Ovaska, "A hybrid artificial immune optimization method," *International Journal of Computational Intelligence Systems*, vol. 2, no. 3, pp. 249–256, 2009.
- [94] D. Guo, J. Wang, J. Huang, R. Han, and M. Song, "Chaotic-NSGA-II: an effective algorithm to solve multi-objective optimization problems," *Proceedings of the International Conference on Intelligent Computing and Integrated Systems (ICISS '10)*, pp. 20–23, 2010.
- [95] R. Caponetto, L. Fortuna, S. Fazzino, and M. G. Xibilia, "Chaotic sequences to improve the performance of evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 3, pp. 289–304, 2003.
- [96] G. Zhang, "Quantum-inspired evolutionary algorithms: a survey and empirical study," *Journal of Heuristics*, vol. 17, no. 3, pp. 303–351, 2010.
- [97] E. G. Bakhoun and C. Toma, "Specific mathematical aspects of dynamics generated by coherence functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.
- [98] E. G. Bakhoun and C. Toma, "Dynamical aspects of macroscopic and quantum transitions due to coherence function and time series events," *Mathematical Problems in Engineering*, vol. 2010, Article ID 428903, 13 pages, 2010.
- [99] F. Tao, L. Zhang, Z. H. Zhang, and A. Y. C. Nee, "A quantum multi-agent evolutionary algorithm for selection of partners in a virtual enterprise," *Manufacturing Technology*, vol. 59, no. 1, pp. 485–488, 2010.

Review Article

Motion Analysis of Live Objects by Super-Resolution Fluorescence Microscopy

Chunyan Yao,^{1,2} Jianwei Zhang,² Guang Wu,^{3,4} and Houxiang Zhang⁵

¹Key Laboratory of E&M, Zhejiang University of Technology, Ministry of Education and Zhejiang Province, Hangzhou 310014, China

²Department of Informatics, University of Hamburg, 22527 Hamburg, Germany

³Guangxi Academy of Sciences, 98 Daling Road, Nanning 530007, China

⁴DreamSciTech Consulting, Shenzhen 518054, China

⁵Department of Technology and Nautical Sciences, Aalesund University College, 26025 Aalesund, Norway

Correspondence should be addressed to Chunyan Yao, yao@informatik.uni-hamburg.de

Received 1 September 2011; Accepted 26 September 2011

Academic Editor: Carlo Cattani

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

Motion analysis plays an important role in studying activities or behaviors of live objects in medicine, biotechnology, chemistry, physics, spectroscopy, nanotechnology, enzymology, and biological engineering. This paper briefly reviews the developments in this area mostly in the recent three years, especially for cellular analysis in fluorescence microscopy. The topic has received much attention with the increasing demands in biomedical applications. The tasks of motion analysis include detection and tracking of objects, as well as analysis of motion behavior, living activity, events, motion statistics, and so forth. In the last decades, hundreds of papers have been published in this research topic. They cover a wide area, such as investigation of cell, cancer, virus, sperm, microbe, karyogram, and so forth. These contributions are summarized in this review. Developed methods and practical examples are also introduced. The review is useful to people in the related field for easy referral of the state of the art.

1. Background

Motion analysis for microscopic live objects has become a new approach to understanding complex living information of the microworld [1]. Fluorescence microscopy imaging has rapidly evolved in these years and provided new means for studying microbial behavior, cell motion, or intracellular processes *in vivo*. The super-resolution techniques allow the observation of many biological structures not resolvable in conventional fluorescence microscopy. New advances in these techniques now give them the ability to image three-dimensional (3D) structures, measure interactions, and record dynamic processes in living cells [2–5]. However, since such studies often generate a large volume of noisy image data that cannot be analyzed efficiently and reliably by human observation, it is a critical issue to develop computing methods for automatic motion analysis. Many tracking techniques exist in the computing science but often fail to yield satisfactory results in the case of high object densities, high noise levels, and complex motion patterns [6].

Cellular life can be described as a dynamic equilibrium of a highly complex network of interacting molecules [7]. It is not sufficient to only know the identity of the participants in a cellular process, but questions such as where, when, and for how long also have to be addressed to understand the mechanism being investigated.

Recent improvements in methods of single-particle fluorescence tracking have permitted detailed studies of molecular motion on the nanometer scale due to its high selectivity, sensitivity, simplicity, and fastness [8]. The high microscopy sensitivity resolves the signal of a single fluorescence-labeled bimolecular within a living cell. A time resolution of milliseconds for imaging weakly fluorescent cellular structures like small organelles, vesicles, or even single molecules is already available [9]. This introduces tools to a burgeoning field of nanotechnology for analysis of microtubules, DNAs, proteins, and other biochemical particles.

Motion analysis for microscopic live objects is a key issue in this field. One aim is the recognition of selected

targets in the image and tracking them in time [10]. Single-particle tracking is powerful in measuring particle motion from video-microscopy image sequences [11, 12]. A system actually needs to observe the type of cells and their movement in long image sequences [13].

In the survey, advances in motion analysis for microscopic live objects are briefly reviewed. Overall, significant progress has been made in several issues, as well as being integrated with other methods for biomedical applications.

The scope of this paper is restricted to motion analysis for microscopic live objects by image processing [14, 15], especially in the field of cellular research. Although this topic has attracted researchers as early as since the 1980s, this survey concentrates on the contributions of the latest three years when the single-molecule fluorescence microscopy or super-resolution fluorescence imaging became practically available. It includes only some representative or important works from recent years.

The current paper has four more sections. Section 2 introduces some application examples. Section 3 briefly shows the principle of particle filters. Section 4 lists the relevant research problems and typical methods. Section 5 includes a conclusion and a discussion of future trends.

2. Application Examples

2.1. Virion Study. Optical studies have revealed that virions move laterally on the plasma membrane, but the complexity of the cellular environment has prevented access to the molecular dynamics of early virus-host couplings, which are important for cell infection [16, 17]. Kukura et al. discussed a technique for studying virus-membrane interactions and for resolving nanoscopic dynamics of individual biological nanoobjects [18]. Helmuth et al. uses a shape reconstruction method that leads to better discrimination between different endosomal virus entry pathways and to more robust, accurate, and self-consistent quantification of endosome shape features [19].

For 3D tracking of GLUT4 vesicles, Wu et al. present an algorithm, where mobile granules are segmented from a time-lapse image stack, and Kalman filter is used to estimate the granules for reducing searching range of reliable tracking [20]. Schelhaas et al. studied the lateral mobility of individual, incoming human papillomavirus type 16 pseudoviruses bound to live HeLa cells by single-particle tracking [21, 22].

In [23], eight tracking approaches are evaluated for testing real microscopy images of HIV-1 particles. The probabilistic approaches based on independent particle filters are found to be superior to the deterministic schemes as well as to the approaches based on a mixture of particle filters.

2.2. Cellular Analysis. A tracking algorithm is tested in [10] on solid-lipid nanoparticles diffusing within cells and on lymphocytes diffusing in lymphonodes. It appears useful for the cellular and in vivo image processing where little a priori assumption on the type, the extent and the variability of

particle motions, can be done. Another example of single-particle tracking is proposed in [13] with application to platelet adhesion under flow.

2.3. Subcellular Analysis. Determination of the subcellular localization and dynamics is an important step towards the understanding of multimolecular complexes in a cellular context [24]. Live imaging of subcellular structures is essential for the understanding of cellular processes. Algorithms have been proposed in the community for detecting fluorescently labeled subcellular objects in microscope images [25]. However, reconstruction of subcellular structures from images remains a major challenge in the field [19].

Transport of intracellular organelles along the microtubule cytoskeleton is in a bidirectional manner [26]. The study of protein dynamics is essential for the understanding the multi-molecular complexes at subcellular levels [27]. For understanding of microtubules and their associated proteins, dynamic interactions are routinely observed in vitro on the level of single molecules, mainly using a geometry in which labeled motors move on surface-immobilized microtubules [28]. Machan and Hof discussed several key questions of lateral mobility investigation in planar lipid membranes, including the influence of membrane and aqueous phase composition, choice of a fluorescent tracer molecule, frictional coupling between the two membrane leaflets and between membrane and solid support [29]. A simulation and estimation framework is proposed in image-processing field for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery [30].

2.4. Activity in Nucleus. The nucleus is a well-organized and highly compartmentalized organelle, and this organization is intimately related to nuclear function [31]. Dange et al. discussed the potential of tracking single RNAs and proteins in the nucleus [7]. The dynamics, localization, and interaction rates are vital to the understanding of cellular life. They provide a review of the HIV life cycle, which provides an opportunity to study mechanisms deeply integrated within the structure of the nucleus.

Gene delivery helps treatment of many diseases. Underlying that the rational design of gene-delivery vectors can discover the individual steps of the gene-delivery pathway. With fluorescence microscopy, it is possible to isolate individual steps along the gene-delivery pathway to characterize the mechanisms of cellular binding, cellular internalization, and nuclear entry [32].

3D imaging of bacterial protein distribution and neuron dendritic morphology is discussed in [33]. Super-resolution method is proposed in [34] for orientation estimation and localization of fluorescent dipoles using 3D steerable filters.

2.5. Molecular Mechanism. Dynamic properties of proteins have crucial roles in understanding protein function and molecular mechanism within cells [35]. Researchers often need to accurately determine the motion of single-molecule trajectories [36]. It is also interesting to study the behavior of motor proteins and associated organelle transport within

a cell [11, 37]. Molecular motors such as kinesin, myosin, and F-1-ATPase are responsible for many important cellular processes. These motor proteins exhibit nanometer-scale, stepwise movements on millisecond timescales. Methods to measure these small and fast movements with high spatial and temporal resolution require relatively complicated systems [38]. Colocalization of two or more molecules is an essential aspect of many biological molecular processes and single-molecule technologies for investigating these processes in live cells [39]. An imaging method is developed in [40] to quantitatively detect the colocalization of two species of individual molecules.

3. Particle Tracking

3.1. Particle Filtering. Probabilistic tracking methods have recently shown several advantages with better integration of spatial and temporal information, and the possibility to more effectively incorporate prior knowledge. Smal et al. propose a fully automated particle filtering algorithm for the tracking of many subresolution objects in fluorescence microscopy image sequences [6]. It involves a track management procedure and allows the use of multiple dynamics models.

Villa et al. developed a particle-tracking algorithm optimized for low signal noised images with minimum requirements on the target size and without a priori knowledge of the motion type. The particle tracking is performed by building, from a stack of accumulative difference images, a single 2D image in which the motion of the whole set of the particles is coded in time [10].

Many single-particle tracking algorithms deliver subpixel accurate measurements with noisy data corresponding to sub-10-nm resolution. Image-correlation techniques have been shown to be the most accurate method of tracking extended objects. Saunter proposes a method for experimentally determining the accuracy of image-correlation-based tracking and demonstrates the possibility of making measurements accurate to 5 nm when working with extended objects within live cells [11].

3.2. Multiple Model Filtering. A multiple model tracking method is proposed in [41] with multidimensional assignment. They combine an interacting multiple model filter, multidimensional assignment, particle occlusion handling, and merge-split event detection together. The advantage of a multidimensional assignment is that both spatial and temporal information can be used by using several later frames as reference. The multiple model filter, which is used to maintain and predict the state of each track, contains several models which correspond to different types of biologically realistic movements.

Godínez et al. developed deterministic and probabilistic approaches for multiple virus tracking in multichannel fluorescence microscopy images [23]. The probabilistic approaches are based on a mixture of particle filters and independent particle filters.

Genovesio et al. propose a method to detect and track multiple moving biological spotlike particles showing

different kinds of dynamics [42]. It can extract and analyze information such as number, position, speed, movement, and diffusion phases of particles. After a detection stage performed by a 3D undecimated wavelet transform, prediction of spots' future states is accomplished with an interacting multiple model algorithm which includes several models corresponding to different biologically realistic movement types. Then the filters are updated to compute final estimations.

For tracking problem of several hundreds of objects, a framework is provided in [43] with general information about vesicle transport, that is, traffic flows between origin and destination regions. Traffic estimation is accomplished by adapting the advances in network tomography. A method is proposed in [44] to detect and track multiple moving biological spotlike particles showing different kinds of dynamics in image sequences.

3.3. 3D Nanoscopy. Various biophysical studies require high spatial and temporal resolution in vitro and also in vivo [38]. It remains a challenge to precisely localize single molecules in 3D [45, 46]. Tang et al. applies near-isotropic 3D optical nanoscopy with photon-limited chromophores to 3D imaging of bacterial protein distribution and neuron dendritic morphology with subdiffraction resolution [3, 33].

Huang et al. demonstrated 3D stochastic optical reconstruction microscopy by using optical astigmatism to determine both axial and lateral positions of individual fluorophores with nanometer accuracy [2]. The construction of a 3D image can achieve a resolution of 20 to 30 nanometers in the lateral dimensions and 50 to 60 nanometers in the axial dimension. This allows to resolve the 3D morphology of nanoscopic cellular structures.

Yu et al. argued that the bright fluorescence can yield a theoretical particle tracking uncertainty of less than 1 nm. A lateral tracking uncertainty of 1-2 nm is determined from analysis of trajectories of fixed and freely diffusing particles. Axial position information for 3D particle tracking is obtained by defocused imaging [47].

4. Research Issues

4.1. Fluorescence Microscopy. Most biological molecules are less than 5–10 nm in diameter, and getting molecular details requires imaging at this scale. The basic research in cell biology and in medical sciences is mainly based on confocal fluorescence [10]. Nowadays, a great deal of attention in biomedical and pharmaceutical technology is going to the development of nanoscopic particles to efficiently deliver nucleic acids to target cells. Despite the great potential of nucleic acids for treatment of various diseases, progress in the field is fairly slow [48]. The resolution of conventional optical microscopy is constrained to about 200–500 nm due to the diffraction limit, but the recent developed super-resolution fluorescence imaging, which is based on single-molecule localization and image reconstruction, offers a comparatively simple way to achieve a substantially improved optical resolution down to similar to 20–50 nm in the image plane [49–51]. The recent stochastic optical reconstruction

microscopy makes use of single-molecule imaging methods and photo-switchable fluorescent probes to temporally separate the spatially overlapping images of individual molecules. An image is acquired over a number of imaging cycles. This allows the position to be determined with nanometer accuracy [51]. Single-molecule fluorescence microscopy has become one of the most popular methods in the single-molecule toolbox [52]. In [53], a 26-ms time resolution and a spatial accuracy of 5 nm in each dimension are achieved. The resulting high-resolution trajectories reveal not only heterogeneity among vesicles but also heterogeneity within single-vesicle trajectories. Fluorescence imaging with one-nanometer accuracy and single-molecule high-resolution colocalization are used to monitor the diffusive behavior of synthetic molecular walkers at the single-molecule level. Michelotti discussed the imaging methods and experimental challenges of very low velocities (e.g., 3 nm/min) of nanowalkers [8]. The state of the art in single-molecule tools including fluorescence spectroscopy, tethered particle microscopy, optical and magnetic tweezers, and atomic force microscopy can be found in the survey by Walter et al. [54].

4.2. Detection. Determination of the position of a cell, termed localization, is of paramount importance in achieving reliable and robust motion analysis. Achieving high-level descriptors such as dynamics and activities is possible if the position is known and accurately tracked. Aguet et al. introduce a method for the joint estimation of position and orientation of dipoles, based on the representation of a physically realistic image formation model as a 3D steerable filter. They establish theoretical, localization-based resolution limits on estimation accuracy and experimentally show that the position accuracy is 5 nm and the orientation accuracy is 2 degrees [34].

Smal et al. evaluate the performance of the most frequently used detection methods for quantitative comparison [55]. Seven unsupervised and two supervised methods are involved. The experiments are carried out on synthetic images of three different types, for which the ground truth was available, as well as on real image data sets acquired for two different biological studies. The results suggest that for very high noise images the supervised learning methods perform best overall.

4.3. Intracellular Process. Ensemble measurements are not sufficient to describe individual steps of molecular mobility, spatial-temporal resolution, kinetic parameters, and geographical mapping. It is important to find where individual steps exactly occur for better understanding of the living cell. The nucleus is with many highly complex multiorder processes, such as replication, transcription, splicing, and so forth and provides a complicated, heterogeneous landscape. Single-molecule tracking has become more and more attractive [7]. Jung et al. reported the work on diffusion of oriented single molecules with switchable mobility in networks of long unidimensional nanochannels [56].

Using dynamic techniques to track, manipulate, and probe motor proteins is crucial in providing new insights

[1]. To understand the regulation of intracellular transport through quantitative analysis of the motion of organelles in a controlled environment, Soppina et al. present a simple and reliable method that uses avidin-coated magnetic beads to prepare microtubules labeled at the minus end. It demonstrated video-rate high-resolution imaging of single cellular organelles moving along plus and minus directions on labeled microtubules [26].

4.4. Segmentation and Count. Ruusuuvuori carried out a comparative study including eleven-spot detection or segmentation algorithms from various application fields [25]. The experimentally derived images permit a comparison of method performance in realistic situations where the number of objects varies within image set. The study finds major differences in the performance of different algorithms, in terms of object counts and segmentation accuracies. An automatic segmentation and tracking method is designed in [44] to enable quantitative analyses of cellular shape and motion from 4D microscopy data.

A method in [13] includes functions of automatic segmentation methodology which removes operator bias, platelet recognition across the series of images based on a probability density function. It can be integrated to analyze the platelet trajectories to obtain relevant information, such as deposition and removal rates, displacement distributions, pause times, and rolling velocities.

4.5. Movement and Tracking. It is often required to localize particles in a live cell to a certain accuracy to study their localization-related functions [57, 58]. Particle tracking has seen numerous applications in live-cell imaging studies of subcellular dynamics. Establishing correspondence between particles in a sequence of frames with high particle density, particles merging and splitting, particles entering and exiting the frame, temporary particle disappearance, and an ill-performing detection algorithm is the most challenging task [41]. Smal et al. propose a completely automatic tracker based on a Bayesian probabilistic framework. It better exploits spatiotemporal information and prior knowledge than common approaches, which yields more robust tracking [59].

Sbalzarini and Koumoutsakos present a 2D feature tracking algorithm for the automated detection and quantitative analysis of particle trajectories as recorded by video imaging in cell biology [60]. The tracking requires no a priori modeling of the motion, it is self-initializing, it discriminates spurious detections, and it can handle temporary occlusion as well as particle appearance and disappearance from the image.

To extract the maximum information from a sequence of fluorescence images, Yoon et al. describe a Bayesian-based inference approach, based on a transdimensional sequential Monte Carlo method that utilizes spatiotemporal information. The method allows accurate tracking of molecules over long trajectories even with low signal/noise ratio and in the presence of fluorescence blinking and photobleaching [61].

The movement trajectories in fluorescence video microscopy can be computationally analyzed in terms of

diffusion rate and mode of motion [21]. The trajectories play a role in the analysis of living-cell dynamics. Boulanger et al. develop a general simulation framework to produce image sequences showing small moving spots in interaction and corresponding to intracellular dynamics and trafficking in biology [30]. Wieser and Schutz describe strategies how to make use of single-molecule trajectories for deducing information about nanoscopic structures in a live cell context, for example, elucidating the plasma membrane organization [9].

For simultaneously determining multiple trajectories of single molecules from sequential fluorescence images, Claytor et al. developed a procedure for accurately monitoring surface motion under ambient conditions [36]. The tracking algorithm is computationally nondemanding and does not assume a model for molecular motion.

4.6. Dynamics. Investigation of lipid lateral mobility in biological membranes and their artificial models provides information on membrane dynamics and structure where methods based on optical microscopy are convenient for such investigations [29].

The motion of the tagged locus is observed and analyzed to extract quantitative information regarding its dynamics. Levi and Gratton reviewed recent advances in chromatin dynamics in interphase. They introduced the basis of particle-tracking methods and trajectory analysis and summarized what has been learnt by using this technology in the context of chromatin dynamics [31]. Nitzsche reviewed the recent methods related to gliding motility assays in conjunction with 3D-nanometry [28]. They provide practical advice on how to set up gliding assays, acquire high-precision data from microtubules and attached quantum dots, and analyze data by 3D-nanometer tracking. Nanoscale tracking of single particles in fixed cells was demonstrated in [47], where a range of complex behaviors, possibly due to binding/unbinding dynamics, is observed.

For direct visualization of the protein dynamics, Wang et al. combined total internal reflection fluorescence microscopy with oblique illumination fluorescence microscopy to observe the movement of membrane-anchored green fluorescence proteins in living cells [35].

Kukura et al. present a colocalization methodology that combines scattering interferometry and single-molecule fluorescence microscopy to visualize both position and orientation of single quantum dot-labeled Simian virus 40 particles [18]. Using nanometer spatial with 8 ms temporal resolution, they observed sliding and tumbling motions during rapid lateral diffusion on supported lipid bilayers and repeated back and forth rocking between nanoscopic regions separated by 9 nm.

4.7. Events and Behaviors. Detection of meaningful events in spatiotemporal fluorescence image sequences is certainly important in cellular analysis, for example, membrane transport and gene-delivery process [32]. Pécot et al. propose an original patch-based Markov modeling to detect spatial irregularities in fluorescence images with low false

alarm rates [27]. Toprak and Selvin described some of the most commonly used fluorescence imaging tools to measure nanoscale movements and the rotational dynamics of biomolecules [62].

4.8. Reconstruction. Shape reconstruction of 2D/3D sub-cellular structures from live cell helps us to understand the whole object. Helmuth et al. presented model-based algorithm to reconstruct outlines of subcellular structures using a subpixel representation [19, 63]. From 2D images, 3D tracking can also be realized by utilizing the exponential decay of the fluorescence intensity excited by the evanescent field [64].

5. Conclusion and Future Directions

This paper summarizes recent developments in motion analysis for microscopic live objects in cellular research. Typical contributions are addressed for localization, tracking, count, estimation, reconstruction, modeling, cell analysis, and so forth. Representatives are listed for people to have a general overview of state of the art. A number of methods are introduced on motion analysis for microscopic live objects. Most of the cited works are published reports in the last three years.

Although motion analysis for microscopic live objects has been developed in many applications as an important approach to observation and diagnosis, many problems still exist in its development for biomedical engineering. Researchers are exerting efforts not only in simple localization, but also in improving the future methods in other aspects.

- (1) Dealing with noisy data: in fluorescence microscopy, the blurred noisy images acquired are complex to analyze. Hidden Markov models facilitate extraction of the sequence of hidden states from noisy data through construction of probabilistic models. Since constraints resulting from short data sets and poisson-distributed photons in like fluorescence can limit good statistics, additional information criteria such as peak localization error and chi-square probabilities can incorporate theoretical constraints in principle [65].
- (2) Accurate detection: in live-cell fluorescence microscopy imaging, the biological images generally need to detect many subresolution objects. Indeed, complex interactions between a large number of small moving particles in a complex scene cannot be easily modeled, limiting the performance of object detection and tracking algorithms [30]. Object detection and tracking often perform poorly in the case of noisy image data [59]. In live-cell imaging by fluorescence microscopy, the signal-to-noise ratio can be extremely low, making automated spot detection a very challenging task [55].
- (3) Reliable tracking: reliability is a great concern in practical applications. Motion analysis relies on many conditions and parameters. Related techniques rely

on existing noise statistics, initial positions, and sufficiently good approximation. Quantitative analysis of dynamic processes in living cells requires tracking of hundreds of bright spots in noisy image sequences [59]. Factors, such as low signal/noise ratio, unknown number of particles, and fluorophore blinking and photobleaching, make accurate tracking over long trajectories very difficult [61]. Existing tracking methods rarely work when many small and poorly distinguishable objects interact. Pécot et al. proposes that another way of tracking that consists in determining the full trajectories of all the objects, can be more relevant [24]. Techniques like background subtraction and Kalman filter are also helpful for reliable tracking [20]. Rogers et al. designed an algorithm to accurately track the motion of low-contrast particles against a background with large variations in light levels. The method is based on a polynomial fit of the intensity around each feature point and is especially suitable for tracking endogenous particles in the cell, imaged with bright field, phase contrast or fluorescence optical microscopy. It can also simultaneously track particles of all different sizes and shapes [66].

- (4) Information analysis: advanced methods in computer vision and image processing are helpful to extract interesting information. Data analysis requires deep collaboration between biologists and computer scientists. It is a challenge to transform the vast amounts of unstructured data into quantitative information for the discovery of cellular behaviors and the rigorous testing of mechanistic hypotheses [67, 68].

Acknowledgments

This work was supported by National Natural Science Foundation of China (51075367), Natural Science Foundation of Zhejiang Province (Y1090931, 2010R10006, 2010C33095), and the Doctoral Program of Higher Education (20070337002).

References

- [1] C. Veigel and C. F. Schmidt, "Moving into the cell: single-molecule studies of molecular motors in complex environments," *Nature Reviews Molecular Cell Biology*, vol. 12, no. 3, pp. 163–176, 2011.
- [2] B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy," *Science*, vol. 319, no. 5864, pp. 810–813, 2008.
- [3] B. Huang, M. Bates, and X. Zhuang, "Super-resolution fluorescence microscopy," *Annual Review of Biochemistry*, vol. 78, pp. 993–1016, 2009.
- [4] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, "Multicolor super-resolution imaging with photo-switchable fluorescent probes," *Science*, vol. 317, no. 5845, pp. 1749–1753, 2007.
- [5] J. Kajstura, M. Rota, S. R. Hall et al., "Evidence for human lung stem cells," *New England Journal of Medicine*, vol. 364, no. 19, pp. 1795–1806, 2011.
- [6] I. Smal, E. Meijering, K. Draegestein et al., "Multiple object tracking in molecular bioimaging by Rao-Blackwellized marginal particle filtering," *Medical Image Analysis*, vol. 12, no. 6, pp. 764–777, 2008.
- [7] T. Dange, A. Joseph, and D. Grünwald, "A perspective of the dynamic structure of the nucleus explored at the single-molecule level," *Chromosome Research*, vol. 19, no. 1, pp. 117–129, 2011.
- [8] N. Michelotti, C. de Silva, A. E. Johnson-Buck, A. J. Manzo, and N. G. Walter, "A bird's eye view. Tracking slow nanometer-scale movements of single molecular nano-assemblies," *Methods in Enzymology*, vol. 475, no. C, pp. 121–148, 2010.
- [9] S. Wieser and G. J. Schütz, "Tracking single molecules in the live cell plasma membrane-Do's and Don'ts," *Methods*, vol. 46, no. 2, pp. 131–140, 2008.
- [10] C. E. Villa, M. Caccia, L. Sironi et al., "Accumulative difference image protocol for particle tracking in fluorescence microscopy tested in mouse lymphonodes," *PLoS ONE*, vol. 5, no. 8, Article ID e12216, 2010.
- [11] C. D. Saunter, "Quantifying subpixel accuracy: an experimental method for measuring accuracy in image-correlation-based, single-particle tracking," *Biophysical Journal*, vol. 98, no. 8, pp. 1566–1570, 2010.
- [12] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha, "Advances in single-molecule fluorescence methods for molecular biology," *Annual Review of Biochemistry*, vol. 77, pp. 51–76, 2008.
- [13] M. MacHin, A. Santomaso, M. Mazzucato et al., "Single particle tracking across sequences of microscopical images: application to platelet adhesion under flow," *Annals of Biomedical Engineering*, vol. 34, no. 5, pp. 833–846, 2006.
- [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] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.
- [16] D. N. Fredricks, T. L. Fiedler, and J. M. Marrazzo, "Molecular identification of bacteria associated with bacterial vaginosis," *New England Journal of Medicine*, vol. 353, no. 18, pp. 1899–1911, 2005.
- [17] S. A. Fischer, M. B. Graham, M. J. Kuehnert et al., "Transmission of lymphocytic choriomeningitis virus by organ transplantation," *New England Journal of Medicine*, vol. 354, no. 21, pp. 2235–2249, 2006.
- [18] P. Kukura, H. Ewers, C. Müller, A. Renn, A. Helenius, and V. Sandoghdar, "High-speed nanoscopic tracking of the position and orientation of a single virus," *Nature Methods*, vol. 6, no. 12, pp. 923–927, 2009.
- [19] J. A. Helmuth, C. J. Burckhardt, U. F. Greber, and I. F. Sbalzarini, "Shape reconstruction of subcellular structures from live cell fluorescence microscopy images," *Journal of Structural Biology*, vol. 167, no. 1, pp. 1–10, 2009.
- [20] X. P. Wu, J. Y. Li, Y. K. Xu, K. D. Xu, and X. X. Zheng, "Three-dimensional tracking of GLUT4 vesicles in TIRF microscopy," *Journal of Zhejiang University: Science A*, vol. 9, no. 2, pp. 232–240, 2008.

- [21] M. Schelhaas, H. Ewers, M. L. Rajamäki, P. M. Day, J. T. Schiller, and A. Helenius, “Human papillomavirus type 16 entry: retrograde cell surface transport along actin-rich protrusions,” *PLoS Pathogens*, vol. 4, no. 9, Article ID e1000148, 2008.
- [22] S. Y. Chen and Y. F. Li, “Determination of stripe edge blurring for depth sensing,” *IEEE Sensors Journal*, vol. 11, no. 2, Article ID 5585653, pp. 389–390, 2011.
- [23] W. J. Godinez, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, “Deterministic and probabilistic approaches for tracking virus particles in time-lapse fluorescence microscopy image sequences,” *Medical Image Analysis*, vol. 13, no. 2, pp. 325–342, 2009.
- [24] T. Pécot, C. Kervrann, and P. Bouthemy, “Network tomography-based tracking for intracellular traffic analysis in fluorescence microscopy imaging,” in *Proceedings of the 1st International Conference on Bio-inspired Systems and Signal Processing (Biosignals ’08)*, vol. 1, pp. 154–161, January 2008.
- [25] P. Ruusuvaari, T. Äijö, S. Chowdhury et al., “Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images,” *BMC Bioinformatics*, vol. 11, article 248, 2010.
- [26] V. Soppina, A. K. Rai, and R. Mallik, “Simple non-fluorescent polarity labeling of microtubules for molecular motor assays,” *BioTechniques*, vol. 46, no. 7, pp. 543–549, 2009.
- [27] T. Pécot, C. Kervrann, S. Bardin, B. Goud, and J. Salamero, “Patch-based Markov models for event detection in fluorescence bioimaging,” *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI ’08)*, vol. 11, no. 2, pp. 95–103, 2008.
- [28] B. Nitzsche, V. Bormuth, C. Bräuer et al., “Studying kinesin motors by optical 3D-nanometry in gliding motility assays,” *Methods in Cell Biology*, vol. 95, no. C, pp. 247–271, 2010.
- [29] R. Macháň and M. Hof, “Lipid diffusion in planar membranes investigated by fluorescence correlation spectroscopy,” *Biochimica et Biophysica Acta—Biomembranes*, vol. 1798, no. 7, pp. 1377–1391, 2010.
- [30] J. Boulanger, C. Kervrann, and P. Bouthemy, “A simulation and estimation framework for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery,” *Medical Image Analysis*, vol. 13, no. 1, pp. 132–142, 2009.
- [31] V. Levi and E. Gratton, “Chromatin dynamics during interphase explored by single-particle tracking,” *Chromosome Research*, vol. 16, no. 3, pp. 439–449, 2008.
- [32] C. K. Payne, “Imaging gene delivery with fluorescence microscopy,” *Nanomedicine*, vol. 2, no. 6, pp. 847–860, 2007.
- [33] J. Tang, J. Akerboom, A. Vaziri, L. L. Looger, and C. V. Shank, “Near-isotropic 3D optical nanoscopy with photon-limited chromophores,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 22, pp. 10068–10073, 2010.
- [34] F. Aguet, S. Geissbuhler, I. Marki, T. Lasser, and M. Unser, “Super-resolution orientation estimation and localization of fluorescent dipoles using 3-D steerable filters,” *Optics Express*, vol. 17, no. 8, pp. 6829–6848, 2009.
- [35] C. Wang, G. Fu, J. Wang, G. Wang, Y. Cheng, and Z. Z. Xu, “Direct visualization of the dynamics of membrane-anchor proteins in living cells,” *Journal of Microscopy*, vol. 229, no. 1, pp. 67–77, 2008.
- [36] K. Claytor, S. Khatua, J. M. Guerrero, A. Tcherniak, J. M. Tour, and S. Link, “Accurately determining single molecule trajectories of molecular motion on surfaces,” *Journal of Chemical Physics*, vol. 130, no. 16, Article ID 164710, 2009.
- [37] Y. Zhao, I. Gaidarov, and J. H. Keen, “Phosphoinositide 3-kinase C2 links clathrin to microtubule-dependent movement,” *Journal of Biological Chemistry*, vol. 282, no. 2, pp. 1249–1256, 2007.
- [38] H. Ueno, S. Nishikawa, R. Iino et al., “Simple dark-field microscopy with nanometer spatial precision and microsecond temporal resolution,” *Biophysical Journal*, vol. 98, no. 9, pp. 2014–2023, 2010.
- [39] I. E. G. Morrison, I. Karakikes, R. E. Barber, N. Fernández, and R. J. Cherry, “Detecting and Quantifying Colocalization of Cell Surface Molecules by Single Particle Fluorescence Imaging,” *Biophysical Journal*, vol. 85, no. 6, pp. 4110–4121, 2003.
- [40] I. Koyama-Honda, K. Ritchie, T. Fujiwara et al., “Fluorescence imaging for monitoring the colocalization of two single molecules in living cells,” *Biophysical Journal*, vol. 88, no. 3, pp. 2126–2136, 2005.
- [41] L. Feng, Y. Xu, Y. Yang, and X. Zheng, “Multiple dense particle tracking in fluorescence microscopy images based on multidimensional assignment,” *Journal of Structural Biology*, vol. 137, no. 2, pp. 219–228, 2011.
- [42] A. Genovesio, T. Liedl, V. Emiliani, W. J. Parak, M. Coppey-Moisán, and J. C. Olivo-Marin, “Multiple particle tracking in 3-D+t microscopy: method and application to the tracking of endocytosed quantum dots,” *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1062–1070, 2006.
- [43] T. Pécot, C. Kervrann, and P. Bouthemy, “Minimal paths and probabilistic models for origin-destination traffic estimation in live cell imaging,” in *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI ’08)*, pp. 843–846, May 2008.
- [44] J. C. Olivo-Marin, “An overview of image analysis in multidimensional biological microscopy,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’06)*, pp. V1173–V1176, May 2006.
- [45] 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.
- [46] S. Y. Chen and Y. F. Li, “Vision sensor planning for 3-D model acquisition,” *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 35, no. 5, pp. 894–904, 2005.
- [47] J. Yu, C. Wu, S. P. Sahu, L. P. Fernando, C. Szymanski, and J. McNeill, “Nanoscale 3D tracking with conjugated polymer nanoparticles,” *Journal of the American Chemical Society*, vol. 131, no. 51, pp. 18410–18414, 2009.
- [48] K. Braeckmans, K. Buyens, B. Naeye et al., “Advanced fluorescence microscopy methods illuminate the transfection pathway of nucleic acid nanoparticles,” *Journal of Controlled Release*, vol. 148, no. 1, pp. 69–74, 2010.
- [49] S. van de Linde, S. Wolter, and M. Sauer, “Single-molecule photoswitching and localization,” *Australian Journal of Chemistry*, vol. 64, no. 5, pp. 503–511, 2011.
- [50] B. M.I. Flier, M. Baier, J. Huber et al., “Single molecule fluorescence microscopy investigations on heterogeneity of translational diffusion in thin polymer films,” *Physical Chemistry Chemical Physics*, vol. 13, no. 5, pp. 1770–1775, 2011.
- [51] M. Bates, B. Huang, M. J. Rust, G. T. Dempsey, W. Wang, and X. Zhuang, “Sub-diffraction-limit imaging with stochastic

- optical reconstruction microscopy,” *Springer Series in Chemical Physics*, vol. 96, pp. 400–415, 2010.
- [52] J. Hohlbein, K. Gryte, M. Heilemann, and A. N. Kapanidis, “Surfing on a new wave of Single-molecule fluorescence methods,” *Physical Biology*, vol. 7, no. 3, Article ID 031001, 2010.
- [53] M. C. Konopka and J. C. Weisshaar, “Heterogeneous motion of secretory vesicles in the actin cortex of live cells: 3D tracking to 5-nm accuracy,” *Journal of Physical Chemistry A*, vol. 108, no. 45, pp. 9814–9826, 2004.
- [54] N. G. Walter, C. Y. Huang, A. J. Manzo, and M. A. Sobhy, “Do-it-yourself guide: how to use the modern single-molecule toolkit,” *Nature Methods*, vol. 5, no. 6, pp. 475–489, 2008.
- [55] I. Smal, M. Loog, W. Niessen, and E. Meijering, “Quantitative comparison of spot detection methods in fluorescence microscopy,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, Article ID 5109713, pp. 282–301, 2010.
- [56] C. Jung, J. Kirstein, B. Platschek et al., “Diffusion of oriented single molecules with switchable mobility in networks of long unidimensional nanochannels,” *Journal of the American Chemical Society*, vol. 130, no. 5, pp. 1638–1648, 2008.
- [57] Z. J. Lü, J. Z. Lu, Y. Q. Wu, and L. Y. Chen, “Introduction to theories of several super-resolution fluorescence microscopy methods and recent advance in the field,” *Progress in Biochemistry and Biophysics*, vol. 36, no. 12, pp. 1626–1634, 2009.
- [58] W. E. Moerner, “Single-molecule mountains yield nanoscale cell images,” *Nature Methods*, vol. 3, no. 10, pp. 781–782, 2006.
- [59] I. Smal, K. Draegestein, N. Galjart, W. Niessen, and E. Meijering, “Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: application to microtubule growth analysis,” *IEEE Transactions on Medical Imaging*, vol. 27, no. 6, pp. 789–804, 2008.
- [60] I. F. Sbalzarini and P. Koumoutsakos, “Feature point tracking and trajectory analysis for video imaging in cell biology,” *Journal of Structural Biology*, vol. 151, no. 2, pp. 182–195, 2005.
- [61] W. Y. Ji, A. Bruckbauer, W. J. Fitzgerald, and D. Klenerman, “Bayesian inference for improved single molecule fluorescence tracking,” *Biophysical Journal*, vol. 94, no. 12, pp. 4932–4947, 2008.
- [62] E. Toprak and P. R. Selvin, “New fluorescent tools for watching nanometer-scale conformational changes of single molecules,” *Annual Review of Biophysics and Biomolecular Structure*, vol. 36, pp. 349–369, 2007.
- [63] M. Petrou, M. Jaward, S. Chen, and M. Briers, “Super-resolution in practice: the complete pipeline from image capture to super-resolved subimage creation using a novel frame selection method,” *Machine Vision and Applications*. In press.
- [64] D. Lin, W. Ma, X. Liu, Y. Sun, and Y. Wu, “Three-dimensional imaging and tracking for constrained Brownian motion of single nano-scaled particles in solution,” in *Proceedings of the 1st International Congress on Image and Signal Processing (CISP ’08)*, vol. 4, pp. 244–248, May 2008.
- [65] S. Jung and R. M. Dickson, “Hidden Markov analysis of short single molecule intensity trajectories,” *Journal of Physical Chemistry B*, vol. 113, no. 42, pp. 13886–13890, 2009.
- [66] S. S. Rogers, T. A. Waigh, X. Zhao, and J. R. Lu, “Precise particle tracking against a complicated background: polynomial fitting with Gaussian weight,” *Physical Biology*, vol. 4, no. 3, pp. 220–227, 2007.
- [67] 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, 12 pages, 2012.
- [68] J. F. Dorn, G. Danuser, and G. Yang, “Computational processing and analysis of dynamic fluorescence image data,” *Methods in Cell Biology*, vol. 85, pp. 497–538, 2008.

Research Article

Modeling and Representation of Human Hearts for Volumetric Measurement

Qiu Guan,¹ Wanliang Wang,¹ and Guang Wu^{2,3}

¹ College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

² Guangxi Academy of Sciences, 98 Daling Road, Nanning 530007, China

³ DreamSciTech Consulting, Shenzhen 518054, China

Correspondence should be addressed to Wanliang Wang, wwl@zjut.edu.cn

Received 24 July 2011; Accepted 28 August 2011

Academic Editor: Carlo Cattani

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

This paper investigates automatic construction of a three-dimensional heart model from a set of medical images, represents it in a deformable shape, and uses it to perform volumetric measurements. This not only significantly improves its reliability and accuracy but also makes it possible to derive valuable novel information, like various assessment and dynamic volumetric measurements. The method is based on a flexible model trained from hundreds of patient image sets by a genetic algorithm, which takes advantage of complete segmentation of the heart shape to form a geometrical heart model. For an image set of a new patient, an interpretation scheme is used to obtain its shape and evaluate some important parameters. Apart from automatic evaluation of traditional heart functions, some new information of cardiovascular diseases may be recognized from the volumetric analysis.

1. Introduction

The research to diagnose and prevent cardiovascular diseases becomes more important than ever. Thanks to the newly developed technologies in medical imaging and computing, automatic evaluation of patient hearts now becomes possible. This is very useful for diagnosis and treatment. If all kind of cardiovascular diseases were cured, human life could be much longer [1]. Modeling and volumetric measurement have the capability to improve the diagnostic value of cardiac images. Ventricular volume and representation make physician to evaluate the consequences of myocardial infarction according to a glance at the anatomy of the ventricles. Besides, Parameters rely on left ventricular volume such as the ejection fraction, which is a kind of measurement of ventricular ability to pump oxygenated blood through the body. Since cardiologists rely on these values as indicators of cardiac malfunction, it is useful for us to calculate the ventricular volumes [2–6].

Researchers have attempted much in this issue, such as medical image acquisition, image processing, feature enhancement and extraction, cardiac boundary segmentation, parameter computation, functional formulation, disease as-

essment, and so forth. A number of methods have also been attempted for carrying out these works. The contributions can be extensively explored in the literature [7–12]. Typically, Yamamuro et al. carried out a project on two-dimensional (2D) image processing [7]. They evaluate accuracy of cardiac functional analysis with multidetector-row computed tomography (CT) and segmental reconstruction algorithm over a range of heart rates. Various functional parameters of the left ventricle are measured, and they are correlated and agreed with those obtained with magnetic resonance imaging (MRI). For the single-photon emission computed tomography, Germano et al. have developed an algorithm to quantitatively measure left ventricular ejection fraction from gated ^{99m}Tc-sestamibi myocardial perfusion images [8]. The algorithm operates in the three-dimensional (3D) space and uses gated short-axis image volumes. It segments the ventricle, estimates and displays endocardial and epicardial surfaces for all gating intervals in the cardiac cycle, calculates the relative left ventricular cavity volumes, and derives the global ejection fraction from the end-diastolic and end-systolic volume. Results show that the automatic segmentation and contouring of the ventricle were very successful of the experimental studies. Relevant programs were also

developed to provide clinically useful additional information to complement myocardial perfusion studies in hospitals. However, the limitation of these existing works lies in the lack of cardiac shape completeness, correctness or regularity, and fitting reliability.

This paper is concerned with modeling and analysis of 3D cardiac model (mainly the ventricle) of human heart and fitting the model to other patients' 3D cardiac image volumes to form the patient's subject-specified model. The model building includes semiautomated registration of shapes for all patients in the database, alignment, and decomposition into a flexible model. For a new case, the model is used to segment and fit 3D cardiac images. Finally quantitative functional analysis of the left ventricle is followed, which plays a very important role in the automatic/aided diagnosis of cardiac diseases. The method utilized in this research is a model-based approach. It includes two meanings. Firstly, the heart model is built by statistical analysis of many existing images from the database which are obtained from several hospitals, which creates a general heart model and possible variation. Secondly, the cardiac shape of a new patient is evaluated from the deviation of its shape from normal cases.

The remainder of this paper is organized as follows. In Sections 2 and 3, we introduce the method for model creation and training. Sections 4 and 5 describe issues of model representation and interpretation. Mathematic method of volumetric measurement is given in Section 6. Example experimental results are given in Section 7. Finally, conclusions are drawn in Section 8.

2. Model Creation

The key to quantitative analysis of cardiac functions is to recognize the dynamical and deformable cardiac shape from medical images. In recent decades, researchers have contributed many ideas for segmenting, fitting, and modeling of deformable shapes [13, 14]. The most frequently used ones are Deformable Models, Classification, Global Search, Snakes, Level-set, Eigen-faces and Eigen-patches, and so forth. Some of them are based on minimization of an energy function (e.g., Snakes). However, the energy function method searches the object boundary stopping on most rapid changes of gray-level image profiles. Due to the complexity of deformable object in human body where many other organs are located near it, the gray-level gradients sometimes cannot be used to describe the organ anatomical edges actually. Another typical method, the eigen-patches, is also used to model regions where the shape is assumed to be fixed. However, the problem comes when the medical organ is usually not fixed, for example, when the heart is beating. These require a method for not only modeling the object shape but also its shape variation. From our investigation, a method based on a flexible model is adopted to provide principled means to efficiently parameterize a cardiac shape and its variability. The model is initiated from the active shape modeling method described by Cootes et al. [15–17]. It not only satisfies the requirements of cardiac shape modeling and analysis, but also allows dimensionality reduction of the model for reducing the implementation cost.

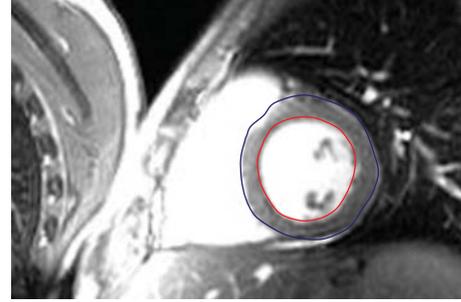


FIGURE 1: Segmentation of ventricular shapes from its background. In the training step, experienced doctors tell where the accurate boundaries are. The program thus creates a trained model representing the general shape of the hearts.

In the first stage, a statistical model has to be created with a given set of cardiac examples for representing the general shape and its variation. For digital computation, each shape in the training set is represented by a set of labeled feature points, which must be consistent from one shape to another. For instance, a number of 3D points on a ventricle shape should always correspond to the same locations in the biological or anatomical sense. Figure 1 illustrates a segmentation program for training a general model of the ventricular shape. The segmentation results are also used as the input of the standard for evaluating automatic shape fitting by the proposed model-based approach.

To build a model of the general shape, it requires labeled training images to represent correspondences among the shape examples. Automatic landmarking methods for this purpose have been thoroughly studied by researchers in recently years. For example, Izard et al. presented a method for landmarking MR images in registering brain structures from different images using a generic algorithm [18]. With a set of such labeled training examples, we need to align them into a common coordinate frame. The Generalized Procrustes Analysis can be used to align training shapes and minimize the sum of squared distances to the mean of the set. In fact, it is to find the transformations T_i which minimize

$$G = \sum |\mathbf{m} - T_i(\mathbf{x}_i)|^2, \quad (1)$$

where

$$\mathbf{m} = \frac{1}{n} \sum T_i(\mathbf{x}_i), \quad |\mathbf{m}| = 1, \quad (2)$$

and \mathbf{x}_i is a specific shape example in the training set which is represented by a $3n$ element vector (for n points of landmarks in 3D space), that is,

$$\mathbf{x} = (x_1, y_1, z_1, \dots, x_n, y_n, z_n)^T. \quad (3)$$

The aligned training set forms a cloud in the 3D space and can be considered as a sample from a probability density function. To reduce the computation cost and memory requirement, we use principal component analysis (PCA) to pick out the main axes of the cloud, and model only the first few, which account for the majority of the variation. For

the ventricular model, the first 50 principal eigen vectors are good enough to represent the shape and variation. The general model is then represented as

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{b}, \quad (4)$$

where $\bar{\mathbf{x}}$ is the mean of the aligned shapes, \mathbf{b} is a $3n \times t$ matrix whose columns are unit vectors along the principal axes, and \mathbf{b} is a t element vector of shape parameters. This means the shape dimension is reduced from $3n$ to t by PCA analysis.

This creates a statistical model like the point distribution model, and such a model is used in the flexible model framework to locate new examples in new images. By varying the shape parameters in \mathbf{b} within limits learnt from the training set, we can generate new plausible shapes. Usually the variance of b_j is λ_j (the eigen value of the j th largest in the matrix).

3. Model Training by Genetic Algorithm

Genetic algorithm (GA) is introduced as a computational analogy of adaptive systems. It is modeled loosely on the principles of the evolution via natural selection, employing a population of individuals that undergo selection in the presence of variation-inducing operators such as mutation and recombination (crossover). A fitness function is used to evaluate individuals, and reproductive success varies with fitness, and therefore GA is a better way in global search, and passed few years witness its widely applications. There are definitely some laws and sequence of species existing in the nature. Taking medical images, for example, when the artificial factors are excluded, the scale and rotation variety of required images obey the normal distribution. Besides, the information we get from the normal distribution can efficiently help us establish a more reasonable model.

Generally, the statistical model uses the first shape of the training set or an arbitrary shape to be the first meanshape. However, such meanshape is not a ‘‘mean’’ finally. Using GA to form the first meanshape is an alternative choice. We can then get the information of the normal distribution and make the meanshape so as to improve the searching efficiency. Below are some strategies that can be adopted in the fitness function [14, 19].

Strategy 1 (obtain the model parameters by GA). As the amount of GA arithmetic parameters should be controlled when considering the factor of time. The first shape can be used as initial meanshape to form the model, and the model parameters are generated using the GA algorithm. The sum of Euclidean distances between each shape and the meanshape can be taken as the fitness function

$$f_1 = \sqrt{\sum_{i=1}^n |(\bar{\mathbf{x}} + P\mathbf{b}_g) - \mathbf{x}_i|^2}. \quad (5)$$

Strategy 2 (generate the shape directly from GA). Directly using the coordinates of the shape points together as the input parameter of GA, the sum of Euclidean distances between

each shape and the meanshape can be used in the fitness function. It is an efficient method for the shape without many points. Its computation time will be increased rapidly with the more points

$$f_2 = \sqrt{\sum_{i=1}^n \left((x_g - x_i)^2 + (y_g - y_i)^2 \right)}. \quad (6)$$

Strategy 3 (generate each point one by one). Searching the optimum point using GA arithmetic with each point as the input parameter, the fitness of GA is the sum of the distance between the points generated by the GA and the other corresponding points in the training set. Suitable weights are added to reflect the significance of some shapes in the training set. All the points are fitted together as the first mean shape to participate in the align procedure

$$f_3 = \sum_{i=1}^n \sqrt{\left((x_g - x_i)^2 + (y_g - y_i)^2 \right)}. \quad (7)$$

GA searches for the optimum solution, and the running time expands more or less when using GA for model generation. Fortunately, this problem is not very serious in the modeling process.

4. Model Representation

In this paper, a B-Spline surface model is used to represent the shape and get its volume [20, 21]. By applying matrix form of B-Spline, we obtain a polynomial B-Spline representation with two unitary parameters. Polynomials of B-Spline surfaces make the integral possible. This approach provides an actual volume of B-Spline surface, and it is also convenient and quick. Actually, a part of the B-Spline surface can be represented by the following matrix form:

$$s_{i,j}(t, w) = T_{k_1} M_{i,u}^{k_1+1} V_{i,j}^h (M_{j,v}^{k_2+1})^T W_{k_2}^T, \quad (8)$$

where T , W , and M are basis matrixes in the nonempty intervals for the B-Spline surface. V contains the control points:

$$V_{i,j}^{k_1,k_2} = \begin{bmatrix} \mathbf{V}_{i-k_1,j-k_2}^{k_1,k_2} & \mathbf{V}_{i-k_1,j-k_2+1}^{k_1,k_2} & \cdots & \mathbf{V}_{i-k_1,j}^{k_1,k_2} \\ \mathbf{V}_{i-k_1+1,j-k_2}^{k_1,k_2} & \mathbf{V}_{i-k_1+1,j-k_2+1}^{k_1,k_2} & \cdots & \mathbf{V}_{i-k_1+1,j}^{k_1,k_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{V}_{i,j-k_2}^{k_1,k_2} & \mathbf{V}_{i,j-k_2+1}^{k_1,k_2} & \cdots & \mathbf{V}_{i,j}^{k_1,k_2} \end{bmatrix}. \quad (9)$$

Equation (8) can also be rewritten as (10) or split into three scalar equations (11)

$$S_{i,j}(t, w) = \sum_{l=0}^{k_2} \sum_{r=0}^{k_1} B_{i,j}(r, l) t^r w^l, \quad (10)$$

$$\begin{aligned}
X(t, w) &= \sum_{i=k_1}^m \sum_{j=k_2}^n \sum_{l=0}^{k_2} \sum_{r=0}^{k_1} B_{i,j}^x(r, l) t^r w^l, \\
Y(t, w) &= \sum_{i=k_1}^m \sum_{j=k_2}^n \sum_{l=0}^{k_2} \sum_{r=0}^{k_1} B_{i,j}^y(r, l) t^r w^l, \\
Z(t, w) &= \sum_{i=k_1}^m \sum_{j=k_2}^n \sum_{l=0}^{k_2} \sum_{r=0}^{k_1} B_{i,j}^z(r, l) t^r w^l.
\end{aligned} \tag{11}$$

5. Interpretation

After we get a statistical model trained from the sets of examples, it is ready to interpret new images. The heart is located in the chest between the lungs behind the sternum and above the diaphragm. It is surrounded by the pericardium. It has the great vessels: the superior and inferior vena cava, the pulmonary artery and vein, and the aorta. The aortic arch lies behind the heart. The esophagus and the spine lie further behind the heart. This knowledge can help the computer to roughly put the cardiac model into a new image volume keeping not too far from the true position for accelerating the interpreting process using the created model.

An iterative method is used for matching the model to images. It iteratively deforms to fit to image volume of the ventricle. The shapes are constrained by statistical derivations to vary only in ways seen in a training set of labeled examples. In addition to the shape model, we require models of the image appearance around each model point. It can be built to represent the statistical variation of the gradient along profiles through the points, normal to the boundary curve at that point. The true boundary position can be found by computing the distance for the statistical profile moving along the image profile [22–25]. Finally the ventricular shape of the patient is fitted by repeating the following two steps until convergence: (1) look along normals through each model point to find the best local match for the model of the image appearance at that point (with minimum distance); (2) update the pose and shape parameters to best fit the model instance to the found points. This is to say that the goal of model fitting is to search best candidate image points near the model and update global transformation, T , and parameters, \mathbf{b} , to minimize

$$\begin{aligned}
f &= |\mathbf{X} - T(\bar{\mathbf{x}} + \mathbf{P}\mathbf{b})|^2 \\
&= |\mathbf{X} - T(\bar{\mathbf{x}} + \mathbf{P}\mathbf{b}; X_c, Y_c, Z_c, s, \alpha, \beta, \gamma)|^2,
\end{aligned} \tag{12}$$

where \mathbf{X} is the temporary model obtained in the immediate steps. This minimization can be achieved by some nonlinear optimizers with iterative approaches. Finally the pose parameters in T are fixed, and we get the corresponding shape parameters of the patient.

6. Volumetric Measurement

There are mainly three traditional approaches to calculate the ventricular volume. First, the ventricular volume is represented as the volume of a simple shape (e.g., truncated el-

lipse) or a combination of different figures. This method is simple to perform while the volume is coarse. Secondly, the ventricular volume is regarded as the sum of multiple smaller volumes of similar configuration. Thirdly, the ventricular volume is represented as the volume of B-Spline surface, and the volume is usually computed by using numerical integral such as Simpson's rule and Gauss's rule.

To calculate the volume of B-Spline surfaces, the polynomial expressions of B-Spline surface are given [20]. Let $\partial S/\partial t$ and $\partial S/\partial w$ denote the derivatives of $S(u, v)$ with respect to t and w , respectively. The volume can be determined by

$$\begin{aligned}
V &= \sum_{i=k_1}^m \int_0^1 A(z(w)) z'(w) dw \\
&= \sum_{i=k_1}^m \sum_{j=k_2}^n \iint_0^1 Y(t, w) \times \frac{\partial X(t, w)}{\partial t} \times \frac{\partial Z(t, w)}{\partial w} dt dw \\
&= \sum_{i=k_1}^m \sum_{j=k_2}^n \iint_0^1 \sum_{l_1=0}^{(3 \times k_2 - 1)} \sum_{l_2=0}^{(3 \times k_1 - 1)} C_{i,j}(l_1, l_2) \times t^{l_2} \times w^{l_1} dt dw.
\end{aligned} \tag{13}$$

With a $k_1 \times k_2$ order B-Spline surface, it can be finally written as

$$V = \sum_{i=k_1}^m \sum_{j=k_2}^n \left(\sum_{l_1=1}^{(3 \times k_2 - 1)} \sum_{l_2=1}^{(3 \times k_1 - 1)} C_{i,j}(l_1, l_2) \times \frac{1}{l_2 + 1} \times \frac{1}{l_1 + 1} \right), \tag{14}$$

where $C_{i,j}$ is a $3k_1 \times 3k_2$ matrix which is the product of the three polynomials:

$$\begin{aligned}
C_{i,j}(l_1, l_2) &= \sum_{\substack{f=0 \\ d=l_1-f \\ 0 \leq d \leq k_2}}^{2k_2-1} \sum_{\substack{f_2=0 \\ d_2=l_2-f_2 \\ 0 \leq d_2 \leq k_1-1}}^{2k_1} \left(\sum_{\substack{r=0 \\ s=f-r \\ 0 \leq s \leq k_2-1}}^{k_2} \sum_{\substack{r_2=0 \\ s_2=f_2-r_2 \\ 0 \leq s_2 \leq k_1}}^{k_1} B_{i,j}^y(r, r_2) \right. \\
&\quad \left. \times s \times B_{i,j}^x(s, s_2) \right) \\
&\quad \times d_2 \times B_{i,j}^z(d, d_2).
\end{aligned} \tag{15}$$

7. Experiments

Medical images give functional information about the heart while having less information on its anatomy. As it is well known that one of the main diagnostic parameters of interest for physician is its volume. In this paper, some experiments are carried out to construct the left ventricular surfaces fitted by B-spline model (Figure 2). From one cardiac cycle, the changes in volume can be obtained while the heart beats. The ventricular volumes are determined by the algorithms. Table 1 shows some of the volumes sampled from a cardiac cycle.

TABLE 1: Volumes at different phases in a cardiac cycle.

| Time (ms) | 100 | 300 | 500 | 700 |
|--------------------|------|------|------|------|
| Endocardium volume | 70.7 | 49.2 | 85.2 | 98.6 |
| Epicardium volume | 262 | 203 | 294 | 352 |

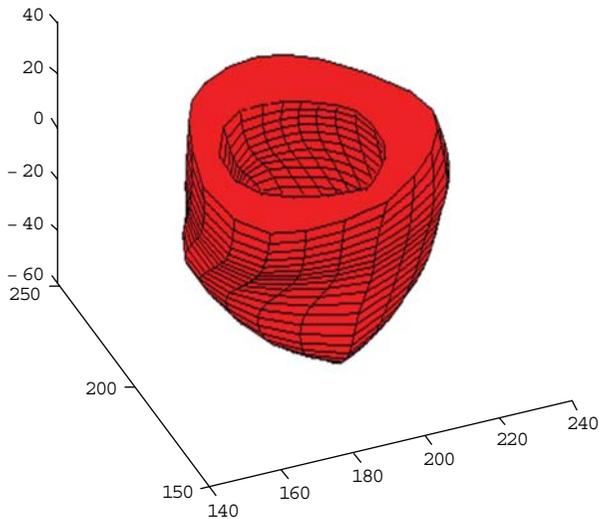


FIGURE 2: The B-spline surface model of a left ventricle.

8. Conclusions

This paper presented a model-based approach for volumetric analysis of human hearts, especially for the ventricles, which is very important for diagnosis and treatment of cardiovascular diseases. The method is based on a flexible model combined with genetic strategies. Based on the flexible model trained from hundreds of patient images, a new patient will be actively analyzed to obtain its individual shape. We also adopted an efficient method for representation of 3D surfaces and provided a corresponding volumetric measurement algorithm of the B-Spline surface. The volumetric algorithm of a B-Spline model is important in working out other functional parameters of human hearts. The pipeline proposed in this paper takes advantage of complete segmentation of the heart shape. It can automatically construct a 3D model from a set of medical images. This not only significantly improves its reliability but also makes it possible to derive valuable information to doctors, such as dynamic volumetric measurements.

Acknowledgments

Thanks to Z. Xu, T. T. Jiang, J. Zhang, Y. Q. Xu, Y. H. Du, and C. Jiang for their helpful research and laboratory work in the project during their study in the university. This work was supported by the National Natural Science Foundation of China (NSFC nos. 61103140 and 61105073) and Zhejiang Provincial S&T Department (2010R10006, 2010C33095).

References

- [1] 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.
- [2] D. Moroni, S. Colantonio, O. Salvetti, and M. Salvetti, "Heart deformation pattern analysis through shape modelling," *Pattern Recognition and Image Analysis*, vol. 19, no. 2, pp. 262–270, 2009.
- [3] X. Huang and D. N. Metaxas, "Metamorphs: deformable shape and appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1444–1459, 2008.
- [4] R. I. Ionasec, I. Voigt, B. Georgescu et al., "Patient-specific modeling and quantification of the aortic and mitral valves from 4-D cardiac CT and TEE," *IEEE Transactions on Medical Imaging*, vol. 29, no. 9, Article ID 5458068, pp. 1636–1651, 2010.
- [5] S. Y. Chen and Q. Guan, "Parametric shape representation by a deformable NURBS model for cardiac functional measurements," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 480–487, 2011.
- [6] T. Arts, F. W. Prinzen, T. Delhaas, J. R. Milles, A. C. Rossi, and P. Clarysse, "Mapping displacement and deformation of the heart with local sine-wave modeling," *IEEE Transactions on Medical Imaging*, vol. 29, no. 5, Article ID 5437350, pp. 1114–1123, 2010.
- [7] M. Yamamuro, E. Tadamura, S. Kubo et al., "Cardiac functional analysis with multi-detector row CT and segmental reconstruction algorithm: comparison with echocardiography, SPECT, and MR imaging," *Radiology*, vol. 234, no. 2, pp. 381–390, 2005.
- [8] G. Germano, H. Kiat, P. B. Kavanagh et al., "Automatic quantification of ejection fraction from gated myocardial perfusion SPECT," *Journal of Nuclear Medicine*, vol. 36, no. 11, pp. 2138–2147, 1995.
- [9] C. Lorenz and J. V. Berg, "A comprehensive shape model of the heart," *Medical Image Analysis*, vol. 10, no. 4, pp. 657–670, 2006.
- [10] M. E. Barley, A. M. Galea, A. A. Armoundas, T. S. Rosbury, and G. B. Hirschman, "Validation of a novel catheter guiding method for the ablative therapy of ventricular tachycardia in a phantom model," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, Article ID 4663616, pp. 907–910, 2009.
- [11] S. Y. Chen, "Cardiac deformation mechanics from 4D images," *Electronics Letters*, vol. 43, no. 11, pp. 609–611, 2007.
- [12] G. Luo and P. A. Heng, "LV shape and motion: B-spline-based deformable model and sequential motion decomposition," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 3, pp. 430–446, 2005.
- [13] O. Ecabert, J. Peters, H. Schramm et al., "Automatic model-based segmentation of the heart in CT images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, Article ID 4505365, pp. 1189–1202, 2008.
- [14] H. Lu, S. Liu, W.-L. Wang, and S. Y. Chen, "Generation of a point distribution model using genetic algorithms," in *Proceedings of the 4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications*, Tunisia, March 2007.
- [15] T. F. Cootes and C. J. Taylor, "Statistical model of appearance for computer vision," Tech. Rep., University of Manchester, Manchester, UK, 2004.
- [16] A. Andreopoulos and J. K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI," *Medical Image Analysis*, vol. 12, no. 3, pp. 335–357, 2008.

- [17] J. Koikkalainen, T. Tolli, K. Lauerma et al., "Methods of artificial enlargement of the training set for statistical shape models," *IEEE Transactions on Medical Imaging*, vol. 27, no. 11, Article ID 4591396, pp. 1643–1654, 2008.
- [18] C. Izard, B. Jedynek, and C. Stark, "Automatic landmarking of magnetic resonance brain images," in *Proceedings of the SPIE International Symposium on Medical Imaging*, vol. 5747, pp. 1329–1340, San Diego, Calif, USA, 2005.
- [19] C. McIntosh and G. Hamarneh, "Genetic algorithm driven statistically deformed models for medical image segmentation," in *Proceedings of the Genetic and Evolutionary Computation Conference, (GECCO '06)*, Image Processing and Computer Vision, Seattle, Wash, USA, July 2006.
- [20] T. T. Jiang, S. Y. Chen, and Y. Xu, "3-D representation and volumetric measurement of human heart from a cylindrical B-spline surface model," in *Proceedings of the International Conference on BioMedical Engineering and Informatics*, pp. 765–769, Sanya, Hainan, China, May 2008.
- [21] B. Zhang and J. F. M. Molenbroek, "Representation of human head with bi-cubic B-Spline technique based on the laser scanning technique in 3D surface anthropometry," *Applied Ergonomics*, vol. 35, no. 5, pp. 459–465, 2004.
- [22] S. Y. Chen and Y. F. Li, "Determination of stripe edge blurring for depth sensing," *IEEE Sensors Journal*, vol. 11, no. 2, Article ID 5585653, pp. 389–390, 2011.
- [23] S. Y. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.
- [24] A. I. Veress, W. P. Segars, B. M. W. Tsui, and G. T. Gullberg, "Incorporation of a left ventricle finite element model defining infarction into the XCAT imaging phantom," *IEEE Transactions on Medical Imaging*, vol. 30, no. 4, pp. 915–927, 2011.
- [25] K. Punithakumar, I. B. Ayed, A. Islam, I. G. Ross, and S. Li, "Tracking endocardial motion via multiple model filtering," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 8, Article ID 5471238, pp. 2001–2010, 2010.

Research Article

Medicoeconomic Index for Photo-Induced Skin Cancers

Ioana Chiorean,¹ Liana Lupşa,¹ Luciana Neamţiu,² and Maria Crişan³

¹ Faculty of Mathematics and Computer Science, Babeş-Bolyai University, 400084 Cluj-Napoca, Romania

² Department of Epidemiology and Biostatistics, Cancer Institute Ion Chiricuţă, 40445 Cluj-Napoca, Romania

³ Department of Histology, University of Medicine and Pharmacy "Iuliu Haţieganu", Cluj-Napoca, Romania

Correspondence should be addressed to Maria Crişan, mcrisan7@yahoo.com

Received 31 July 2011; Accepted 10 September 2011

Academic Editor: Carlo Cattani

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

Like in every type of cancer, in skin cancer the efficiency of the medical treatment is very important. In the present paper, a Bayesian model for the management of this disease is given, and a medical index to measure the effectiveness of treatment from medical, economical, and quality of life point of view is presented, taking into account some of the patients characteristics.

1. Introduction

During the last 10–15 years an alarming increase of the photo-induced cutaneous cancers was registered nationally and internationally (1.000.000 new cases in 2007 in the USA), reaching epidemic proportions in some areas. At the moment, the number of patients with cutaneous cancers has significantly increased in Romania. The real incidence of this pathology is unknown, partly due to the lack of an integrated national or regional database. The phenomenon is more frequent in the case of persons who uncontrollably expose themselves to ultraviolet rays for cosmetic or therapeutic reasons or due to improper work conditions, without protection outfits, but it depends also on other reasons, as photo type, age, gender, and so forth. Since we witness a continuous process of ozone layer alteration, global warming, and increase in the pollution degree, the prevention of photo-induced cutaneous cancers and elaboration of protocols for precocious diagnosis and innovative treatment appear as a priority issue of public health (see [1]).

One strategy used in medicine to identify and treat a disease is by screening. Its purpose is to identify in an early stage some malady in a community, enabling timely intervention and management in order to reduce suffering and mortality. In the specific case of skin cancer, this strategy also functions. In order to obtain good results, the measurements have to be taken from different points of view. For

instance, most published researches are based on the study of cost and effectiveness of a treatment (see [2–6]), taking into account different characteristics of the patients. The present paper proposes to determine the efficiency of the treatment also from the quality of life point of view. The results are obtained by means of a Bayesian approach for determining the involved parameters, and, as a measure, the economical index is obtained.

2. Bayesian Regression Model

We consider a sample of n individuals with melanoma in different stages, participating in a clinical trial and denote by (Ef_i) the effectiveness data, by (C_i) cost data, and by (Q_i) the quality of life for each patient i , $i \in \{1, \dots, n\}$. These n patients are considered to be split in two different groups: those who are under treatment and those who are not. The results of the clinical trial, in terms of effectiveness, costs, and quality of life, are not determined only by the type of treatment received, and so it is necessary to consider a series of possible covariates that may influence the above results. Such covariates include the patient's photo type, age, state of health at the time of the clinical trial, gender, and other characteristics that depend on the type of clinical trial under analysis. We define X as an $n \times (k + 1)$ matrix of covariates, where each column (X_i) refers to one covariate. The first column is a column of ones referring to

the constant (see [7, 8]). We seek, therefore, to explain the results obtained (Ef_i , C_i , and Q_i), as a linear combination of the k covariates considered (the patient's individual characteristics and the type of treatment received). For this purpose, we propose a Bayesian multiple linear regression model in which the perturbation terms (u_i , v_i , and w_i) are assumed to be Gaussian, independent, and identically distributed with a mean of 0 and variances of σ_1^2 , σ_2^2 , and σ_3^2 , respectively,

$$\begin{aligned} Ef_i &= \beta_0 + \beta_1 \cdot X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_{k-1,i} + \beta_{T,i} + u_i, \\ C_i &= \delta_0 + \delta_1 \cdot X_{1,i} + \delta_2 X_{2,i} + \dots + \delta_{k-1,i} + \delta_{T,i} + v_i, \\ Q_i &= q_0 + q_1 \cdot X_{1,i} + q_2 X_{2,i} + \dots + q_{k-1,i} + q_{T,i} + w_i, \end{aligned} \quad (1)$$

where the vectors $\beta = (\beta_0, \beta_1, \dots, \beta_{k-1}, \beta_T)$, $\delta = (\delta_0, \delta_1, \dots, \delta_{k-1}, \delta_T)$, $q = (q_0, q_1, \dots, q_{k-1}, q_T)$, and the accuracy values $\tau_1 = \sigma_1^{-2}$, $\tau_2 = \sigma_2^{-2}$, $\tau_3 = \sigma_3^{-2}$ are the parameters of the model.

The k covariates considered for which data are available need not to be explicative of the effectiveness, the costs, and the quality of life, and so the above general model could be corrected by eliminating those covariates that do not explain effectiveness and cost.

The first step to be taken in estimating the parameters is to determine the likelihood function of effectiveness $l_e(Ef | \beta, \tau_1)$, $l_c(C | \delta, \tau_2)$, $l_q(Q | q, \tau_3)$, where $Ef = (Ef_1, \dots, Ef_n)$, $C = (C_1, \dots, C_n)$, and $Q = (Q_1, \dots, Q_n)$. In this stage both costs, effectiveness and qualities of life are assumed to present a normal distribution, and so the likelihood functions are represented by the following expressions:

$$\begin{aligned} l(Ef, C, Q | \beta, \delta, q, \tau_1, \tau_2, \tau_3) \\ = l_e(Ef | \beta, \tau_1) \cdot l_c(C | \delta, \tau_2) \cdot l_q(Q | q, \tau_3), \end{aligned} \quad (2)$$

where

$$\begin{aligned} l_e(Ef | \beta, \tau_1) &\propto \tau_1^{N/2} \cdot \exp\left\{\frac{\tau_1}{2} (Ef - X\beta)' (Ef - X\beta)\right\}, \\ l_c(C | \delta, \tau_2) &\propto \tau_2^{N/2} \cdot \exp\left\{\frac{\tau_2}{2} (C - X\delta)' (C - X\delta)\right\}, \\ l_q(Q | q, \tau_3) &\propto \tau_3^{N/2} \cdot \exp\left\{\frac{\tau_3}{2} (Q - Xq)' (Q - Xq)\right\}. \end{aligned} \quad (3)$$

Assuming model (1) from a Bayesian point of view, we must specify the prior distribution for the $3 \cdot k + 6$ parameters of the model. The prior distribution represents expert information about the set of model parameters before the sample observations are analysed. We propose a normal (N), respectively, gamma (G) form for the base prior and assume independence between the coefficients β , δ , q and precision terms τ_1 , τ_2 , τ_3 . Therefore,

$$\begin{aligned} \pi(\beta, \tau_1) &= \pi_{e,1}(\beta) \cdot \pi_{e,2}(\tau_1), \\ \pi(\delta, \tau_2) &= \pi_{c,1}(\delta) \cdot \pi_{c,2}(\tau_2), \\ \pi(q, \tau_3) &= \pi_{q,1}(q) \cdot \pi_{q,2}(\tau_3), \end{aligned} \quad (4)$$

where

$$\begin{aligned} \pi_{e,1}(\beta) &\approx N(\beta_0, V_1^{-1}), & \pi_{c,1}(\delta) &\approx N(\delta_0, V_2^{-1}), \\ \pi_{q,1}(q) &\approx N(q_0, V_3^{-1}), \\ \pi_{e,2}(\tau_1) &\approx G(a_1, b_1), & \pi_{c,2}(\tau_2) &\approx G(a_2, b_2), \\ \pi_{q,2}(\tau_3) &\approx G(a_3, b_3). \end{aligned} \quad (5)$$

The parameters β_0 , V_1 , δ_0 , V_2 , q_0 , V_3 , a_1 , b_1 , a_2 , b_2 , a_3 , b_3 , which determine the prior distribution, are defined on the basis of the initial information (available when the analysis begins). The joint posterior distribution of the parameters β , δ , q , τ_1 , τ_2 , τ_3 , given the data (Ef, C, Q), can be calculated from (4), using Bayes' theorem

$$\begin{aligned} \pi(\beta, \tau_1 | Ef) &\propto \tau_1^{(n+2a_1)/2-1} \\ &\cdot \exp\left\{-\frac{1}{2} \left[(Ef - X\beta)' (Ef - X\beta) \right. \right. \\ &\quad \left. \left. + (\beta - \beta_0) V_1^{-1} (\beta - \beta_0) + 2b_1 \tau_1 \right] \right\}, \\ \pi(\delta, \tau_2 | C) &\propto \tau_2^{(n+2a_2)/2-1} \\ &\cdot \exp\left\{-\frac{1}{2} \left[(C - X\delta)' (C - X\delta) \right. \right. \\ &\quad \left. \left. + (\delta - \delta_0) V_2^{-1} (\delta - \delta_0) + 2b_2 \tau_2 \right] \right\}, \\ \pi(q, \tau_3 | Q) &\propto \tau_3^{(n+2a_3)/2-1} \\ &\cdot \exp\left\{-\frac{1}{2} \left[(Q - Xq)' (Q - Xq) \right. \right. \\ &\quad \left. \left. + (q - q_0) V_3^{-1} (q - q_0) + 2b_3 \tau_3 \right] \right\}. \end{aligned} \quad (6)$$

Inferences about quantities of interest must be based on these posterior distributions. The prior distributions used here are not the only possible choices.

3. The Medicoeconomical Evaluation

The standard measure used to compare only the cost and effectiveness of treatments is the incremental cost-effectiveness ratio (ICER). Nevertheless, this measure presents severe interpretational problems, as well as difficulties in estimating the confidence or credibility intervals. The incremental net benefit (INB) has been proposed as an alternative to ICER (see [8]). The INB of treatment 1 (new) versus treatment 0 (actual or control) is defined as

$$\text{INB}(R_c) = R_c (\mu_1 - \mu_2) - (v_1 - v_2), \quad (7)$$

where μ 's and v 's are the mean effectiveness and cost of the respective treatments. The value R_c is interpreted as the cost that decision-makers are willing to accept in order to increase the effectiveness of the treatment applied by one unit. Thus, analyzing whether the alternative treatment is

more cost effective than the control treatment is equivalent to determining whether $INB(R_c)$ is positive.

For our problem, we use the medicoeconomical index denoted by $MEI(\lambda_e, \lambda_c, \lambda_q)$ introduced in [3], according to

$$MEI(\lambda_e, \lambda_c, \lambda_q) = \lambda_e(\mu_1 - \mu_0) + \lambda_c(\nu_1 - \nu_0) + \lambda_q(\gamma_1 - \gamma_0), \quad (8)$$

where μ 's, ν 's, and γ 's are the mean effectiveness, cost, and quality of life of the respective treatments. The values λ_e , λ_c , and λ_q are weights ($0 \leq \lambda_e \leq 1$, $0 \leq \lambda_c \leq 1$, $0 \leq \lambda_q \leq 1$, $\lambda_e + \lambda_c + \lambda_q = 1$) which reflect the importance we give to each criteria (the medication effect, the medication costs, the quality of life).

Acknowledgment

The study is part of the National Research Program CNCSIS-IDEI, no. 2624.

References

- [1] M. Crisan, C. Cattani, C. Badea et al., "Modelling cutaneous senescence process, computational science and its applications," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '10)*, vol. 6017, part 2, pp. 215–224, Springer, 2010.
- [2] A. Briggs, K. Claxton, and M. Sculpher, *Decision Modelling for Health Economic Evaluation*, Oxford, UK, Oxford University Press, 2006.
- [3] I. Chiorean, L. Lupşa, and L. Neamţiu, "Markov models for the simulation of cancer screening process," in *Proceedings of the AIP Conference: International Conference on Numerical Analysis and Applied Mathematics (ICNAAM '08)*, vol. 1048 of *Numerical Analysis and Applied Mathematics*, pp. 143–146, 2008.
- [4] A. O'Hagan and J. W. Stevens, "A framework for cost-effectiveness analysis from clinical trial data," *Health Economics*, vol. 10, no. 4, pp. 303–315, 2001.
- [5] F. J. Vázquez-Polo and M. A. Negrín-Hernández, "Incorporating patients' characteristics in cost-effectiveness studies with clinical trial data: a flexible Bayesian approach," *SORT*, vol. 28, no. 1, pp. 87–108, 2004.
- [6] A. R. Willan and A. H. Briggs, *Statistical Analysis of Cost-Effectiveness Data*, John Wiley & Sons, 2006.
- [7] A. H. Briggs, "A Bayesian approach to stochastic cost-effectiveness analysis," *Health Economics*, vol. 8, no. 3, pp. 257–261, 1999.
- [8] L. Neamţiu, *Cost-Effectiveness Study for Cervical Screening Program: A Bayesian Approach*, Bilateral Workshop, Königswinter, Germany, 2007.

Research Article

Target Contour Recovering for Tracking People in Complex Environments

Jianhua Zhang,¹ Sheng Liu,² Y. F. Li,³ and Jianwei Zhang¹

¹ Group TAMS, Department of Informatics, University of Hamburg, 22527 Hamburg, Germany

² College of Information Engineering, Zhejiang University of Technology, 18 Chaowang Road, Hangzhou 310014, China

³ Department of System Engineering and Engineering Management, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong

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

Received 13 June 2011; Accepted 18 August 2011

Academic Editor: Shengyong Chen

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

Recovering people contours from partial occlusion is a challenging problem in a visual tracking system. Partial occlusions would bring about unreasonable contour changes of the target object. In this paper, a novel method is presented to detect partial occlusion on people contours and recover occluded portions. Unlike other occlusion detection methods, the proposed method is only based on contours, which makes itself more flexible to be extended for further applications. Experiments with synthetic images demonstrate the accuracy of the method for detecting partial occlusions, and experiments on real-world video sequence are also carried out to prove that the method is also good enough to be used to recover target contours.

1. Introduction

In a visual tracking system, the accuracy of object detection and localization usually greatly affects the performance of tracking system and thus it has been one of the key links of the whole system. However, a moving target object is often occluded by background or other moving objects, which makes the target difficult to be detected and localized. Especially when the target is a deformable object [1, 2], it is more difficult to distinguish the occluded parts from the deformable parts [3]. There are a lot of efforts made in detecting and handling the occlusion problem in the literature [4–6]. Most of these methods adopt color [7, 8], appearance [9–11], texture [12], or motion [13] features as clues to detect occlusions. But these methods often tend to fail in detecting and localizing the target when the occluder has the same color, appearance, or texture as the occluded targets or when it has occluded the targets for a long time. However, using contour as a clue can overcome this shortage. For example, as shown in Figure 1, although the contours of occluder and occluded object are completely the same, the overlapped composite contour of two objects is obviously different from that of the individual contour.

In this paper a novel method based on contour is presented to detect the partial occlusions of the target and to recover the complete shape of the target in a tracking system. But the existence of noise and deformation on contour increases the difficulty of occlusion detection. Therefore, the practical method of this study is elaborately developed to avoid the influence of noise and deformation. Our method mainly consists of the following steps. First, the models of shape to be detected will be statistically constructed from several viewpoints. Second, the obtained contours of target are used to match the shape models. During the matching process, a morphing-detecting strategy is introduced to detect the occluded regions. After an initial matching by integral invariants [14], the contours are morphed from the models to the current contours. The partial occlusions can be detected by comparing the times of the break change of contours with a threshold. Third, the recovered shape can be obtained by shape registration (e.g., thin plate spline [15]) after excluding the detected occlusions. There are several advantages of this method: (i) with the help of the morphing process, it can overcome the insufficiency of information when the partial occlusion is detected only by object contours. (ii) The partial occlusion regions are

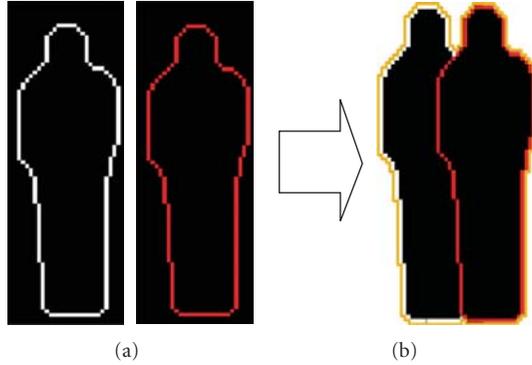


FIGURE 1: An example of composite contours. (a) is two similar contours of pedestrians. (b) is the composite contours. It is obvious that the contour in (b) is different from that in (a).

determined by the accumulated evidence during morphing process, which can improve the robustness and accuracy of the results. (iii) A new descriptor insensitive to noise is presented to precisely describe the change of object boundary.

1.1. Related Works. Since occlusion is a common but difficult problem in a tracking system, many efforts have been made to solve the problem by using a large variety of features of image [10, 16]. As one of the most popular features, color is often adopted as a clue. But when the tracked objects have similar color to background and other objects, it will cause ambiguity and result in tracking failure. Therefore, some extra assumptions have to be made in some methods that the color of target objects must be distinguished from other objects [7, 8]. Nguyen and Smeulders used appearance model for occluded object tracking. But it can only handle occlusions within a short time [10]. Yilmaz et al. used both the distance among the objects and the change of the object size to detect the occlusions [9]. Although it is an effective detecting strategy, only the coarse location of the occluded parts can be determined, which causes the more complex steps and less accurate result for the recovering of missing object parts. Pan and Hu proposed a method by detecting the nontarget region in ROI to handle the position of the occlusion by using of the spatiotemporal context and motion constraint [11]. This method enhanced the robustness under occlusions. However, the authors acknowledged that it still has some failure modes when the appearances of the moving objects are similar to that of occluder. As a kind of informative feature, texture is also used for occlusion detection. For example, Zhu et al. developed a method for objects detection and occlusion estimation by finding the corresponding parts with similar texture feature and spatial feature in consecutive frames [12]. More recently, Stein and Hebert [13] presented a local detection of occlusion boundaries by using motion as clues in short video clips. But when two objects have the same motions, failure will be inevitable.

Unlike these methods mentioned above, the proposed method uses only the contour of the targets to detect

occlusions and recover the exact contour of them. By using the proposed method, one can accurately localize the occluded parts in object contours without the help of any other features of image. Since this method does not need color, texture, and motion information, it will not result in failure when such information is ambiguous. It can also succeed in localizing a target occluded from a long time owing to the ability of recovering target shape directly from object contours.

The remainder of this paper is organized as follows. In Section 2, the building process of target models is described in detail. We give details of occlusions detection and contours recovering in Section 3. Experimental results and conclusions are sketched in Sections 4 and 5, respectively.

2. Building Target Models

In this study, our goal is to detect the occlusions on the target contours and then use them to exactly recover the whole target contours in the tracking system; in particular, we focus our efforts on occlusion detection and contour recovering of pedestrians. Since in videos the pedestrians may appear in different viewpoints, there is no uniform contour model which can be simply used to represent the contour exactly. Therefore, three sets of models have been built for representing a pedestrian from three different viewpoints, which are the front, the left, and the right models. These models are most typical and fundamental from which the contours from other viewpoints can be deformed.

First, we extract pedestrian contours in the training set from three different viewpoints. Each set of three-viewpoint contours has two postures, which are the standing posture and the walking posture, respectively. According to their viewpoints, these contours are aligned, respectively. The average contours are calculated from these aligned contours. The ultimate contours of models are obtained by smoothing these average contours. Because the contours viewed from the back are very similar to those viewed from the front, we assume that the models of them are also the same. The model of the left has the same posture as the model of the right but their directions are opposite. Therefore, three sets of models with six postures are constructed for a pedestrian. As shown in Figure 2, the first row is the subset of the training set and the second row is our six postures of the models.

Second, we compare a target contour obtained from a tracking system with each of the six models. We match the target contour to six models using the method in [14] to calculate the dissimilarities between the target and each of the six models. The model which has the minimal dissimilarity with the target contour is chosen as the matching model. Manay et al. introduced local area integral invariants to match shapes. A local area integral invariant is defined by the area bounded by a circle and the shape boundary. Because the local area integral invariant is much more robust than other invariants, it can obtain more accurate matching results, even if there exist the deformations on subparts, missing parts, or noise. Figure 3 shows an example of local area integral

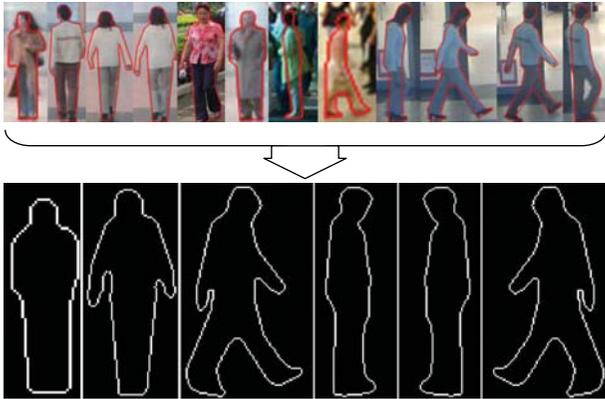


FIGURE 2: The models of a pedestrian. The first row is the subset of our training set. The second row shows the models. The two on the left are the models from the front viewpoint, the two in the middle are the models from the right viewpoint, and the two on the right are the models from the left viewpoint.

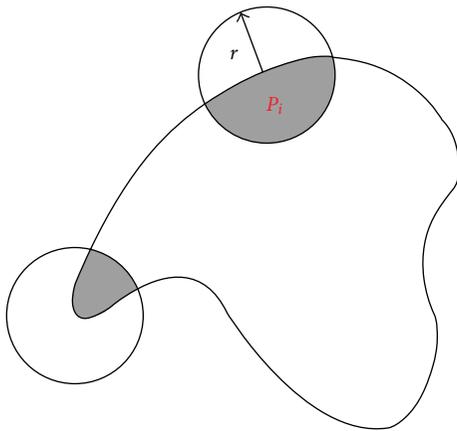


FIGURE 3: The local area integral invariant. The r is the radius of circle, and P_i is the centre. The gray region is the local area integral invariant which is bounded by the circle and object boundary.

invariant. In the proposed method, we adopt normalized local area integral invariant obtained by dividing it by the whole circle area as shape descriptor and use it to compute the dissimilarity between the target contours and models.

3. Occlusion Detection

Nowadays, there are many successful methods that can be used to segment the foreground from the background and obtain the target contours. Although these obtained contours may be not the exact object boundaries, the proposed method still can be used to correctly detect the occlusions and recover the target contours. Since we focus on the occlusion detection, to obtain the target contours will not be described in this paper. In this section, we will describe the occlusion detection in detail.

Given contours of the foreground, whether and where the contours are occluded should be concluded by comparing them with models. It is quite difficult to directly detect

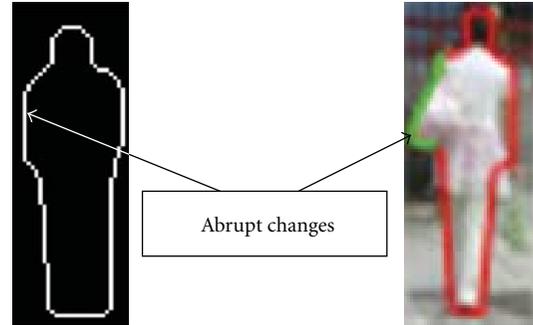


FIGURE 4: The changes caused by deformation and by occlusion. The red lines mean the gradual changes and the green lines are the abrupt changes.

the occlusion since the difference between the change of the contours and the occlusions is very ambiguous. But fortunately, inspired by the observation, in general we found that the changes in neighboring parts of a deformed boundary are gradual but the changes caused by occlusions are abrupt, as illustrated by red and green lines in Figure 4.

Since one part of an object is always connected with its neighbor parts, the change of this part will inevitably affect its neighbors by the action of internal force. The affection is tapering off along with the increasing distances between the changed part and its neighbors. Thus, it causes the gradual changes of contour. But when there are partial occlusions on object contours the gradual changes will be broken and abrupt changes will arise on the object contours of occluded parts relative to those of unoccluded parts. Note that the abrupt change mentioned here means continuously abrupt change, which must be distinguished from the change caused by articulation movement. When articulations move, only the contours around articulations will change, while the other parts will not change. Therefore, the abrupt change caused by articulation movement cannot last and will not be detected as occlusion by our method.

Hence, the occluded parts can be detected by comparing the changes between the model of object contours and corresponding parts of the target, and consequently the location of occlusions can be found according to the detected abrupt changes. An intuitive method is to match the target with the model and directly compute the changes of each pair of corresponding contours to find the abruptly changed parts. However, such a method is crude and sensitive to noise that it will lead to the incorrect results when there are mismatched parts. Therefore, for precisely detecting and localizing the occlusion parts, it needs a more effective and more robust strategy. The proposed method introduces a morphing process in which the model is gradually deformed to the target. At each deforming step, the model is changed slightly, and the influence by mismatching can be eliminated by smoothing deformed contours. And we accumulate the abrupt change times for each part of model during morphing process, which is insensitive to noise. In this way, the proposed method can completely overcome the shortages of the crude method mentioned above.

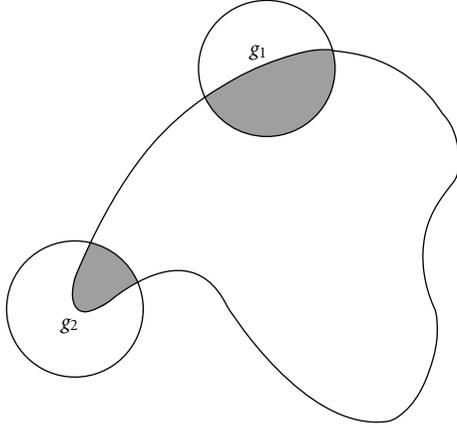


FIGURE 5: Use LAII to indicate the capability of shape preserving. The LAIs of all points in segment g_1 are similar, but those of all points in segment g_2 are different. Hence the capability of shape preserving of g_1 is larger than that of g_2 .

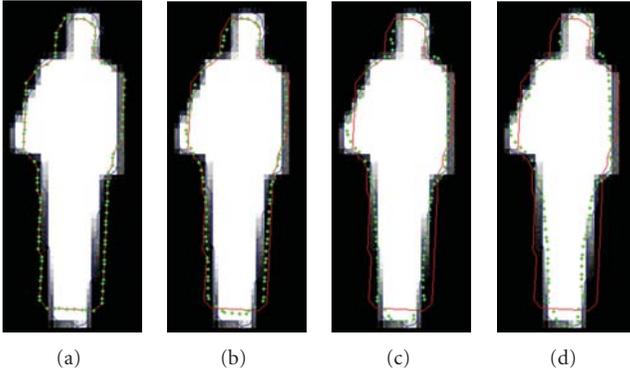


FIGURE 6: The intermediate contours. The blue shape is the contour of target object, and the red shape is the model. The green points mean the intermediate contours which are moving from model to target. (a) is the initial positions of the intermediate contours. (b) and (c) are the middle positions of the intermediate contours. (d) is the final positions.

Our method mainly consists of three stages. Firstly, after the contours of target are obtained by background subtraction (i.e., [17, 18]), the model is sampled as points set and a spring-mass model is constructed to denote a morphing process which deforms the model contour to the target step by step. Secondly, after the intermediate positions of points of the points set are obtained, the changes between two contours, which are obtained from the two adjacent steps, are calculated and then the abrupt changes can be detected accurately. Lastly, the decision of occlusion and its location will be concluded by accumulating evidence.

3.1. Construct the Spring-Mass Model. We firstly sample the matching model to form points set in which the distances between two neighbor points are equal along the contour. Denote the points set as $P = \{p_1, p_2, \dots, p_n\}$. Each point in P is treated as a mass, and the line connecting two adjacent

points is regarded as a spring. Then the spring stiffness which is assumed as a constant for each spring needs to be set. We define the term “self-similarity of segment” as the similarity between small neighbors of one point and large neighbors of this point. We then make the following assumption: if the self-similarity of segment of one point is larger, the capability of this point preserving its geometrical shape is stronger. This means that its change is expected to be small when the object contours are deformed. For example, there are two points p_1 and p_2 on two segments of object contours, g_1 and g_2 , respectively. As shown in Figure 5, two circles bounding the segments have the same radius. The curvatures of all points on g_1 are similar but those of g_2 are not, which makes the probability of the segment g_1 preserving its geometrical shape higher than that of g_2 when the object contours are deformed. Thus, the “stiffness” of one segment with higher capability of shape preserving is much larger than that of the segments with lower capability of shape preserving. Therefore, the sum of curvatures in the neighborhood of one point can be used to measure this segment “stiffness.” However, the curvature is quite sensitive to noise since it is a kind of derivative of contours. Hence, we introduce the local area integral invariant (LAI) to measure the shape-preserving capability of segment, as shown in Figure 5.

For each point in P , k LAIs, corresponding to k radiuses, r_1 to r_k with ascending sort, are calculated. The j th LAII of the i th point is denoted as A_{ij} . Thus the *self-similarity* of segment of shape in the neighborhood of p_i is calculated as follow:

$$C_i = \sum_{j=1}^{K-1} |A_{i,j} - A_{i,j+1}|. \quad (1)$$

Then the stiffness of spring between the i th and $(i + 1)$ th points can be defined as

$$s_i = \alpha \left(\frac{C_i + C_{i+1}}{2} \right), \quad (2)$$

where α is a constant for adjusting the stiffness which can be set to 1 in general. The spring stiffness in different parts of an object is also assumed to be unable to change a lot. Therefore some constraints must be added to avoid the large variation of the stiffness among the adjacent parts. Equation (3) is used to ensure the stiffness changes gradually:

$$s_i = \begin{cases} s_i & \frac{1}{t} \leq \frac{s_i}{s_{i-1}} < t, \\ ts_i & \frac{s_i}{s_{i-1}} < \frac{1}{t}, \\ \frac{s_i}{t} & \frac{s_i}{s_{i-1}} > t, \\ s_1 & i = 1, \end{cases} \quad (3)$$

where t is a constant used to restrict the maximal variation of stiffness between two adjacent parts (in our experiments, t is set to 1.1).

After obtaining these stiffness constants, we then calculate the external force and move the spring-mass model in the next subsection.

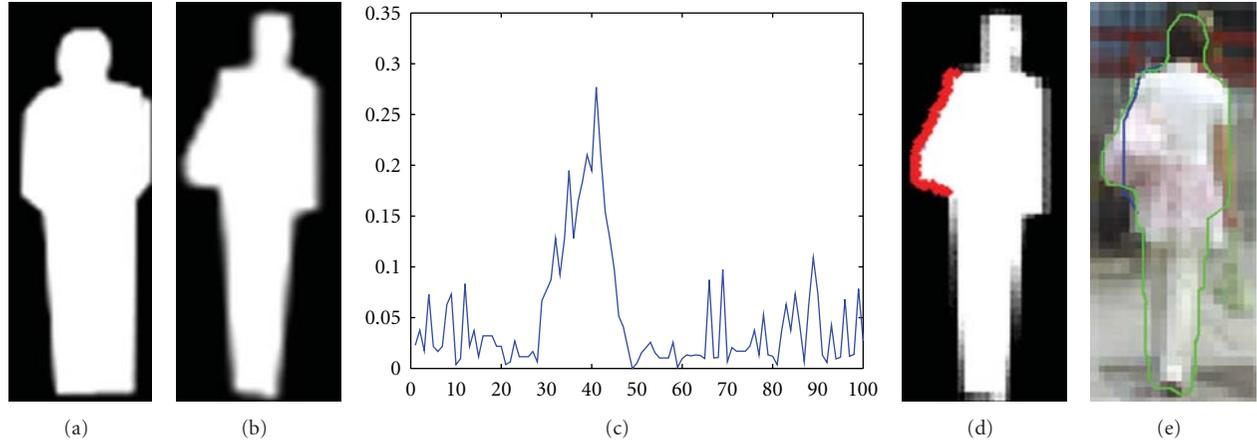


FIGURE 7: An example of detection and location of occlusion region. (a) and (b) are the template object and the target object, respectively. (c) plots the accumulating evidence of each point. There is obviously one continuous segment with high reliability which is corresponding to the occlusion region plotted by red line in (d). (e) is the result of shape recovering. The green line is the target contours with occlusions and the blue line is the recovered contours. Note that the blue line is visible in occluded parts and is invisible in normal parts as it is superposition with green line.

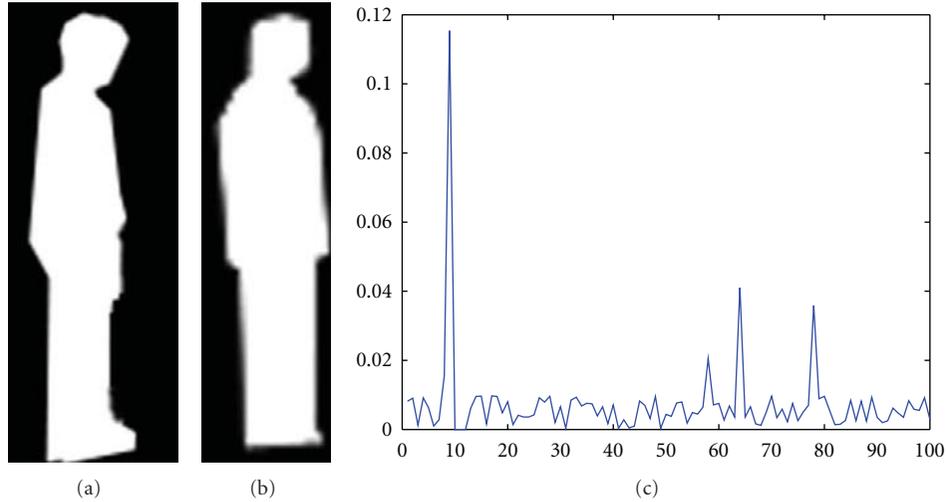


FIGURE 8: An example of detection and location of occlusion region. (a) and (b) are the template object and the target object, respectively. (c) plots the accumulating evidence of each point. Our method detects the target that has not been occluded since there are no continuous segments with high reliability.

3.2. *Obtain the Intermediate Contours.* For moving the spring-mass model, there must be external forces acting on the mass point. Now the method to calculate these external forces is illustrated in detail. After matching the target contours with object shape model by local area integral invariants, each point in point set P has a corresponding point in target contours. But there are some mismatching points which can be classified into two categories. In the first category, the mismatching points in the model are corresponding to the occluded parts of the target which make up the pseudocontour of the target. In the second category, the mismatching points belong to the parts of unoccluded contours. The main difference between the two categories is that the mismatching points in the first category are continuous since those points locating in the occluded parts are all mismatched but those in the second category

are not. The first category is used as a clue to detect the abrupt changes in our method. But the second category will affect the detected results. Therefore, we employ the thin plate spline to refine the matching result. We deform the model to the target according to the initially matching result. Denote the deformed model as P' . The refined matching result can be obtained by finding the nearest corresponding point q_j on the target for each point p'_i on P' . After refining, there are only few second category mismatching points. The disturbance of them will be eliminated during the morphing process by slightly smoothing the deformed model. The following equation is defined to calculate the external force acting on the i th point on template:

$$f_i = \beta \|q_j - p_i\| (q_j - p_i), \quad (4)$$

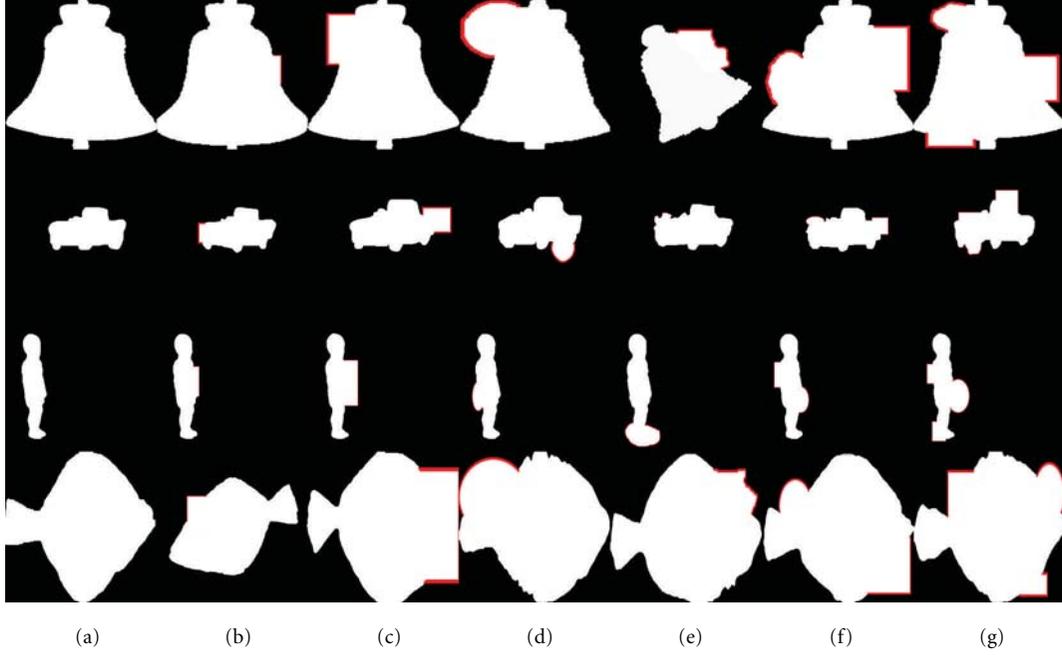


FIGURE 9: Some examples of occluded images based on MPEG-7 dataset. (a) is selected as the template image. (b), (c), (d), and (e) are the occluded images by single object under different occlusion region. (f) is the occluded image by two objects. Images in (g) are occluded by three objects. The red line indicates the occlusion region.

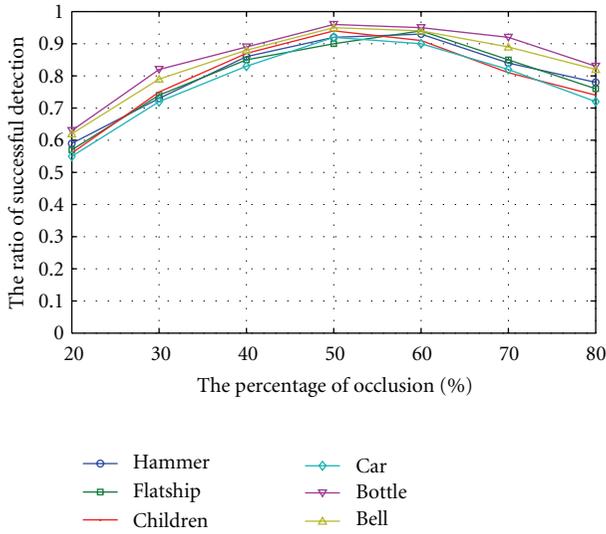


FIGURE 10: The average ratio of successful detection of occlusion region of each class image.

where the magnitude of f_i is determined by $\beta \|q_j - p_i\|$, and the direction is determined by $(q_j - p_i)$. β is a constant for all points which controls the magnitude of external force (in our experiments it is set to 0.1).

The next statement will explain the movement principle of the model under the action of the external force. The movement of each point in P is determined by three forces: the external force, the internal force produced by the adjacent

spring, and the friction force. Mathematically, the principle of movement can be expressed as

$$F_i^{\text{ext}} - F_i^{\text{int}} - c v_i = m_i a_i, \quad (5)$$

where F_i^{ext} is the external force, calculated by (4), and F_i^{int} is the internal force which is formulated in (7). The friction force $c v_i$ is proportioned to the velocity of point p_i , in which c is a constant. Denote l_i as the displacement of p_i and $v_i = dl_i/dt$ as the velocity of point p_i . $a_i = d^2 l_i/dt^2$ is the acceleration of this point. Then (5) is converted to the following equation:

$$F_i^{\text{ext}} - F_i^{\text{int}} - c \frac{dl_i}{dt} - m_i \frac{d^2 l_i}{dt^2} = 0. \quad (6)$$

In our spring-mass model, each point is connected with two springs. Therefore, the internal force can be formulated as

$$F_{i1}^{\text{int}} = s_{i-1} \left(\frac{\|p_i - p_{i-1}\|}{\|p'_i - p'_{i-1}\|} - 1 \right) (p'_i - p'_{i-1}),$$

$$F_{i2}^{\text{int}} = s_i \left(\frac{\|p_i - p_{i+1}\|}{\|p'_i - p'_{i+1}\|} - 1 \right) (p'_i - p'_{i+1}), \quad (7)$$

$$F_i^{\text{int}} = F_{i1}^{\text{int}} + F_{i2}^{\text{int}}.$$

Here p'_{i-1} , p'_i , and p'_{i+1} are the instantaneous positions corresponding to p_{i-1} , p_i , and p_{i+1} at current time, respectively. Equation (6) is an ordinary differential equation. By

solving it, the displacement (also new position), velocity, and acceleration of each point at each time can be obtained. The intermediate contours at any time then consist of the positions of these moved points at the corresponding time. Because the spring-mass model will not automatically stop until the sum of external force is equal to zero, we manually stop it when the distance between the moving contours and the target contours reaches its minimum. Figure 6 illustrated the process. As mentioned above, there are still a few second class mismatching points after refining processing which will influence the morphing process. To deal with it, Gaussian convolution is introduced to smooth the intermediate contours after the new position of each point is obtained.

3.3. Detect and Accumulate the Abrupt Changes. After obtaining the intermediate contours, we then compute e_i , the change of the i th point between two adjacent intermediate contours, P' and P'' :

$$e_i = |d'_i - d'_i| |\theta'_i - \theta'_i|, \quad (8)$$

where d' is the distance between p'_i and p'_{i+1} which are the i th and $(i+1)$ th points in P' . θ'_i stands for the angle between p'_i and p'_{i+1} . d'_i and θ'_i are the corresponding distance and angle in P'' . Equation (8) measures the change by multiplying the difference of distance and that of angle. Since e_i is computed from the difference of distance and the difference of angle, this kind of measurement is invariant to scale, rotation, and translation. But it may be sensitive to noise. We modify (8) to be more robust, and then we have the following equation:

$$e'_i = \frac{\sum_{j=i-k}^{i+k} e_j \cdot \exp(e_j - \bar{e})}{2k+1}, \quad (9)$$

where $\bar{e} = (1/N) \sum_{i=1}^N e_i$ is the average change of all points. In (9), when e_j is larger (smaller) than \bar{e} , the value of e_j will be enlarged (lessened). Thus, it will amplify the large change and suspend the small change and lead to more robust results to detect the abrupt changes. And using the average change of $2k+1$ points, which are k former and latter points of the i th point, respectively, can further improve its robustness. There are $h+1$ e'_i computed from the model contours, h intermediate contours, and one target contour. And then the amount of abrupt change times for each point is accumulated:

$$v_i = \sum_{i=1}^{h+1} f(e'_i), \quad (10)$$

where $f(e'_i)$ is an indicator function of the abrupt change:

$$f(e'_i) = \begin{cases} 1 & e'_i - T_s \geq 0, \\ 0 & e'_i - T_s < 0, \end{cases} \quad (11)$$

where T_s is a threshold to indicate that the change is abrupt or gradual (in our experiments T_s is set to $\mu + 2\sigma$, where μ is the average of e'_i and σ is the variance of it [19]). If v_i is

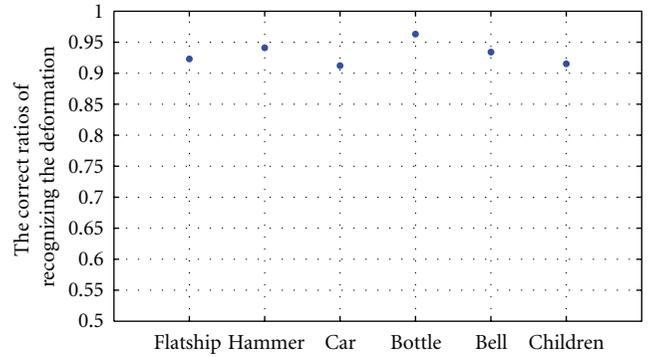


FIGURE 11: The correct ratios of recognizing the deformation.

larger than a threshold T_v (in our experiments T_v is set to 2μ , and μ is the average of v_i), the point p_i may be in an occluded region. Note that the occluded region must have a certain extent. Therefore, if one point is in an occluded region, a certain number of neighbor points must also be in the same occluded region. Thus, we can set another threshold T_n (in our experiments T_n is set to $n/10$, where n is the number of points in template point set) which indicates the least number of points in an occluded region. Finally, if the amount of sequence points all of which v_i are large than T_v is not less than T_n , the sequence points indicate an occluded region (see Figure 7 as an example). If there is not any sequence points all of which v_i are larger than T_v is larger than T_n , the target contour is an deformed contours without occlusions. An example is illustrated in Figure 8.

3.4. Recover Target Contours. After detecting the occlusion regions, the target contour then can be recovered conveniently. The parts of model corresponding to the parts of target contours detected as occluded regions are excluded at first. We then sample a certain number of points from the target contours excluded occlusions and the corresponding points in the shape model as anchor points. The thin plate spline is employed to warp the shape model to target contours. Figure 7(e) illustrated an example of contours recovering. For more accurately recovering, we can use more complex methods [20, 21]. For example, the shape context [22] is an excellent shape register method based on a sample point set which can increase the accuracy of the recovered shape. The integral invariant is also an excellent shape matching and register method, but in our method we do not employ it to recover the target contours since it needs the closed contours.

4. Experiments

We firstly evaluate the performance of occlusion detection method on an occluded version of MPEG-7 dataset [23] in which we added occlusion regions of different sizes and numbers to the images (Figure 9 shows some examples).

We select six classes of images as testing images, which are the bell, the car, the children, the flatship, the bottle, and the hammer. From each class one image is chosen as

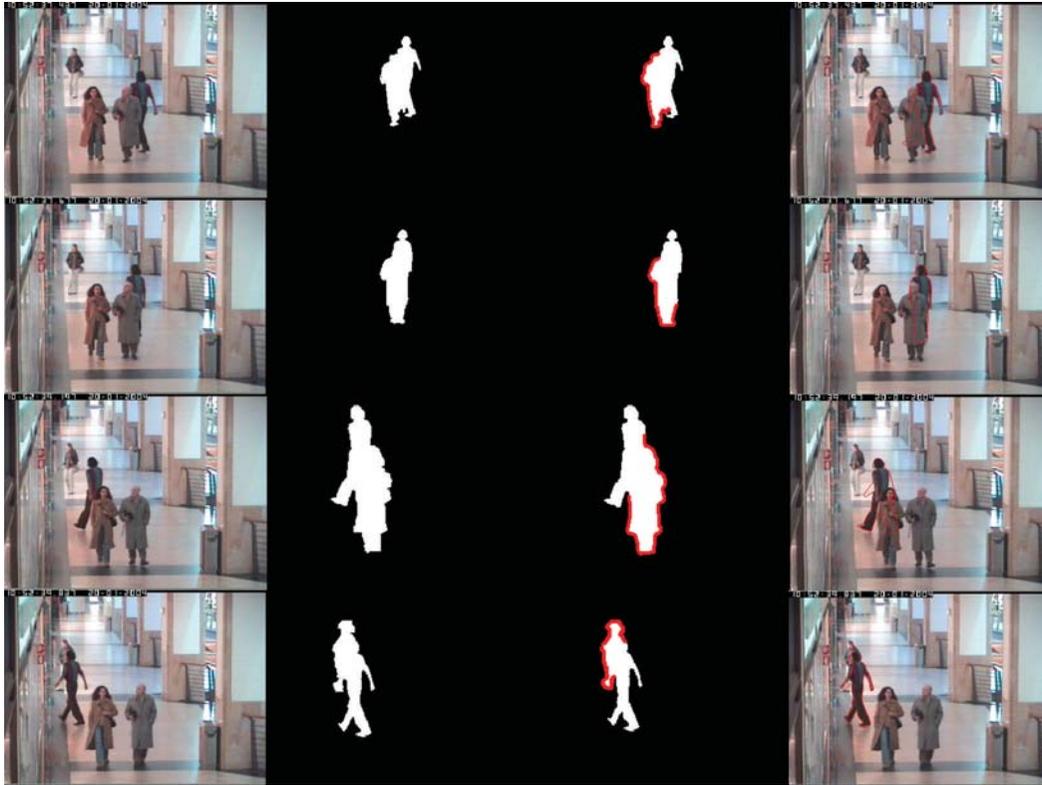


FIGURE 12: Examples of occlusion detection and target shape recovering in real-world video sequences. The frame numbers from top to bottom are 1983, 1989, 2027, and 2043, respectively. The columns from left to right are original frames, foreground contours, results of occlusion detection, and results of shape recovering, respectively. In the second column, we have eliminated other foreground contours which are not connected to our target for clearer display. Note that the results of occlusion detection are quite exact but the result of shape recovering in the third row is failure.

the template (see Figure 9(a)), and the other images are occluded by shapes of different sizes and numbers. For each image in a class, the object boundary is occluded by one rectangle, ellipse, and random shape, respectively, and the occlusion region in contour ranges from 20 to 80 percent (see Figures 9(b), 9(c), 9(d), and 9(e)). Each image is also occluded by two or three shapes simultaneously. It brings about multiocclusion regions, and the amount of occlusion contour also ranged from 20 to 80 percent (see Figures 9(f) and 9(g)). After obtaining the occlusion parts detected by our method, we compare them with the actual occlusion regions. If the overlapped parts are not less than 90 percent and not more than 110 percent, we believe the detection is successful. Otherwise the detection fails. Figure 10 shows the average ratio of successful detection of all six classes' images. When the occlusion region is too small, the ratio of successful detection is low since it is rather ambiguous between the small-scale partial occlusions and deformation. But with the occlusion regions increasing, the ratio of successful detection is also mounting up until 60 percent. When the occlusion region is more than 60 percent, the successful ratio will decrease from the highest ratio since the normal boundary is too small to distinguish the occlusion part. But we can regard the occluder as the model which is occluded by target, by which the new occlusion region is less than 60 percentage and the high ratio of successful detection can be obtained.

Because it is quite easy to confuse the partial occluded contours with deformed contours, the performance of the proposed method to correctly recognize the deformed object contours must also be evaluated. By choosing one image as model in a class, our method is used to detect other normal images in the same class. If there is no occlusion region after the detection, it shows that our method can correctly recognize the deformed object boundary. Figure 11 illustrates these correct ratios of the six classes' objects.

It can be seen from Figure 11 that the bottle class has the highest recognition ratio since its shape is the simplest and the degree of deformation is the smallest. Both the children class and car class have the relative lower recognition ratio since their shapes are more complicated. Note that in Figure 10, these two classes also gain the relative lower ratios of successful detection. But a complex object can be decomposed into the combination of simple object by some shape decomposition method (i.e., [24, 25]) to improve the performance of our method.

Next, we apply the proposed method to detect occlusions and recover target shape in a tracking system. Figure 12 shows the occlusion detection results and contours recovering results of the left viewpoint model, respectively. In Figure 12, each row corresponds to one frame in real-world video. From left to right, the columns are the original images, foreground contours around the target, the

occlusion contours detected by the proposed method, and the recovered contours of target. We extracted the foreground by using background subtraction and neglected those too small patches. All six models of pedestrian are matched to the contours of each disconnected foreground region for computing the dissimilarities. The model corresponding to the minimal dissimilarity is chosen for detecting the occlusions and recovering target contours. If the minimal dissimilarity is larger than a threshold, we believe that this foreground region does not contain the target and ignore it before we match the next foreground region.

As shown in Figure 12, the results of occlusion detection are accurate since the percentage of occluded parts is just between 30% and 70%. This is in line with the previous experiments. The results of shape recovering are not as accurate as occlusion detection since we just use the TPS method to recover target shapes. To improve this, one can employ some more complicated methods (i.e., shape context [22]).

5. Conclusions and Discussions

The main contribution of this paper is that a novel method is presented to detect partial occlusion based on target contours. With the help of this method, satisfactory performances in both synthetic and real-world image sequences are achieved for object tracking and target contours recovering. In our method contour matching is taken as a morphing process from which sufficient information is obtained to detect the partially occluded regions and the influence of mismatching and noise can also be overcome. The method only needs object contours, which can be used in the situation where color, texture, and motion information cannot be obtained exactly.

However, there are still some further works for improving the performance of algorithm. As mentioned in Section 4, a relatively low detection ratio was observed when the occlusion region is small due to the extreme ambiguity between the deformed contours and small-scale occluded contours. When the object to be detected is complex, the method needs to be further enhanced, for example, by employing some shape decomposition methods to decompose a contour to a certain number of simple objects and then detect them separately. During the processing of target contours recovering, we simply use the thin-plate spline to register the model to target. To improve this, some more complex methods can be adopted.

Acknowledgments

This work was supported by the National Natural Science Foundation of China and Microsoft Research Asia (60870002) and Zhejiang Provincial Natural Science Foundation (R1110679, Y1110688).

References

- [1] 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.
- [2] S. Y. Chen, "Cardiac deformation mechanics from 4D images," *Electronics Letters*, vol. 43, no. 11, pp. 609–611, 2007.
- [3] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: a survey," *ACM Computing Surveys*, vol. 38, no. 4, p. 13, 2006.
- [4] E. Lobaton, R. Vasudevan, R. Bajcsy, and R. Alterovitz, "Local occlusion detection under deformations using topological invariants," in *Proceedings of the European Conference on Computer Vision*, pp. 101–114, 2010.
- [5] A. Ayvaci, M. Raptis, and S. Soatto, "Occlusion detection and motion estimation with convex optimization," in *Proceedings of Advances in Neural Information Processing Systems*, 2010.
- [6] S. Zhang, J. Wu, Y. Tian, S. Liu, and X. Sun, "Robust visual tracking based on occlusion detection and particle redistribution," in *Proceedings of the 2nd International Conference on Internet Multimedia Computing and Service*, 2010.
- [7] N. Amézquita, R. Alquézar, and F. Serratosa, "Dealing with occlusion in a probabilistic object tracking method," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–8, June 2008.
- [8] T. Yang, S. Z. Li, Q. Pan, and J. Li, "Real-time multiple objects tracking with occlusion handling in dynamic scenes," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 970–975, June 2005.
- [9] A. Yilmaz, X. Li, and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1531–1536, 2004.
- [10] H. T. Nguyen and A. W. M. Smeulders, "Fast occluded object tracking by a robust appearance filter," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1099–1104, 2004.
- [11] J. Pan and B. Hu, "Robust occlusion handling in object tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2007.
- [12] L. Zhu, J. Zhou, and J. Song, "Tracking multiple objects through occlusion with online sampling and position estimation," *Pattern Recognition*, vol. 41, no. 8, pp. 2447–2460, 2008.
- [13] A. N. Stein and M. Hebert, "Local detection of occlusion boundaries in video," *Image and Vision Computing*, vol. 27, no. 5, pp. 514–522, 2009.
- [14] S. Manay, D. Cremers, B. W. Hong, A. J. Yezzi, and S. Soatto, "Integral invariants for shape matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1602–1617, 2006.
- [15] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. v, pp. 567–585, 1992.
- [16] B. J. Matuszewski, J. K. Shen, and L. K. Shark, "Elastic image matching with embedded rigid structures using spring-mass system," in *Proceedings of the International Conference on Image Processing*, pp. 937–940, September 2003.
- [17] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 246–252, June 1999.

- [18] Z. Zivkovic and F. Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006.
- [19] H. J. Zhang, J. H. Wu, and S. W. Smoliar, "System for automatic video segmentation and key frame extraction for video sequences having both sharp and gradual transitions," US Patent 5635982, 1997.
- [20] S. Y. Chen, W. Huang, and C. Cattani, "Traffic dynamics on complex networks: a survey," *Mathematical Problems in Engineering*. In press.
- [21] M. Li, C. Cattani, and S. Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 13 pages, 2011.
- [22] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.
- [23] L. J. Latecki, R. Lakaemper, and U. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 424–429, June 2000.
- [24] K. Koara, A. Nishikawa, and F. Miyazaki, "Hierarchical part decomposition method of articulated body contour, and its application to human body motion measurement," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, pp. 2055–2061, November 2000.
- [25] J. M. Lient, J. Keyser, and N. M. Amato, "Simultaneous shape decomposition and skeletonization," in *Proceedings of the ACM Symposium on Solid and Physical Modeling*, pp. 219–228, June 2005.

Research Article

An Adaptive Coding Pass Scanning Algorithm for Optimal Rate Control in Biomedical Images

Hsi-Chin Hsin,¹ Tze-Yun Sung,² and Yaw-Shih Shieh²

¹Department of Computer Science and Information Engineering, National United University, Miaoli City 36003, Taiwan

²Department of Electronics Engineering, Chung Hua University, Hsinchu City 30012, Taiwan

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

Received 13 June 2011; Accepted 15 August 2011

Academic Editor: Shengyong Chen

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

High-efficiency, high-quality biomedical image compression is desirable especially for the telemedicine applications. This paper presents an adaptive coding pass scanning (ACPS) algorithm for optimal rate control. It can identify the significant portions of an image and discard insignificant ones as early as possible. As a result, waste of computational power and memory space can be avoided. We replace the benchmark algorithm known as postcompression rate distortion (PCRD) by ACPS. Experimental results show that ACPS is preferable to PCRD in terms of the rate distortion curve and computation time.

1. Introduction

With the rapid growth of modern telemedicine techniques, high-efficiency image-processing algorithms are essentially in great demand [1, 2]. Wavelet transform (WT) [3–5] provides many desirable properties that are beneficial to the biomedical image applications. In general, the significant wavelet coefficients of an image are clustered within each subband; this property is the main idea behind the embedded block coding with optimized truncation (EBCOT) algorithm, which is the core of JPEG2000 [6, 7].

EBCOT adopts the post compression rate distortion (PCRD) algorithm for optimal rate control, which demands a large amount of memory space and leads to waste of computational power [8]. Du et al. utilized a fixed scan order to reduce the computation time [9], which however lacks adaptation. Vikram et al. proposed a scheme to estimate the rate distortion slopes of code blocks for adaptive rate control [10]. Fang et al. proposed a precompression scheme to avoid unneeded computations [11]. An et al. proposed a fractional upward shift for lossless image compression [12]. Auli-Llinas et al. proposed nonuniform adaptation for bit plane coding [13]. Balster et al. proposed inter-based adaptation to estimate the distortion value [14]. Xue et al. distributed the target bit rates adaptively to reduce the computational complexity [15].

This paper proposes an adaptive coding pass scanning (ACPS) algorithm to rearrange the coding passes of an image such that the most significant one has priority over others. In comparison with the complicated approach used in [10–15], which is either time consuming or hardware demanding, ACPS makes good use of the MQ table that is available at both encoder and decoder. As a result, there is no need to store and transmit any information about the contribution of coding passes, and it has the advantage of saving a lot of computation time, memory space and transmission time, which is preferable for many applications [16–20].

The remainder of this paper proceeds as follows. In Section 2, EBCOT is reviewed briefly. Section 3 describes the proposed ACPS algorithm. Experimental results are presented in Section 4. Conclusion is given in Section 5.

2. Review of EBCOT

The embedded block coding with optimized truncation (EBCOT) algorithm adopts wavelet transform for subband decomposition. It is a two-tier algorithm. Tier-1 performs bit plane coding (BPC) followed by arithmetic coding (AC). Tier-2 aims for optimal rate control. Figure 1 depicts the block diagram of EBCOT encoder. Three coding passes,

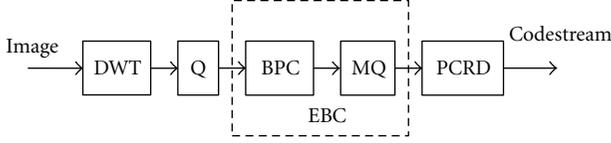


FIGURE 1: Block diagram of the EBCOT encoder.

namely the significance propagation (SP) pass, the magnitude refinement (MR) pass and the cleanup (CU) pass are defined in EBCOT. For a wavelet coefficient that is currently insignificant, if any of the 8 neighboring coefficients are significant, it is coded in the SP pass using the significance coding operation; otherwise, it is coded in the CU pass using the cleanup coding operation. If the coefficient becomes significant, its sign is then coded using the sign coding operation. The significant coefficients are recursively updated using the magnitude refinement coding operation in the MR pass. The bit streams of the coding passes can be further compressed by AC with the probability models stored in the MQ table.

EBCOT applies the postcompression rate distortion (PCRD) algorithm for optimal rate control. Specifically, let $\{B_i\}$ be the code blocks of an image. B_i is coded from the most significant bit plane to the least significant bit plane, and terminates at some point, says n_i , resulting in a bit rate denoted by $R_i^{n_i}$. All of 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 optimal truncation points by minimizing the total distortion given by

$$D = \sum_i D_i^{n_i}, \quad (1)$$

with the following rate constraint:

$$R = \sum_i R_i^{n_i} \leq R_c, \quad (2)$$

where R_c is the target bit rate. As all of the code blocks including insignificant ones must be coded before PCRD, the drawbacks are wastes of computational power, execution time, and memory space.

3. Image Coding with Adaptive Coding Pass Scanning

For the telemedicine applications, the following question is often raised. Is there useful information available at both encoder and decoder based on which the contribution of coding passes can be obtained? If so, there is no need to transmit the scanning order of coding passes, from encoder to decoder, for optimal rate control. Motivated by the MQ coder, we propose a novel scheme to estimate the rate distortion slope (RDS) of coding passes. Only the coding passes with significant RDS are used to generate the final code stream, and the unneeded ones with insignificant RDS can be discarded before EBCOT tier-1. Thus, it has the

advantage of saving computational power, memory space, and transmission time.

3.1. Adaptive Coding Pass Scanning Algorithm. In [16], we proposed the CBRDE algorithm to arrange the code blocks of an image adaptively. Specifically, let $b_{ij}(m, n)$ be a binary function of position (m, n) in the i th code block, at the j th bit plane, and $B_{ij} = \bigcup_{m,n} \{b_{ij}(m, n)\}$. The RDS of B_{ij} is defined as

$$S_{ij} = \frac{E[D_{ij}]}{E[R_{ij}]},$$

$$\text{where } E[D_{ij}] = \sum_{m,n} \text{prob}(b_{ij}(m, n) = 1),$$

$$E[R_{ij}] = \sum_{m,n} H(b_{ij}(m, n)), \quad (3)$$

$$H(b) = -\text{prob}(b = 1) \cdot \log_2(\text{prob}(b = 1))$$

$$-\text{prob}(b = 0) \cdot \log_2(\text{prob}(b = 0)),$$

$$\text{prob}(b_{ij}(m, n) = 1) = \begin{cases} Qe_Value, & \text{if MPS} = 0, \\ 1 - Qe_Value, & \text{if MPS} = 1, \end{cases}$$

D_{ij} and R_{ij} are the distortion decreased and the coding bits increased for B_{ij} , respectively, $E(\circ)$ is the expectation operation, $H(\circ)$ is the entropy operation, and Qe_Value is the probability of less probable symbol (LPS), which can be obtained from the MQ table directly.

At each bit plane, each code block is coded with a sequence of three coding passes, that is, SP followed by MR and CU. All of the coding passes are possible truncation points. It is noted that more than 75% of the coding passes of an image are optimal truncation points, and more than 97% of which are either consecutive or one coding pass away from each other [21]. Motivated by the above, the adaptive coding pass scanning (ACPS) algorithm is thus proposed. As shown in Figure 2, if the RDS of the MR pass is greater than that of the SP pass, the MR pass should be performed immediately after the SP pass according to the PCRD algorithm, and the optimal segment should contain the SP and MR passes. However, unlike the CU pass, the MR pass may take place before the SP pass at each bit plane. In order to improve the compression performance, the above example is modified as follows. The scan order of SP and MR passes is rearranged such that the one with greater RDS should be coded first, which leads to the ACPS algorithm. Figure 3 depicts the flowchart of ACPS, in which CBRDE is used to estimate the RDS of coding passes.

3.2. Image Coder with ACPS. As a large amount of energy is in the LL subband, it should be coded first. For the coding passes of the non-LL subbands, the one with the greatest RDS should be coded as early as possible. Hence, we propose an image coder with ACPS. Block diagrams of the proposed encoder and decoder are shown in Figures 4 and 5, respectively. The encoding steps are as follows.

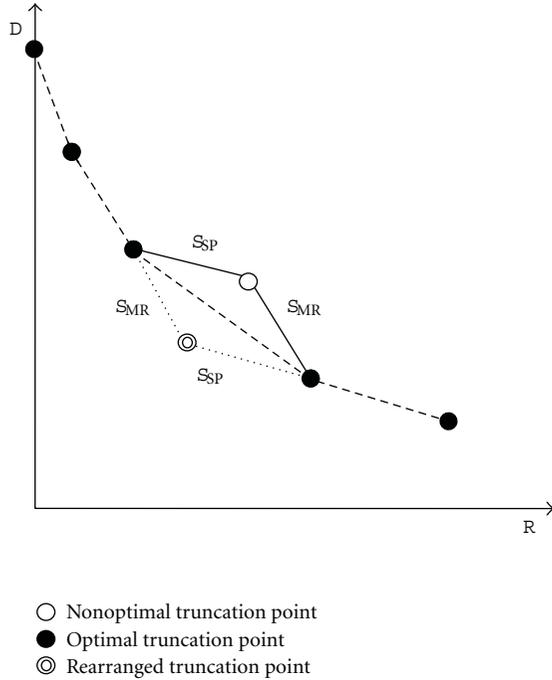


FIGURE 2: RD curves formed by nonoptimal truncation point (empty circle), optimal truncation points (dark circles), and rearranged truncation point (double-ring circle) S_{SP} and S_{MR} are the RDS of the SP and MR passes in a code block.

Step 1. Initialization. Decompose an image into wavelet subbands, quantize the transform coefficients, and divide each subband into code blocks. Initially, the fixed scan order [21] is used. More specifically, wavelet subbands are scanned in a zigzag order, from the lowest frequency subband to the highest frequency subband, the code blocks within each subband are scanned from left to right, top to bottom, the first coding pass for each code block is the CU pass, and the scan order at the next bit plane is the SP pass followed by the MR pass and the CU pass.

Step 2. Estimation of RDS. At the current bit plane, estimate the RDS of coding passes by CBRDE. ACPS is then used to rearrange the scan order of coding passes.

Step 3. Adaptive Scan of Coding Passes. As the LL subband contains lots of significant coefficients, the SP and MR passes in the LL subband are a top priority. All the SP and MR passes in the non-LL subbands have priority over the CU passes at the current bit plane.

Step 4. EBCOT tier-1 Encoding. After rearranging the coding passes at the current bit plane, EBCOT tier-1 is used to produce the code streams that can be transmitted to decoder immediately. For the next bit plane, go to step 2, and continue the steps in order. Terminate at any point while the accumulated bit rate reaches the target bit rate.

To show the potential of ACPS, the following correct rate is utilized. Let the SP, MR, and CU passes in the i th code block at the j th bit plane be represented as C_{ij}^0 , C_{ij}^1 , and C_{ij}^2 ,

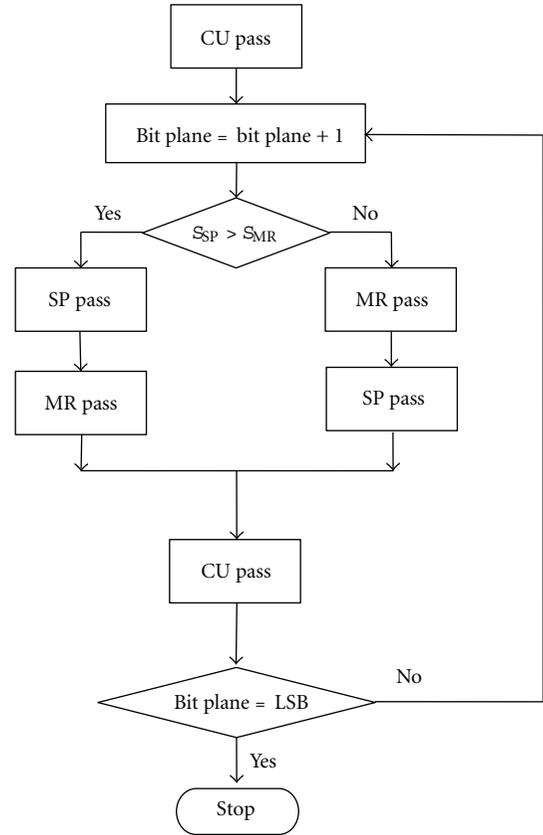


FIGURE 3: Flowchart of the proposed ACPS algorithm.

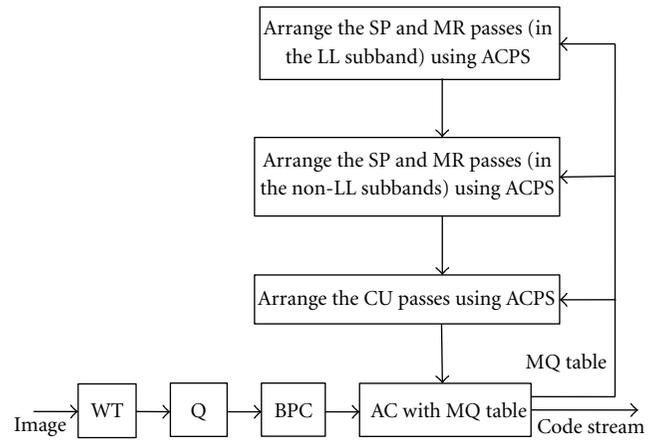


FIGURE 4: Block diagram of the proposed image encoder with ACPS.

respectively. $\hat{P}(C_{ij}^k)$ and $P(C_{ij}^k)$ denote the location indices of C_{ij}^k , $k \in \{0, 1, 2\}$ for the estimated scan order and the true scan order, respectively. We measure the correct rate of ACPS below:

$$CR_j = \left(1 - \frac{1}{N_j^2} \sum_i \sum_{k=0}^2 \left| \hat{P}(C_{ij}^k) - P(C_{ij}^k) \right| \right) \times 100\%, \quad (4)$$

where N_j is the total number of coding passes at the j th bit plane. The results of images shown in Figure 6 are given in

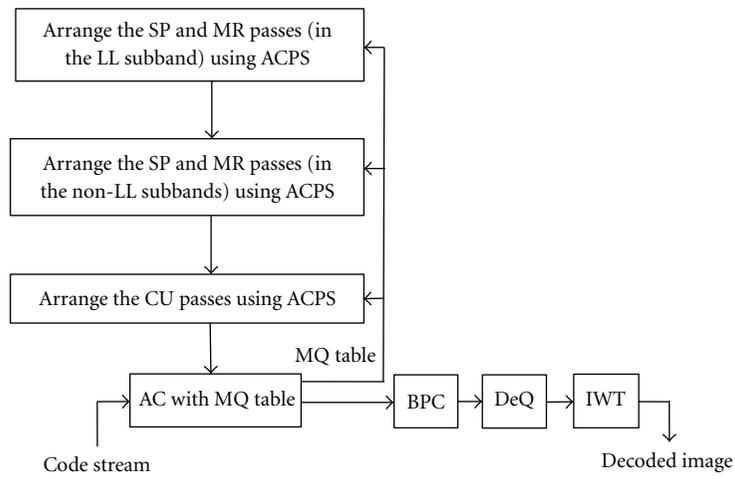


FIGURE 5: Block diagram of the proposed image decoder with ACPS.

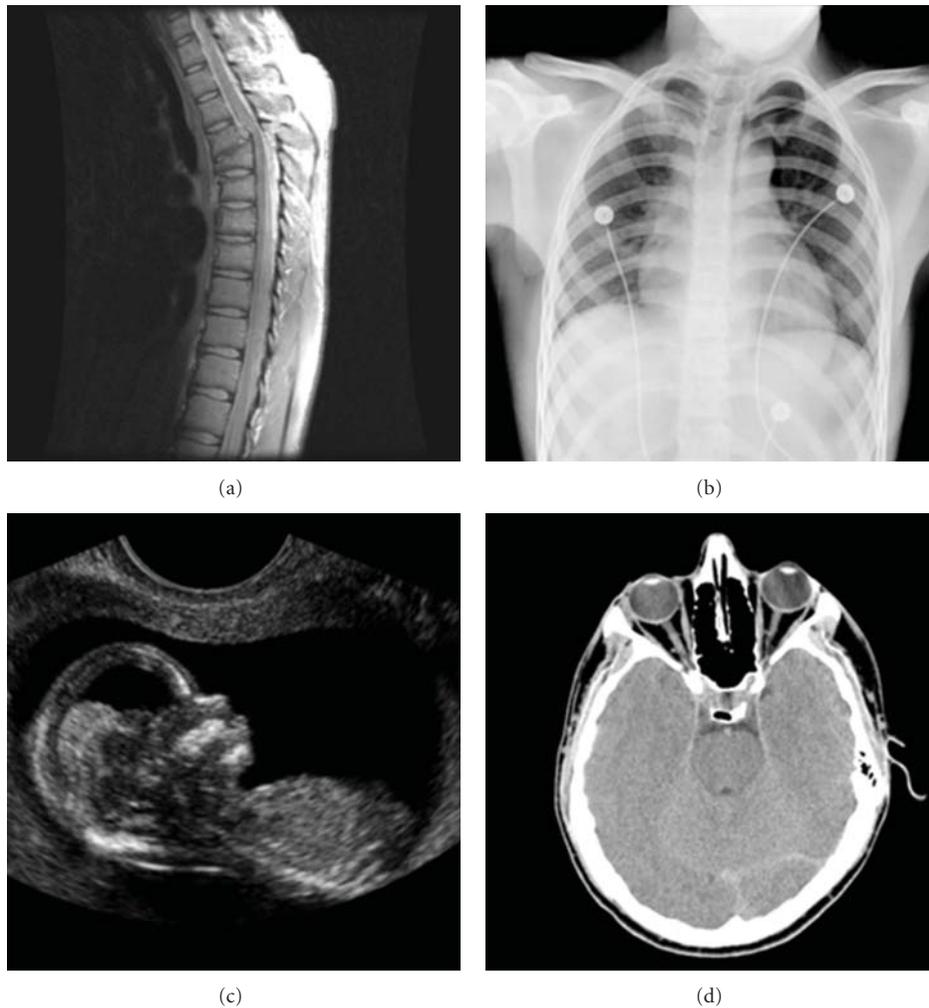


FIGURE 6: Test images; (a) Spine (MRI), (b) Chest (X-ray), (c) Fetus (ultrasonic), and (d) Head (CT).

TABLE 1: The correct rates of ACPS.

| Image | Spine | Chest | Fetus | Head |
|-------------|-------|-------|-------|-------|
| Bit-plane 1 | 100% | 100% | 100% | 100% |
| Bit-plane 2 | 100% | 100% | 100% | 100% |
| Bit-plane 3 | 66.7% | 66.7% | 66.7% | 63.7% |
| Bit-plane 4 | 70.3% | 70.3% | 74.5% | 78.5% |
| Bit-plane 5 | 85.1% | 71.9% | 79.2% | 69.5% |
| Bit-plane 6 | 70.4% | 74.9% | 78.5% | 73.9% |
| Bit-plane 7 | 73.4% | 78.3% | 77.9% | 76.7% |
| Bit-plane 8 | 79.2% | 74.3% | 73.1% | 77.9% |

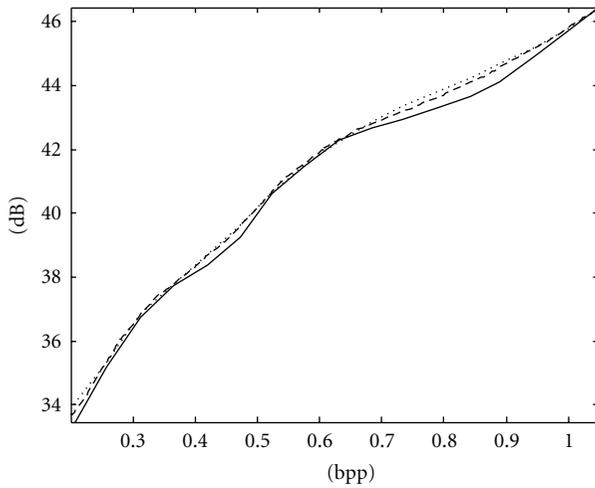


FIGURE 7: RD curves of Spine MRI image using the fixed scan order [21] (solid line), PCRD (dashed line), and ACPS (dotted line).

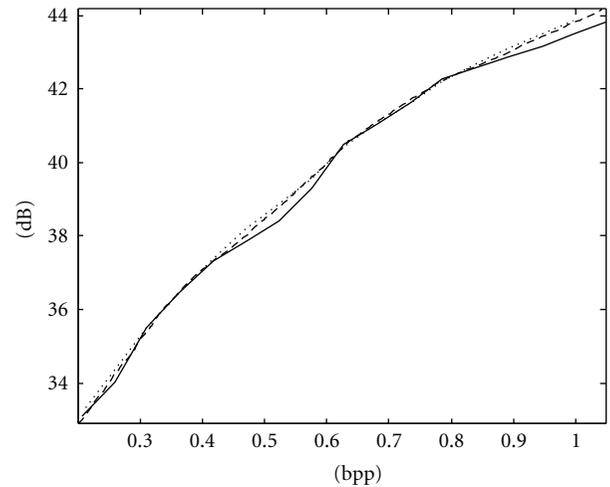


FIGURE 8: RD curves of Chest X-ray image using the fixed scan order [21] (solid line), PCRD (dashed line), and ACPS (dotted line).

Table 1. It is noted that the correct rates for the first two bit planes are extremely high due to the suitable initialization using [21]. For the first few bit planes after the non-LL subbands become significant, the correct rates are less than 70%, because the MQ table is not quite reliable due to only a small number of significant coefficients appearing. However, the correct rates become higher as the MQ table becomes more reliable. The overall correct rate is approximately 80%. It justifies the soundness of ACPS.

4. Experimental Results in Biomedical Image Processing

The proposed image coder with ACPS has been evaluated on various biomedical images. The experimental results of biomedical images shown in Figure 6, namely, spine, chest, fetus, and head, are given in this paper, which are representative of magnetic resonance image (MRI), X-ray, ultrasound, and computed tomography (CT) images, respectively. The 9/7-wavelet filters adopted by the JPEG2000 standard are used. The distortion is defined as the peak-signal-to-noise ratio (PSNR), together with the rate of bits per pixel (bpp) forms the rate distortion (RD) curve.

The RD curves of Spine and Chest images using ACPS and PCRD are comparable and more convex than that using the fixed scan order [21], as shown in Figures 7 and 8, respectively, where the horizontal and vertical axes are the bpp rates and PSNR values. For fetus and head images, ACPS is preferable to both PCRD and the fixed scan order, as shown in Figures 9 and 10, respectively.

Since ACPS takes the place of PCRD, lots of insignificant coding passes can be discarded as early as possible, waste of computation time and memory space can be avoided, and a great improvement in speed results. As shown in Table 2, the speed-up ranges from 30% to 50%.

5. Conclusion

In EBCOT, the code blocks of an image are independently coded with three coding passes, that is, SP followed by MR and CU passes. Any coding passes are possible truncation points, based on which, optimal truncation points can be obtained by using the PCRD algorithm. It requires that all the coding passes are coded and stored in the tier-1 of EBCOT; however, lots of insignificant coding passes are not needed, which in turn are discarded after PCRD. Instead of computing the true RDS, ACPS makes use of the estimated

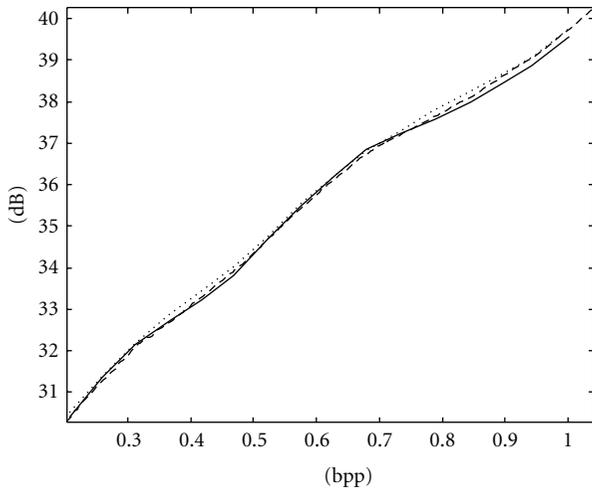


FIGURE 9: RD curves of Fetus ultrasonic image using the fixed scan order [21] (solid line), PCRD (dashed line), and ACPS (dotted line).

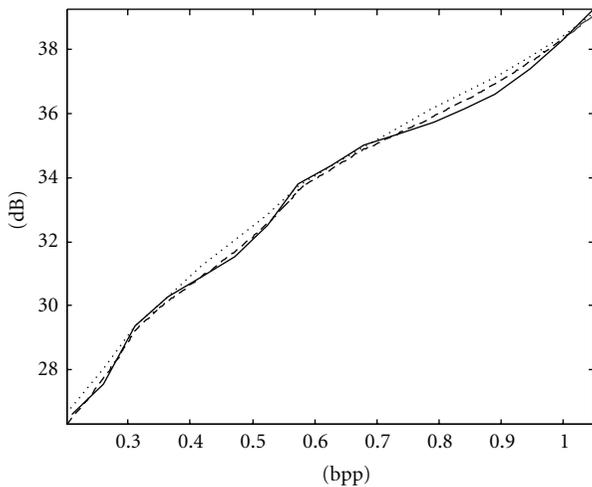


FIGURE 10: RD curves of Head CT image using the fixed scan order [21] (solid line), PCRD (dashed line), and ACPS (dotted line).

TABLE 2: Improvement in speed using ACPS compared to PCRD.

| Image | Speed up |
|--------------------|----------|
| Spine (MRI) | 32.7% |
| Chest (X-ray) | 33.1% |
| Fetus (ultrasonic) | 52.7% |
| Head (CT) | 53.6% |

RDS. It is based on the MQ table, which is available at both encoder and decoder; thus, there is no need to transmit the scan order of the significant coding passes. Experimental results in biomedical image processing show that a great improvement in speed can be obtained by replacing PCRD with ACPS.

Acknowledgment

The National Science Council of Taiwan, Grant no. NSC-99-2221-E-239-034, supported this work.

References

- [1] S. Y. Chen, "Cardiac deformation mechanics from 4D images," *Electronics Letters*, vol. 43, no. 11, pp. 609–611, 2007.
- [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] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.
- [4] C. Cattani and A. Kudreyko, "Harmonic wavelet method towards solution of the Fredholm type integral equations of the second kind," *Applied Mathematics and Computation*, vol. 215, no. 12, pp. 4164–4171, 2010.
- [5] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, vol. 2010, Article ID 408418, 22 pages, 2010.
- [6] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, 2000.
- [7] T. Acharya and P.-S. Tsai, *JPEG2000 Standard for Image Compression*, Wiley-Interscience, New York, NY, USA, 2005.
- [8] M. Dyer, S. Nooshabadi, and D. Taubman, "Design and analysis of system on a chip encoder for JPEG2000," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 2, Article ID 4703231, pp. 215–225, 2009.
- [9] W. Du, J. Sun, and Q. Ni, "Fast and efficient rate control approach for JPEG2000," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 4, pp. 1218–1221, 2004.
- [10] K. N. Vikram, V. Vasudevan, and S. Srinivasan, "Rate-distortion estimation for fast JPEG2000 compression at low bit-rates," *Electronics Letters*, vol. 41, no. 1, pp. 16–18, 2005.
- [11] H. C. Fang, Y. W. Chang, T. C. Wang, C. T. Huang, and L. G. Chen, "High-performance JPEG 2000 encoder with rate-distortion optimization," *IEEE Transactions on Multimedia*, vol. 8, no. 4, Article ID 1658028, pp. 645–653, 2006.
- [12] J. An and Z. Cai, "Efficient rate control for lossless mode of JPEG2000," *IEEE Signal Processing Letters*, vol. 15, pp. 409–412, 2008.
- [13] F. Aulí-Llinàs and M. W. Marcellin, "Distortion estimators for bitplane image coding," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1772–1781, 2009.
- [14] E. J. Balster and W. F. Turri, "Fast PCRD-Opt computation in JPEG2000," in *Proceedings of the International Conference on Signals and Electronic Systems (ICSES '10)*, pp. 431–434, Gliwice, Poland, September 2010.
- [15] J. Xue, C. Li, and N. Zheng, "Proto-object based rate control for JPEG2000: an approach to content-based scalability," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1177–1184, 2011.
- [16] 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.
- [17] S. Y. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.

- [18] 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.
- [19] M. Li and W. Zhao, "Representation of a stochastic traffic bound," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 9, Article ID 5342414, pp. 1368–1372, 2010.
- [20] M. Li, W. S. Chen, and L. Han, "Correlation matching method for the weak stationarity test of LRD traffic," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 181–195, 2010.
- [21] F. Aulí-Llinàs, J. Serra-Sagristà, J. L. Monteagudo-Pereira, and J. Bartrina-Rapesta, "Efficient rate control for JPEG2000 coder and decoder," in *Proceedings of the Data Compression Conference (DCC '06)*, pp. 282–291, Snowbird, Utah, USA, March 2006.

Review Article

Asymptotic Identity in Min-Plus Algebra: A Report on CPNS

Ming Li¹ and Wei Zhao²

¹ School of Information Science & Technology, East China Normal University, No. 500 Dong-Chuan Road, Shanghai 200241, China

² Department of Computer and Information Science, University of Macau, Avenue Padre Tomas Pereira, Taipa, Macau

Correspondence should be addressed to Ming Li, ming_lihk@yahoo.com

Received 23 March 2011; Accepted 20 April 2011

Academic Editor: Carlo Cattani

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

Network calculus is a theory initiated primarily in computer communication networks, especially in the aspect of real-time communications, where min-plus algebra plays a role. Cyber-physical networking systems (CPNSs) are recently developing fast and models in data flows as well as systems in CPNS are, accordingly, greatly desired. Though min-plus algebra may be a promising tool to linearize any node in CPNS as can be seen from its applications to the Internet computing, there are tough problems remaining unsolved in this regard. The identity in min-plus algebra is one problem we shall address. We shall point out the confusions about the conventional identity in the min-plus algebra and present an analytical expression of the asymptotic identity that may not cause confusions.

1. Introduction

We use the term cyber-physical networking systems (CPNS) instead of cyber-physical systems (CPS) as that in Song et al. [1] for the meaning of Internet of Things (IoT) that was stated by Commission of the European Communities [2] or Networks of Things (NoT) as discussed by Ferscha et al. [3], intending to emphasize the point that we are interested in the networking theory in CPS. Communication networks in CPNS include, but are never limited to, the Internet. Physical systems considered in CPNS are heterogeneous, ranging from telemedicine systems to geophysical ones, see, for example, Clifton et al. [4], Traynor [5], Chang [6]. Obviously, data in various physical systems are heterogeneous, see, for example, Chang [6], Goodchild [7], Lai and Xing [8], Mandelbrot [9–11], Hainaut and Devolder [12], Cattani [13–17], Chen et al. [18–22], Mikhael and Yang [23], Bakhoum and Toma [24–26], Li [27–32], Li et al. [33–39], Messina et al. [40], Humi [41], Dong [42], Liu [43], Toma [44], Abuzeid et al. [45], [46–49], Werner [50], and West [51], just naming a few.

There are two challenge issues in CPNS. On the one hand, data models that are irrelevant of statistics of a random function $x(t)$ are greatly desired. On the other hand, theory that may be used to linearize nonlinear data transmission systems but irrelevant of their nonlinearity is particularly

expected, because communication systems, including the Internet, are, in nature, nonlinear due to queuing, see, for example, Akimaru and Kawashima [52], Yue et al. [53], Gibson [54], Cooper [55], Pitts and Schormans [56], McDysan [57], and Stalling [58]. In short, we are interested in *data models that are irrelevant of their statistics and system theory that is irrelevant of the nonlinearity of systems*.

The early work regarding the above in italic may refer to Cruz [59–61], Zhao and Ramamritham [62], Raha et al. [63], Chang [64, 65], Boudec [66], Boudec and Patrick [67], Firoiu et al. [68], and Agrawal et al. [69]. Following Cruz [59, 60], the theory for the above in italic is called network calculus, see, for example, [66, 67], Jiang and Liu [70]. Chang [71] uses the term (σ, ρ) calculus, which is taken as the synonym of network calculus of Cruz in this paper.

The main application area of network calculus is conventionally to computer science, the Internet in particular, see, for example, Wang et al. [72, 73], Li and Zhao [74, 75], Fidler [76], Jiang [77], Jiang et al. [78], Liu et al. [79], Li et al. [80], Li and Kinghtly [81], Burchard et al. [82], Ng et al. [83], Raha et al. [84, 85], Starobinski and Sidi [86], Fukás et al. [87], Jia et al. [88], Golestani [89], and Lenzini et al. [90]. However, we have to emphasize the point that its applications are never limited to computer science. Rather, it is a theory to model data irrelevant of their statistics and to deal with data transmission without the necessity in principle to consider

the nonlinearity of transmission systems, as we shall explain in the next section. Therefore, it may be a promising tool to deal with data and systems in CPNS.

Basically, the fundamental theory of network calculus consists of three parts as described below.

- (i) (σ, ρ) model of arrival data $x(t)$,
- (ii) relationship between $x(t)$, single system (or node or server) $S(t)$ that is usually called service curve, and departure data $y(t)$,
- (iii) departure data $y(t)$ of a series of systems (nodes or servers) $S_n(t)$ ($n = 1, 2, \dots$), driven by arrival data $x(t)$,

where min-plus algebra plays a role, see, for example, [66, 67, 70, 71, 76].

The contributions of this paper are in the following three aspects:

- (i) the problem statement,
- (ii) the proof of the existence of the identity in the min-plus algebra in the domain of generalized functions,
- (iii) the asymptotic expression of the identity.

The rest of paper is organized as follows. Research background is discussed in Section 2. In Section 3, we will brief the min-plus algebra and state the problem regarding the identity in this algebra system. In Section 4, we shall address the existence of the identity in the min-plus algebra. The asymptotic expression of the identity is presented in Section 5. Discussions are given in Section 6, which is followed by conclusions.

2. Research Background

Data in CPNS are heterogeneous. They may be from sensors like radio-frequency identification (RFID), see, for example, [91], Ilie-Zudor et al. [92], Ahuja and Potti [93], data traffic in the Internet [38], transportation traffic (see [94–98]), ocean waves (see [31]), sea level (see [36, 99]), medical signals (see [14]), hydrological data (see [100]), financial data (see [101]), and so on. They may be Gaussian (see [29, 31]) or non-Gaussian (see [12, 102]). They may be in fractional order or integer order. In the case of fractional order, they may be unifractal or multifractal. The sample size of data of interest may be long enough for statistical analysis or very short, for example, a short conversation in mobile phone networks. On the other side, systems are also heterogeneous. Therefore, CPNS challenges us two tough issues. One is in data modeling and the other system modeling. We shall exhibit that the min-plus algebra in network calculus may yet serve as a tool in this regard.

2.1. Network Model. We first explain a single node in CPNS. Then, a model of tandem network is mentioned.

2.1.1. Nonlinearity of Node in CPNS. Denote by N a node in CPNS, see Figure 1. Suppose there are m clients arriving at

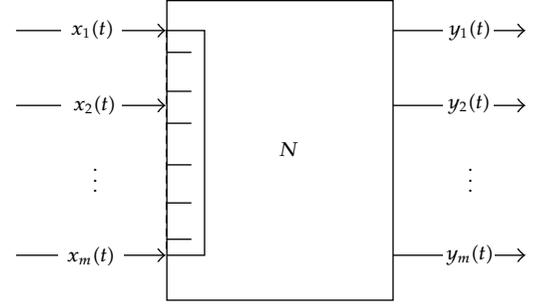


FIGURE 1: Single node in CPNS.

the input of N at time t , see, for example, Starobinski et al. [103].

Without confusions, we use N to represent the operator of node N such that

$$y_i(t) = Nx_i(t), \quad 1 \leq i \leq m. \quad (1)$$

Recall that queuing is a phenomenon often occurring in CPNS. For instance, cars in highways are often queued. Clients in a library for borrowing or returning books need queuing. Suppose client $x_i(t)$ suffers from delay $d_i(t)$. Then,

$$y_i(t) = x_i(t + d_i(t)), \quad 1 \leq i \leq m. \quad (2)$$

Note that $d_i(t)$ is a random variable in two senses. One is

$$d_i(t) \neq d_j(t), \quad 1 \leq i \leq m, 1 \leq j \leq m, i \neq j. \quad (3)$$

The other is

$$d_i(t_1) \neq d_i(t_2), \quad 1 \leq i \leq m, t_1 \neq t_2. \quad (4)$$

Therefore, we have the following remark.

Remark 1 (nonlinearity). A node N in CPNS is usually nonlinear. That is,

$$\sum y_i(t) \neq \sum Nx_i(t), \quad 1 \leq i \leq m. \quad (5)$$

2.1.2. Number of Arrivals is Random. The number of arrivals, denoted by m in Figure 1, is random.

Note 1. We need theory to deal with a nonlinear node N with m arrival clients, where m is a random variable.

2.1.3. Tandem Network Model. A single node previously described is not enough in CPNS since a client may be served by a series of n nodes, which we call tandem network, see Figure 2.

According to Remark 1, each node in Figure 2 is nonlinear. In addition, considering Note 1, we see that the number of arrival clients at the input of each node is random. Some clients may go through from $N1$ to Nn while others may not. For instance, client $x_{1i1}(t)$ leaves the tandem network when it passes through $N1$. Further more, some clients, for example, $x_{21}(t)$, arrive at this tandem network

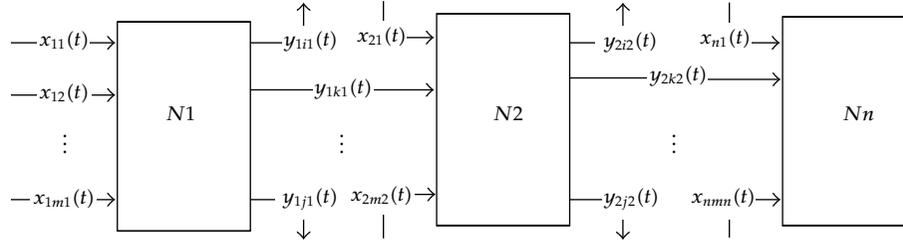


FIGURE 2: Tandem network.

at the input of $N2$. In general, how many clients leave the tandem network at the output of a specific node and how many clients arrive at the input of another specific node are uncertain.

Note 2. We need theory to handle a nonlinear system that is a tandem network as that in Figure 2 to assure the quality of service (QoS) of a specific client or of a specific class of clients within a given period of time.

The above Note 1 and Note 2 propose two challenge tasks in system theory. We shall explain how min-plus algebra is capable of dealing with those tasks late.

2.2. Data Modeling. We consider two classes of data flow. One is arrival data in the aggregated case, or aggregated clients, and the other arrival data of a specific client. In terms of network communications, the former is usually called aggregated arrival traffic while later arrival traffic at connection level. Without confusions, we use the term traffic rather than client.

One of radical properties of arrival traffic (traffic for short) is remarked below.

Remark 2 (positive). Traffic $x_i(t)$ is positive. That is,

$$x_i(t) \geq 0, \quad t \in \mathbb{R}, \quad (6)$$

where \mathbb{R} is the set of real numbers.

Another radical property of traffic is that the maximum of $x_i(t)$ is finite. More precisely, the value of $x_i(t)$ may never be infinite. Thus, we have the following remark.

Remark 3 (finite range). The maximum of $x_i(t)$ is finite. That is,

$$0 \leq x_i(t) \leq x_{i,\max}. \quad (7)$$

Remark 4 (randomness). The function $x_i(t)$ is usually random. This implies that

$$x_i(t_1) \neq x_i(t_2) \quad \text{for } t_1 \neq t_2. \quad (8)$$

2.2.1. Traffic at Connection Level. At connection level, for instance, for the i th connection, traffic is $x_i(t)$. One particularity of $x_i(t)$ is that t for $x_i(t)$ usually lasts within a finite time interval, say, $[0, T]$. The width of the interval may be

short, such as a short conversation like a word ‘‘hello’’ or long, such as a long speech over a network. In any case, it is finite. Modeling $x_i(t)$ with short interval is particularly desired and challenging.

Note 3. In the discrete case, the length of $x_i(t)$ may be too short to the proper statistical analysis of $x_i(t)$ in practice.

Note 4. Without confusions, we use $[0, T]$ to represent the interval in both the continuous case and the discrete one. In the continuous case, $[0, T] \in \mathbb{R}$. In the discrete case, $[0, T] \in \mathbb{Z}$, where \mathbb{Z} is the set of integer numbers, implying $t = 0, 1, \dots, T$. We use $[t_1, t_2]$ to represent an interval the starting point of which is nonzero.

2.2.2. Aggregated Traffic. We adopt Figure 1 to discuss aggregated traffic. At time t , aggregated traffic denoted by $x(t)$ at a node is expressed by

$$x(t) = \sum x_i(t), \quad i = 1, \dots, m. \quad (9)$$

In contrary to $x_i(t)$, the particularity of $x(t)$ is that t for $x(t)$ usually lasts within an interval longer than that of $x_i(t)$. As a matter of fact, if $x_i(t)$ passes through a node, another arrival flow $x_j(t)$ ($j = 1, \dots, m$) may arrive at the node. Consequently, in general, we should consider $t \in (0, \infty)$ for $x(t)$.

2.3. Accumulated Traffic. Traffic, either $x_i(t)$ or $x(t)$, discussed previously is instantaneous one. Data modeling of instantaneous traffic is essential, as we need understanding what its behaviors are at instantaneous time t at the input of a node. However, from the point of view of the service of a node, we also need data modeling of accumulated traffic within a time interval, say, $[0, T]$, without loss of generality, because it is desired for us to understand what the service performance of the node is for the purpose of proper design of a buffer size as well as scheduling policy of the node.

2.3.1. Accumulated Traffic at Connection Level. In the continuous case, the accumulated traffic of $x_i(t)$ within the interval $[0, T]$ is denoted by $X_i(T)$. It is given by

$$X_i(T) = \int_0^T x_i(t) dt, \quad t \in \mathbb{R}. \quad (10)$$

In the discrete case,

$$X_i(T) = \sum_{t=0}^{T-1} x_i(t), \quad t \in \mathbb{Z}. \quad (11)$$

2.3.2. Accumulated Traffic in the Aggregated Case. Denote by $X(T)$ the accumulated traffic in the aggregated case within the interval $[0, T]$. Then, in the continuous case, we have

$$X(T) = \int_0^T x(t)dt, \quad t \in \mathbb{R}. \quad (12)$$

In the discrete case,

$$X(T) = \sum_{t=0}^{T-1} x(t), \quad t \in \mathbb{Z}. \quad (13)$$

The mathematical expressions of $X(T)$ and $X_i(T)$ appear similar except the subscript i . However, $X(T)$ differs from $X_i(T)$ substantially in analysis in methodology. On the one hand, T for $X_i(T)$ should be assumed to be short such that conventional methods in statistics fail to its statistical analysis. On the other hand, T for $X(T)$ may be large enough such that it may be sectioned for the statistical analysis, see, for example, Li et al. [104].

2.3.3. A Basic Property of Accumulated Traffic. One property of accumulated traffic, either $X(T)$ or $X_i(T)$, is the wide sense increasing. By wide sense increasing, we mean that

$$X_i(T_1) \leq X_i(T_2) \quad \text{for } T_1 \leq T_2, \quad (14)$$

or

$$X(T_1) \leq X(T_2) \quad \text{for } T_1 \leq T_2. \quad (15)$$

Therefore, the data functions or series we face with are increasing ones in the wide sense.

2.3.4. (σ, ρ) Model of Data. For $\sigma \geq 0$ and $\rho \geq 0$, the following is called the (σ, ρ) model of data $x_i(t)$,

$$X_i(T) = \int_0^T x_i(u)du \leq \sigma_i + \rho_i T. \quad (16)$$

Note 5. The model expressed by (16) is irrelevant of any information of statistics of $x_i(t)$. The advantage of this model is at the cost of using inequality instead of equality.

Note 6. The model of (16) is simple in computation. Thus, it may be effective in practice, particularly in environments of CPNS, where simple computations are always expected.

For accumulated traffic $X(T)$, we have

$$X(T) = \int_0^T x(u)du \leq \sigma + \rho T. \quad (17)$$

Due to sufficiently large T , we may set the starting time by T_0 . In this case, we have

$$\int_{T_0}^T x(u)du \leq \sigma(T_0) + \rho(T - T_0). \quad (18)$$

Moreover, we are allowed to section the above integral such that

$$\int_{nT}^{(n+1)T} x(u)du \leq \sigma(nT) + \rho(T), \quad n = 0, 1, \dots \quad (19)$$

Without loss of generality, we use (17) to explain σ and ρ .

Remark 5. The parameter σ represents the bound of the burstness or local irregularity of $x(t)$, because

$$0 \leq \lim_{T \rightarrow 0} \int_0^T x(u)du \leq \sigma. \quad (20)$$

Note that the above integral does not make sense if $\lim_{T \rightarrow 0} \int_0^T x(t)dt \neq 0$ for the continuous $x(t)$ even in the field of the Lebesgue's integrals, see Dudley [105], Bartle and Sherbert [106], and Trench [107] for the contents of the Lebesgue's integrals. However, it makes sense when it is considered in the domain of generalized functions, which we shall brief in the following section. A simple way to explain (20) is

$$\lim_{T \rightarrow 0} \int_0^T x(t)dt = \int_0^T \sigma_1 \delta(t)dt, \quad (21)$$

where $\sigma_1 \leq \sigma$ and $\delta(t)$ is the Dirac- δ function.

Remark 6. The parameter ρ represents the bound of the average rate of $X(T)$, because

$$0 \leq \lim_{T \rightarrow \infty} \frac{\int_0^T x(t)dt}{T} \leq \rho = \text{constant}. \quad (22)$$

Remark 7. The parameter σ measures the local property of $x(t)$ while ρ is a measure of global property of $x(t)$.

3. Min-Plus Algebra and Problem Statement

Min-plus convolution is essential in the min-plus algebra. In this section, we first briefly review the conventional convolution in linear systems. Then, we shall visit min-plus convolution. Finally, we shall state the problem in the aspect of identity in the min-plus algebra.

3.1. Conventional Convolution. Denote by p a real number that satisfies $1 \leq p < \infty$. If a function $f(t)$ defined on $[a, b]$, where a is allowed to be $-\infty$ and b is allowed to be ∞ , is measurable and

$$\int_a^b |f(u)|^p du < \infty, \quad (23)$$

we say that $f(t) \in L^p(a, b)$.

Suppose that two functions $f_1(t), f_2(t) \in L^1(-\infty, \infty)$. Then, one says that $f_1(t)$ convolutes $f_2(t)$ if

$$f_1(t) * f_2(t) = \int_{-\infty}^{\infty} f_1(u)f_2(t-u)du, \quad (24)$$

where $*$ is the symbol implying the operation of convolution. We call it conventional convolution so as to distinguish

it from the min-plus convolution we are discussing in this paper.

The conventional convolution is crucial for linear systems, see, for example, Gibson [54], Box et al. [108], Mitra and Kaiser [109], Papoulis [110], Harris [111], Mikusinski [112], Fuller [113], and Bendat and Piersol [114], just naming a few. It has the properties described by the following lemmas.

Lemma 1. *In the algebra system $(L^1; *)$, the conventional convolution is commutative.*

Lemma 2 (closure of $*$). *If $f_1(t), f_2(t) \in L^1$, then $f_1(t) * f_2(t) \in L^1$.*

Lemma 3. *In the algebra system $(L^1; +, *)$, where $+$ implies the ordinary addition, $*$ with respect to $+$ is distributive.*

Lemma 4. *For $a \in \mathbb{R}$, $[af_1(t)] * f_2(t) = f_1(t) * [af_2(t)] = a[f_1(t) * f_2(t)]$.*

Lemma 5. *The identity in $(L^1; *)$ is the Dirac- δ function $\delta(t)$ that is defined by*

$$f(t) = \int_{-\infty}^{\infty} f(u)\delta(t-u)du, \quad (25)$$

where $f(t) \in L^1(-\infty, \infty)$ is continuous at t .

In fact, in the domain of generalized functions, we have

$$\int_{-\infty}^{\infty} \delta(u)du < \infty. \quad (26)$$

Thus, $\delta(t) \in L^1(-\infty, \infty)$ in the sense of generalize functions. Consequently, $\delta(t)$ is taken as the asymptotic identity in $(L^1; *)$ in the domain of generalized functions. Accordingly, the inverse of the conventional convolution discussed by, for instance, Mikusinski [112], Bracewell [115], Huang and Qiu [116], Abutaleb et al. [117], Rhoads and Ekstrom [118], Todoschuck and Jensen [119], and Moreau et al. [120], exists because the necessary and sufficient condition that the inverse of an operation exists is that there exists the identity in that system, see, for example, Korn and Korn [121], Zhang [122], Riley et al. [123], Bronshtein et al. [124], and Stillwell [125], but it should be in the sense of generalized functions. As a matter of fact, the conventional convolution itself is in that sense, see, for example, Smith [126].

Theorem 1. *The algebra system $(L^1; *)$ is a group.*

Proof. First, the operation $*$ is closed in L^1 . Second, $*$ is commutative because, for any $f_1(t), f_2(t), f_3(t) \in L^1(-\infty, \infty)$,

$$f_1(t) * [f_2(t) * f_3(t)] = [f_1(t) * f_2(t)] * f_3(t). \quad (27)$$

Finally, there exists the left identity denoted by $\delta(t)$ and the right one again denoted by $\delta(t)$ in $(L^1; *)$ such that

$$f(t) * \delta(t) = \delta(t) * f(t) \quad \text{for any } f(t) \in L^1(-\infty, \infty). \quad (28)$$

Thus, $(L^1; *)$ is a group. \square

3.2. Min-Plus Convolution. Considering the property of wide sense increasing of accumulated traffic mentioned in Section 2.3, we denote by \mathcal{S} the set that contains all functions that are greater than or equal to zero and that are wide sense increasing.

Definition 1. Let $X_1(t), X_2(t) \in \mathcal{S}$. Then, the following operation is called min-plus convolution:

$$X_1(t) \otimes X_2(t) = \inf_{0 \leq u \leq t} \{X_1(u) + X_2(t-u)\}, \quad (29)$$

where \otimes represents the operation of the min-plus convolution.

Example 1. Let $X(t) = t^2$ for $t > 0$ and 0 elsewhere. Then, $X(t) \otimes X(t) = t^2/2$.

Lemma 6 (closure of \otimes). *Let $X_1(t), X_2(t) \in \mathcal{S}$. Then, $X_1(t) \otimes X_2(t) \in \mathcal{S}$.*

Lemma 7. *The operation \otimes is commutative. That is,*

$$X_1(t) \otimes X_2(t) = X_2(t) \otimes X_1(t) \quad \text{for } X_1(t), X_2(t) \in \mathcal{S}. \quad (30)$$

Define another operation that is denoted by \wedge such that

$$X_1(t) \wedge X_2(t) = \inf[X_1(t), X_2(t)] \quad \text{for } X_1(t), X_2(t) \in \mathcal{S}. \quad (31)$$

Then, we have an algebra system denoted by $(\mathcal{S}, \wedge, \otimes)$ that follows the distributive law.

Lemma 8. *The operation \otimes with respect to \wedge is distributive. That is, for $X_1(t), X_2(t), X_3(t) \in \mathcal{S}$, one has*

$$[X_1(t) \wedge X_2(t)] \otimes X_3(t) = [X_1(t) \otimes X_3(t)] \wedge [X_2(t) \otimes X_3(t)]. \quad (32)$$

The following rule useful in this research is stated as follows.

Lemma 9. *Suppose $K \in \mathbb{R}$. Then, for $X_1(t), X_2(t) \in \mathcal{S}$, one has*

$$[X_1(t) + K] \otimes X_2(t) = X_1(t) \otimes X_2(t) + K, \quad (33)$$

where $+$ is the ordinary addition.

Denote by $I_1(t)$ the conventional identity in the min-plus algebra, which is defined by

$$I_1(t) = \begin{cases} \infty, & t > 0, \\ 0, & t < 0, \end{cases} \quad (34)$$

see [66–70].

It seems quite obvious when one takes $I_1(t)$ as the identity in the min-plus algebra since

$$X(t) \otimes I_1(t) = I_1(t) \otimes X(t) = X(t). \quad (35)$$

However, we shall soon point the contradictions of $I_1(t)$ below.

3.3. *Problem Statement.* Denote by $u(t)$ the Heavyside unit step function. That is,

$$u(t) = \begin{cases} 1, & t > 0, \\ 0, & t < 0. \end{cases} \quad (36)$$

Then, for $K \in \mathbb{R}$, we have

$$Ku(t) = \begin{cases} K, & t > 0, \\ 0, & t < 0. \end{cases} \quad (37)$$

Using (34), we have

$$\begin{aligned} I_1(t) + Ku(t) &= \begin{cases} \infty + K, & t > 0 \\ 0, & t < 0 \end{cases} \\ &= \begin{cases} \infty, & t > 0 \\ 0, & t < 0 \end{cases} \quad (\text{Contradiction 1}) \\ &= I_1(t). \end{aligned} \quad (38)$$

The above is an obvious contradiction regarding the conventional identity defined by (34).

In addition to the above contradiction, we now state another problem regarding (34). As a matter of fact, if we let $X_1(t) = I_1(t)$ and $Ku(t)$ in Lemma 9, then, on the left side of (33) in Lemma 9, we have

$$[I_1(t) + Ku(t)] \otimes X_2(t) = I_1(t) \otimes X_2(t) = X_2(t). \quad (39)$$

On the other side, on the right side of (33) in Lemma 9, we have

$$\begin{aligned} [I_1(t) + Ku(t)] \otimes X_2(t) &= I_1(t) \otimes X_2(t) + Ku(t) \\ &= X_2(t) + Ku(t). \end{aligned} \quad (40)$$

Comparing the right sides of (39) with that of (40) yields another contradiction expressed by

$$X_2(t) = X_2(t) + Ku(t), \quad (\text{Contradiction 2}) \quad (41)$$

The above discussions imply that the definition of the identity of (34) in the min-plus algebra, which is commonly used in literature, see, for example, [66–70], may not be rigorous at least. Therefore, the conventional representation of the identity, that is, (34), may be inappropriate since it may mislead computation results like those in (39) and (40). Consequently, rigorous definition of the identity needs studying.

4. Existence of Identity in Min-Plus Algebra

The problems regarding the definition of the conventional identity, which we stated in Section 3.3, give rise to a question whether or not the identity in the min-plus algebra exists. The answer to this question is rarely seen, to the best of our knowledge. Another question resulted from Section 3.3 is what the rigorous representation of the identity is. We shall provide the answer to the first question in this section. The answer to the second will be explained in the next section.

4.1. *Preliminaries.* We brief some results in generalized functions [127–129] for the purpose of discussing the existence of identity.

Definition 2. Let $\text{supp}(f)$ be the support of a function $f : \mathbb{R} \rightarrow \mathbb{C}$. It implies $\{t : f(t) \neq 0\}$. The function is said to have a bounded support if there exist $a, b \in \mathbb{R}$ such that $\text{supp}(f) \subset [a, b]$.

Definition 3. A function $f : \mathbb{R} \rightarrow \mathbb{C}$ is said to have n time continuous derivatives if its first n derivatives exist and are continuous. If its derivatives of all orders exist and are continuous, f is said to be infinitely differentiable. In this case, f is said to be smooth.

Definition 4. A test function is a smooth $\mathbb{R} \rightarrow \mathbb{C}$ with $\text{supp}(f) \subset [a, b]$. The set of all test functions is denoted by \mathcal{D} .

Definition 5. A linear functional f on \mathcal{D} is a map $f : \mathcal{D} \rightarrow \mathbb{C}$ such that, for $a, b \in \mathbb{C}$ and $\phi, \psi \in \mathcal{D}$, $f(a\phi + b\psi) = af(\phi) + bf(\psi)$.

Definition 6. Denote by (ϕ_n) a sequence of test functions and another test function. We say that $\phi_n \rightarrow \phi$ if the following holds:

- (1) there is an interval $[a, b]$ that contains $\text{supp}(\phi)$ and $\text{supp}(\phi_n)$ for all n ,
- (2) $\lim_{n \rightarrow \infty} \phi_n^{(k)}(t) \rightarrow \phi^{(k)}(t)$ uniformly for $t \in [a, b]$.

Definition 7. A functional f on \mathcal{D} is continuous if it maps every convergent sequence in \mathcal{D} into a convergent sequence in \mathbb{C} . A continuous linear functional f on \mathcal{D} is termed a generalized function. It is often called a distribution in the sense of Schwartz.

Definition 8. A function $f : \mathbb{R} \rightarrow \mathbb{C}$ is locally integrable if $\int_a^b f(t)dt < \infty$ for all a, b .

Lemma 10. Any continuous, including piecewise continuous, function is locally integrable.

Lemma 11 (regular). Any locally integrable function f is a generalized function defined by

$$\langle f, \phi \rangle = \int_{-\infty}^{\infty} f(t)\phi(t)dt < \infty. \quad (42)$$

In this case, f is called regular.

Lemma 12. Any generalized function has derivatives of all orders.

Lemma 13. There exists the Fourier transform of any generalized function.

Definition 9 (rapid function). A function of rapid decay is a smooth function $\phi : \mathbb{R} \rightarrow \mathbb{C}$ such that $t^n \phi^{(r)}(t) \rightarrow 0$ as $t \rightarrow \pm \infty$ for all $n, r \geq 0$, where \mathbb{C} is the space of complex numbers. The set of all functions of rapid decay is denoted by \mathcal{S} .

Lemma 14. *Every function belonging to \mathcal{S} is absolutely integrable.*

4.2. Proof of Existence. Define the norm and inner product of $X \in \mathcal{S}$ by

$$\|X\|^2 = \langle X, X \rangle = \int_0^\infty X^2(u)w(u)du, \quad (43)$$

where $w \in \mathcal{S}$. Combining any $X \in \mathcal{S}$ with its limit yields a Hilbert space that we denote again by \mathcal{S} without confusions.

Let $g \in \mathcal{S}$ be a system function such that it transforms its input $X \in \mathcal{S}$ to the output by

$$y = (X \otimes g) \in \mathcal{S}. \quad (44)$$

Denote the system by the operator L . Then, we purposely force the functionality of L such that it maps an element $X \in \mathcal{S}$ to another element $(X \otimes g) \in \mathcal{S}$. Note that L is a linear operator. In fact, according to Lemma 8, we have

$$L(X \wedge g) = L(X) \wedge L(g). \quad (45)$$

In addition, from Lemma 9, we have

$$L(X + K) = L(X) + K. \quad (46)$$

Therefore, L is a linear mapping from \mathcal{S} to \mathcal{S} .

Denote by \mathcal{L} the space consisting of all such operators by

$$\mathcal{L}(\mathcal{S}, \mathcal{S}) = \mathcal{L}(\mathcal{S}). \quad (47)$$

Then, from Lemmas 8 and 9, one can easily see that $\mathcal{L}(\mathcal{S})$ is a linear space.

Lemma 15 (archimedes criterion). *For any positive real numbers $a > 0$ and $b > 0$, there exists positive integer $n \in \mathbb{Z}$ such that $na > b$ (see [130]).*

Lemma 16 (archimedes). *If $b \in \mathbb{R}$, there exists $n \in \mathbb{Z}$ such that $b < n$ (see [106]).*

Lemma 17. *An operator $T : X \mapsto Y$ is invertible if and only if there exists constant $m > 0$ such that for all $x \in X$, $\|Tx\| \geq m\|x\|$, where X and Y are linear normed spaces (see [131]).*

From the above discussions, we obtain the following theorem.

Theorem 2 (existence). *For $X, g \in \mathcal{S}$ and $X(0) \neq 0$ and $g(0) \neq 0$, if $L(X) = X \otimes g$ or $L_1(g) = g \otimes X$, then both L and L_1 are invertible. Consequently, the identity in the min-plus algebra exists.*

Proof. Consider

$$\begin{aligned} \|LX\| &= \sqrt{\|X \otimes g\|} \\ &= \sqrt{\int_0^\infty \left[\inf_{0 \leq u \leq t} \{X(u) + g(t-u)\} \right]^2 w(u)du}. \end{aligned} \quad (48)$$

Since

$$\inf_{0 \leq u \leq t} \{X(u) + g(t-u)\} \geq \inf \{X(u)\} = X(0) \quad (49)$$

and $X(u) \in \mathcal{S}$, we have

$$0 < X(0) \leq X(u). \quad (50)$$

According to Lemmas 15 and 16, there exists $m > 0$ such that

$$X(0) \geq m^2 X(u). \quad (51)$$

Therefore,

$$\begin{aligned} \|LX\| &\geq \sqrt{\int_0^\infty [\inf \{X(u)\}]^2 w(u)du} \\ &= \sqrt{\int_0^\infty [X(0)]^2 w(u)du} \\ &\geq m \sqrt{\int_0^\infty X(u)^2 w(u)du} = m\|X\|. \end{aligned} \quad (52)$$

Similarly, if $L_1 \in \mathcal{L}(\mathcal{S})$ is such that $L_1(g) = g \otimes X$, we have $\|L_1 g\| \geq m_1 \|g\|$ since $g(0) \neq 0$, where $m_1 > 0$ is a constant. Thus, according to Lemma 17, Theorem 2 holds. \square

Note 7. In Theorem 2, we need the conditions of $X(0) \neq 0$ and $g(0) \neq 0$. Since $X(t)$ and $g(t)$ are wide sense increasing, we need in fact $X(0) > 0$ and $g(0) > 0$.

5. Representation of Identity in Min-Plus Algebra

Express the Dirac- δ function by

$$\delta(t) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \cos(kt). \quad (53)$$

For the purpose of distinguishing the identity we present from the conventional one, we denote $I(t)$ as the identity in what follows instead of $I_1(t)$ as used in Section 3.

Theorem 3 (representation). *The identity in the min-plus algebra is expressed by*

$$I(t) = \lim_{T \rightarrow 0} \left[\frac{2}{T} + \frac{4}{T} \sum_{n=1}^{\infty} \cos\left(\frac{2n\pi t}{T}\right) \right]. \quad (54)$$

Proof. Take the following into account

$$\sum_{n=0}^{\infty} \delta(t - nT) \quad (T > 0). \quad (55)$$

Then, the identity in the discrete case is given by

$$I(k) = \sum_{n=0}^{\infty} \delta(k - nT). \quad (56)$$

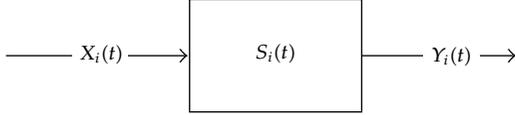


FIGURE 3: Single node with arrival and departure traffic.

FIGURE 4: N tandem nodes with arrival and departure traffic.

The identity in the continuous case is taken as the limit expressed by

$$I(t) = \lim_{T \rightarrow 0} \sum_{n=0}^{\infty} \delta(t - nT). \quad (57)$$

Considering the Poisson's summation formula, we have

$$I(k) = \frac{2}{T} + \frac{4}{T} \sum_{n=1}^{\infty} \cos\left(\frac{2n\pi k}{T}\right). \quad (58)$$

In the limit case,

$$I(t) = \lim_{T \rightarrow 0} \left[\frac{2}{T} + \frac{4}{T} \sum_{n=1}^{\infty} \cos\left(\frac{2n\pi t}{T}\right) \right]. \quad (59)$$

This completes the proof. \square

Remark 8. If one uses the representation in Theorem 3, the contradictions given in (38) and (41) vanish.

Note 8. The identity expressed by (59) is an asymptotic one.

6. Discussions

We mention an application of min-plus algebra to CPNS. Denote by $Y_i(t)$ the accumulated function characterizing the output of the i th node (Figure 3). Then, the min-plus convolution can be used to establish the relationship between $X_i(t)$, $S_i(t)$, and $Y_i(t)$ by

$$Y_i(t) \geq X_i(t) \otimes S_i(t) = \inf_{0 \leq u \leq t} \{S_i(u) + X_i(t - u)\}. \quad (60)$$

Suppose a traffic function passes through N tandem nodes from the first node with the service curve $S_1(t)$ to the N th node with the service curve $S_N(t)$ to reach the destination as indicated in Figure 4. Denote the departure traffic of the N th node by $Y_N(t)$. Then,

$$Y_N(t) \geq X_1(t) \otimes S_N^1(t) = \inf_{0 \leq u \leq t} \{S_N^1(u) + X_1(t - u)\}, \quad (61)$$

where (see [132])

$$S_N^1(t) = S_1(t) \otimes S_2(t) \otimes \dots \otimes S_i(t) \dots \otimes S_N(t). \quad (62)$$

Note 9. Min-plus algebra can be used to linearize a nonlinear system as can be seen from (62). Thus, it may yet be used as a theory in the aspect of data transmission systems in CPNS.

7. Conclusions

We have proposed the problem regarding the conventional identity in the min-plus algebra. In addition, we have presented the proof that the identity in the min-plus algebra exists in the domain of generalized function. Moreover, we have given the asymptotic expression of the identity in the system of min-plus algebra.

Acknowledgment

This work was partly supported by the National Natural Science Foundation of China (NSFC) under the project Grant nos. 60873264, 61070214, and the 973 plan under the Project no. 2011CB302801/2011CB302802.

References

- [1] Z. Song, Y. Q. Chen, C. R. Sastry, and N. C. Tas, *Optimal Observation for Cyber-Physical Systems*, Springer, 2009.
- [2] Commission of the European Communities, "Internet of things—an action plan for Europe," *Journal of International Wildlife Law and Policy*, vol. 12, no. 1-2, pp. 108–120, 2009.
- [3] A. Ferscha, M. Hechinger, A. Riener et al., "Peer-it: stick-on solutions for networks of things," *Pervasive and Mobile Computing*, vol. 43, no. 3, pp. 448–479, 2008.
- [4] G. D. Clifton, H. Byer, K. Heaton, D. J. Haberman, and H. Gill, "Provision of pharmacy services to underserved populations via remote dispensing and two-way videoconferencing," *American Journal of Health-System Pharmacy*, vol. 60, no. 24, pp. 2577–2282, 2003.
- [5] K. Traynor, "Navy takes telepharmacy worldwide," *American Journal of Health-System Pharmacy*, vol. 67, no. 14, pp. 1134–1136, 2010.
- [6] K. T. Chang, *Introduction to Geographical Information Systems*, McGraw-Hill, New York, NY, USA, 2008.
- [7] M. F. Goodchild, "Twenty years of progress: GIScience in 2010," *Journal of Spatial Information Science*, vol. 2010, no. 1, pp. 3–20, 2010.
- [8] T. L. Lai and H. Xing, *Statistical Models and Methods for Financial Markets*, Springer, 2008.
- [9] B. B. Mandelbrot, *Gaussian Self-A nity and Fractals*, Springer, 2001.
- [10] B. B. Mandelbrot, *Multifractals and 1/f Noise*, Springer, 1998.
- [11] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York, NY, USA, 1982.
- [12] D. Hainaut and P. Devolder, "Mortality modelling with Lévy processes," *Insurance*, vol. 42, no. 1, pp. 409–418, 2008.
- [13] C. Cattani, "Harmonic wavelet approximation of random, fractal and high frequency signals," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 207–217, 2010.
- [14] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.
- [15] G. Mattioli, M. Scalia, and C. Cattani, "Analysis of large-amplitude pulses in short time intervals: application to neuron interactions," *Mathematical Problems in Engineering*, vol. 2010, Article ID 895785, 15 pages, 2010.

- [16] C. Cattani, "Shannon wavelets for the solution of integrodifferential equations," *Mathematical Problems in Engineering*, vol. 2010, 22 pages, 2010.
- [17] C. Cattani, "Shannon wavelets theory," *Mathematical Problems in Engineering*, vol. 2008, Article ID 164808, 24 pages, 2008.
- [18] 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.
- [19] S. Y. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.
- [20] S. Y. Chen and Y. F. Li, "Determination of stripe edge blurring for depth sensing," *IEEE Sensors Journal*, vol. 11, no. 2, pp. 389–390, 2011.
- [21] 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.
- [22] S. Chen, J. Zhang, H. Zhang et al., "Myocardial motion analysis for determination of tei-index of human heart," *Sensors*, vol. 10, no. 12, pp. 11428–11439, 2010.
- [23] W. Mikhael and T. Yang, "A gradient-based optimum block adaptation ICA technique for interference suppression in highly dynamic communication channels," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, Article ID 84057, 10 pages, 2006.
- [24] E. G. Bakhoun and C. Toma, "Dynamical aspects of macroscopic and quantum transitions due to coherence function and time series events," *Mathematical Problems in Engineering*, vol. 2010, Article ID 428903, 13 pages, 2010.
- [25] E. G. Bakhoun and C. Toma, "Mathematical transform of traveling-wave equations and phase aspects of quantum interaction," *Mathematical Problems in Engineering*, vol. 2010, Article ID 695208, 15 pages, 2010.
- [26] E. G. Bakhoun and C. Toma, "Relativistic short range phenomena and space-time aspects of pulse measurements," *Mathematical Problems in Engineering*, vol. 2008, Article ID 410156, 20 pages, 2008.
- [27] M. Li, "Change trend of averaged Hurst parameter of traffic under DDOS flood attacks," *Computers and Security*, vol. 25, no. 3, pp. 213–220, 2006.
- [28] M. Li, "A class of negatively fractal dimensional gaussian random functions," *Mathematical Problems in Engineering*, vol. 2011, Article ID 291028, 18 pages, 2011.
- [29] M. Li, "Generation of teletraffic of generalized cauchy type," *Physica Scripta*, vol. 81, no. 2, Article ID 025007, 2010.
- [30] M. Li, "Fractal time series-a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [31] M. Li, "A method for requiring block size for spectrum measurement of ocean surface waves," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 6, pp. 2207–2215, 2006.
- [32] M. Li, "Modeling autocorrelation functions of long-range dependent teletraffic series based on optimal approximation in hilbert space-a further study," *Applied Mathematical Modelling*, vol. 31, no. 3, pp. 625–631, 2007.
- [33] M. Li, W. Zhao, and S. Chen, "MBm-based scalings of traffic propagated in internet," *Mathematical Problems in Engineering*, vol. 2011, Article ID 389803, 21 pages, 2011.
- [34] 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.
- [35] M. Li and W. Zhao, "Detection of variations of local irregularity of traffic under DDOS flood attack," *Mathematical Problems in Engineering*, vol. 2008, Article ID 475878, 11 pages, 2008.
- [36] M. Li, C. Cattani, and S.-Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 13 pages, 2011.
- [37] 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.
- [38] M. Li and S. C. Lim, "Modeling network traffic using generalized cauchy process," *Physica A*, vol. 387, no. 11, pp. 2584–2594, 2008.
- [39] M. Li, S. C. Lim, and Sy. Chen, "Exact solution of impulse response to a class of fractional oscillators and its stability," *Mathematical Problems in Engineering*, vol. 2011, Article ID 657839, 9 pages, 2011.
- [40] A. R. Messina, P. Esquivel, and F. Lezama, "Time-dependent statistical analysis of wide-area time-synchronized data," *Mathematical Problems in Engineering*, vol. 2010, Article ID 751659, 17 pages, 2010.
- [41] M. Humi, "Assessing local turbulence strength from a time series," *Mathematical Problems in Engineering*, vol. 2010, Article ID 316841, 13 pages, 2010.
- [42] M. Dong, "A tutorial on nonlinear time-series data mining in engineering asset health and reliability prediction: concepts, models, and algorithms," *Mathematical Problems in Engineering*, vol. 2010, Article ID 175936, 22 pages, 2010.
- [43] Z. Liu, "Chaotic time series analysis," *Mathematical Problems in Engineering*, vol. 2010, Article ID 720190, 31 pages, 2010.
- [44] G. Toma, "Specific differential equations for generating pulse sequences," *Mathematical Problems in Engineering*, vol. 2010, Article ID 324818, 11 pages, 2010.
- [45] O. M. Abuzeid, A. N. Al-Rabadi, and H. S. Alkhalidi, "Fractal geometry-based hypergeometric time series solution to the hereditary thermal creep model for the contact of rough surfaces using the Kelvin-Voigt medium," *Mathematical Problems in Engineering*, vol. 2010, Article ID 652306, 22 pages, 2010.
- [46] M. Li and W. Zhao, "Visiting power laws in cyber-physical networking systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 302786, p. 13, 2012.
- [47] M. Li and S. C. Lim, "Power spectrum of generalized cauchy process," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 219–222, 2010.
- [48] J. He, H. Qian, Y. Zhou, and Z. Li, "Cryptanalysis and improvement of a block cipher based on multiple chaotic systems," *Mathematical Problems in Engineering*, vol. 2010, Article ID 590590, 14 pages, 2010.
- [49] 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, 2010.
- [50] G. Werner, "Fractals in the nervous system: conceptual implications for theoretical neuroscience," *Frontiers in Fractal Physiology*, vol. 1, p. 28, 2010.
- [51] B. J. West, "Fractal physiology and the fractional calculus: a perspective," *Frontiers in Fractal Physiology*, vol. 1, 2010.
- [52] H. Akimaru and K. Kawashima, *Teletra c: Theory and Applications*, Springer, 1993.

- [53] W. Yue, H. Takagi, and Y. Takahashi, *Advances in Queueing Theory and Network Applications*, Springer, 2009.
- [54] J. D. Gibson, *The Communications Handbook*, IEEE Press, 1997.
- [55] R. B. Cooper, *Introduction to Queueing Theory*, Elsevier, 2nd edition, 1981.
- [56] J. M. Pitts and J. A. Schormans, *Introduction to ATM Design and Performance: With Applications Analysis Software*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2000.
- [57] D. McDysan, *QoS & Traffic Management in IP & ATM Networks*, McGraw-Hill, New York, NY, USA, 2000.
- [58] W. Stalling, *High-Speed Networks: TCP/IP and ATM Design Principles*, Prentice Hall, 2nd edition, 2002.
- [59] R. L. Cruz, "A calculus for network delay—part 1: network elements in isolation part 2," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 114–141, 1991.
- [60] R. L. Cruz, "A calculus for network delay—part 2: network analysis," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 132–141, 1991.
- [61] R. L. Cruz, "Quality of service guarantees in virtual circuit switched networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1048–1056, 1995.
- [62] W. Zhao and K. Ramamritham, "Virtual time csma protocols for hard real-time communication," *IEEE Transactions on Software Engineering*, vol. 13, no. 8, pp. 938–952, 1987.
- [63] A. Raha, S. Kamat, and W. Zhao, "Guaranteeing end-to-end deadlines in ATM networks," in *Proceedings of the 15th International Conference on Distributed Computing Systems*, pp. 60–68, June 1995.
- [64] C. S. Chang, "On deterministic traffic regulation and service guarantees: a systematic approach by filtering," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1097–1110, 1998.
- [65] C. S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 913–931, 1994.
- [66] J. Y. Le Boudec, "Application of network calculus to guaranteed service networks," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1087–1096, 1998.
- [67] J. Y. Le Boudec and T. Patrick, *Network Calculus, A Theory of Deterministic Queueing Systems for the Internet*, Springer, 2001.
- [68] V. Firoiu, J. Y. Le Boudec, D. Towsley, and Z. L. I. Zhang, "Theories and models for internet quality of service," *Proceedings of IEEE*, vol. 90, no. 9, pp. 1565–1591, 2002.
- [69] R. Agrawal, F. Baccelli, and R. Rajan, *An Algebra for Queueing Networks with Time Varying Service and Its Application to the Analysis of Integrated Service Networks*, INRIA, 1998, RR-3435.
- [70] Y. M. Jiang and Y. Liu, *Stochastic Network Calculus*, Springer, 2008.
- [71] C. S. Chang, *Performance Guarantees in Communication Networks*, Springer, 2000.
- [72] H. Wang, J. B. Schmitt, and I. Martinovic, "Dynamic demultiplexing in network calculus—theory and application," *Performance Evaluation*, vol. 68, no. 2, pp. 201–219, 2011.
- [73] S. Q. Wang, D. Xuan, R. Bettati, and W. Zhao, "Toward statistical QoS guarantees in a differentiated services network," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 253–263, 2010.
- [74] C. Z. Li and W. Zhao, "Stochastic performance analysis of non-feedforward networks," *Telecommunication Systems*, vol. 43, no. 3-4, pp. 237–252, 2010.
- [75] M. Li and W. Zhao, "Representation of a stochastic traffic bound," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 9, pp. 1368–1372, 2010.
- [76] M. Fidler, "Survey of deterministic and stochastic service curve models in the network calculus," *IEEE Communications Surveys and Tutorials*, vol. 12, no. 1, pp. 59–86, 2010.
- [77] Y. M. Jiang, "Per-domain packet scale rate guarantee for expedited forwarding," *IEEE/ACM Transactions on Networking*, vol. 14, no. 3, pp. 630–643, 2006.
- [78] Y. M. Jiang, Q. Yin, Y. Liu, and S. Jiang, "Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks," *Computer Networks*, vol. 53, no. 12, pp. 2011–2019, 2009.
- [79] Y. Liu, C. K. Tham, and Y. M. Jiang, "A calculus for stochastic QoS analysis," *Performance Evaluation*, vol. 64, no. 6, pp. 547–572, 2007.
- [80] C. Li, A. Burchard, and J. Liebeherr, "A network calculus with effective bandwidth," *IEEE/ACM Transactions on Networking*, vol. 15, no. 6, pp. 1442–1453, 2007.
- [81] C. Li and E. Knightly, "Schedulability criterion and performance analysis of coordinated schedulers," *IEEE/ACM Transactions on Networking*, vol. 13, no. 2, pp. 276–287, 2005.
- [82] A. Burchard, J. Liebeherr, and S. D. Patek, "A min-plus calculus for end-to-end statistical service guarantees," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4105–4114, 2006.
- [83] J. K. Y. Ng, S. Song, and W. Zhao, "Integrated end-to-end delay analysis for regulated ATM networks," *Real-Time Systems*, vol. 25, no. 1, pp. 93–124, 2003.
- [84] A. Raha, S. Kamat, X. Jia, and W. Zhao, "Using traffic regulation to meet end-to-end deadlines in ATM networks," *IEEE Transactions on Computers*, vol. 48, no. 9, pp. 917–935, 1999.
- [85] A. Raha, W. Zhao, S. Kamat, and W. Jia, "Admission control for hard real-time connections in ATM LANs," *IEEE Proceedings*, vol. 148, no. 4, pp. 1–12, 2001.
- [86] D. Starobinski and M. Sidi, "Stochastically bounded burstiness for communication networks," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 206–212, 2000.
- [87] H. Fukś, A. T. Lawnczak, and S. Volkov, "Packet delay in models of data networks," *ACM Transactions on Modeling and Computer Simulation*, vol. 11, no. 3, pp. 233–250, 2001.
- [88] X. Jia, W. Zhao, and J. Li, "An integrated routing and admission control mechanism for real-time multicast connections in ATM networks," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1515–1519, 2001.
- [89] S. J. Golestani, "Network delay analysis of a class of fair queueing algorithms," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1057–1070, 1995.
- [90] L. Lenzi, E. Mingozzi, and G. Stea, "A methodology for computing end-to-end delay bounds in FIFO-multiplexing tandems," *Performance Evaluation*, vol. 65, no. 11-12, pp. 922–943, 2008.
- [91] INFOS D.4 Networked Enterprise & RFID INFOS G.2 Micro & Nanosystems, in co-operation with the Working Group RFID of the ETP EPOSS, Internet of Things in 2020, Roadmap for the Future[R]. Version 1.1, May 2008.
- [92] E. Ilie-Zudor, Z. Kemény, F. van Blommestein, L. Monostori, and A. van Der Meulen, "A survey of applications and requirements of unique identification systems and RFID techniques," *Computers in Industry*, vol. 62, no. 3, pp. 227–252, 2011.
- [93] S. Ahuja and P. Potti, "An introduction to RFID technology," *Communications and Network*, vol. 2, no. 3, pp. 183–186, 2010.

- [94] R. Wang, L. Zhang, R. Sun, J. Gong, and L. Cui, "A pervasive traffic information acquisition system based on wireless sensor networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 615–621, 2011.
- [95] W. H. K. Lam, S. C. Wong, and H. K. Lo, "Emerging theories in traffic and transportation; and emerging methods for transportation planning and operations," *Transportation Research Part C*, vol. 19, no. 2, pp. 169–171, 2011.
- [96] L. N. Emmanuel, R. T. Antonio, and L. M. Ernesto, "A modeling framework for urban traffic systems microscopic simulation," *Simulation Modelling Practice and Theory*, vol. 18, no. 8, pp. 1145–1161, 2010.
- [97] J. A. Laval, "Hysteresis in traffic flow revisited: an improved measurement method," *Transportation Research Part B*, vol. 45, no. 2, pp. 385–391, 2011.
- [98] B. G. Heydecker and J. D. Addison, "Analysis and modelling of traffic flow under variable speed limits," *Transportation Research Part C*, vol. 19, no. 2, pp. 206–217, 2011.
- [99] P. Santoro, M. Fernández, M. Fossati, G. Cazes, R. Terra, and I. Piedra-Cueva, "Pre-operational forecasting of sea level height for the Río de la Plata," *Applied Mathematical Modelling*, vol. 35, no. 5, pp. 2264–2278, 2011.
- [100] D. She and X. Yang, "A new adaptive local linear prediction method and its application in hydrological time series," *Mathematical Problems in Engineering*, vol. 2010, Article ID 205438, 15 pages, 2010.
- [101] S. V. Muniandy, S. C. Lim, and R. Murugan, "Inhomogeneous scaling behaviors in Malaysian foreign currency exchange rates," *Physica A*, vol. 301, no. 1–4, pp. 407–428, 2001.
- [102] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes, Stochastic Models with Infinite Variance*, Chapman and Hall, New York, NY, USA, 1994.
- [103] D. Starobinski, M. Karpovsky, and L. A. Zakrevski, "Application of network calculus to general topologies using turn-prohibition," *IEEE/ACM Transactions on Networking*, vol. 11, no. 3, pp. 411–421, 2003.
- [104] M. Li, S. Wang, and W. Zhao, "A real-time and reliable approach to detecting traffic variations at abnormally high and low rates," *Springer Lecture Notes in Computer Science*, vol. 4158, pp. 541–550, 2006.
- [105] R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, 2002.
- [106] R. G. Bartle and D. R. Sherbert, *Introduction to Real Analysis*, John Wiley & Sons, 3rd edition, 2000.
- [107] W. F. Trench, *Introduction to Real Analysis*, Pearson Education, 2003.
- [108] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice Hall, 1994.
- [109] S. K. Mitra and J. F. Kaiser, *Handbook for Digital Signal Processing*, John Wiley & Sons, 1993.
- [110] A. Papoulis, *The Fourier Integral and Its Applications*, McGraw-Hill, 1962.
- [111] C. M. Harris, *Shock and Vibration Handbook*, McGraw-Hill, 4th edition, 1995.
- [112] J. Mikusinski, *Operational Calculus*, Pergamon Press, 1959.
- [113] W. A. Fuller, *Introduction to Statistical Time Series*, Wiley, 2nd edition, 1996.
- [114] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedure*, John Wiley & Sons, 3rd edition, 2000.
- [115] R. N. Bracewell, *The Fourier Transform and Its Applications*, McGraw-Hill, New York, NY, USA, 2nd edition, 1978.
- [116] X. Huang and P. H. Qiu, "Blind deconvolution for jump-preserving curve estimation," *Mathematical Problems in Engineering*, vol. 2010, Article ID 350849, 9 pages, 2010.
- [117] A. S. Abutaleb, N. M. Elhamy, and M. E. S. Waheed, "Blind deconvolution of the aortic pressure waveform using the malliavin calculus," *Mathematical Problems in Engineering*, vol. 2010, Article ID 102581, 27 pages, 2010.
- [118] R. L. Rhoads and M. P. Ekstrom, "Removal of interfering system distortion by deconvolution," *IEEE Transactions, Instrumentation and Measurement*, vol. 17, no. 4, pp. 333–337, 1968.
- [119] J. P. Todoeschuck and O. G. Jensen, "Scaling geology and seismic deconvolution," *Pure and Applied Geophysics*, vol. 131, no. 1-2, pp. 273–287, 1989.
- [120] S. Moreau, G. Plantier, J. C. Valière, H. Bailliet, and L. Simon, "Estimation of power spectral density from laser Doppler data via linear interpolation and deconvolution," *Experiments in Fluids*, vol. 50, no. 1, pp. 179–188, 2011.
- [121] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, McGraw-Hill, 1961.
- [122] H. R. Zhang, *Elementary of Modern Algebra*, People's Education Press, China, 1978.
- [123] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical Methods for Physics and Engineering*, Cambridge Press, 2006.
- [124] I. N. Bronshtein, K. A. Semendyayev, G. Musiol, and H. Muehlig, *Handbook of Mathematics*, Springer, 2007.
- [125] J. Stillwell, *Mathematics and Its History*, Springer, 3rd edition, 2010.
- [126] D. C. Smith, "An introduction to distribution theory for signals analysis—part 2. The convolution," *Digital Signal Processing*, vol. 16, no. 4, pp. 419–444, 2006.
- [127] J. Lighthill, *An Introduction to Fourier Analysis and Generalised Functions*, Cambridge University Press, 1958.
- [128] R. P. Kanwal, *Generalized Functions: Theory and Applications*, Birkhauser, 3rd edition, 2004.
- [129] I. M. Gelfand and K. Vilenkin, *Generalized Functions*, vol. 1, Academic Press, New York, NY, USA, 1964.
- [130] A. D. Aleksandro, *Mathematics, Its Essence, Methods and Role*, vol. 3, USSR Academy of Sciences, 1952.
- [131] V. I. Istratescu, *Introduction to Linear Operator Theory*, Marcel Dekker, New York, NY, USA, 1981.
- [132] M. Li and W. Zhao, "Sufficient condition for min-plus deconvolution to be closed in the service-curve set in computer networks," *International Journal of Computers*, vol. 1, no. 3, pp. 163–166, 2007.