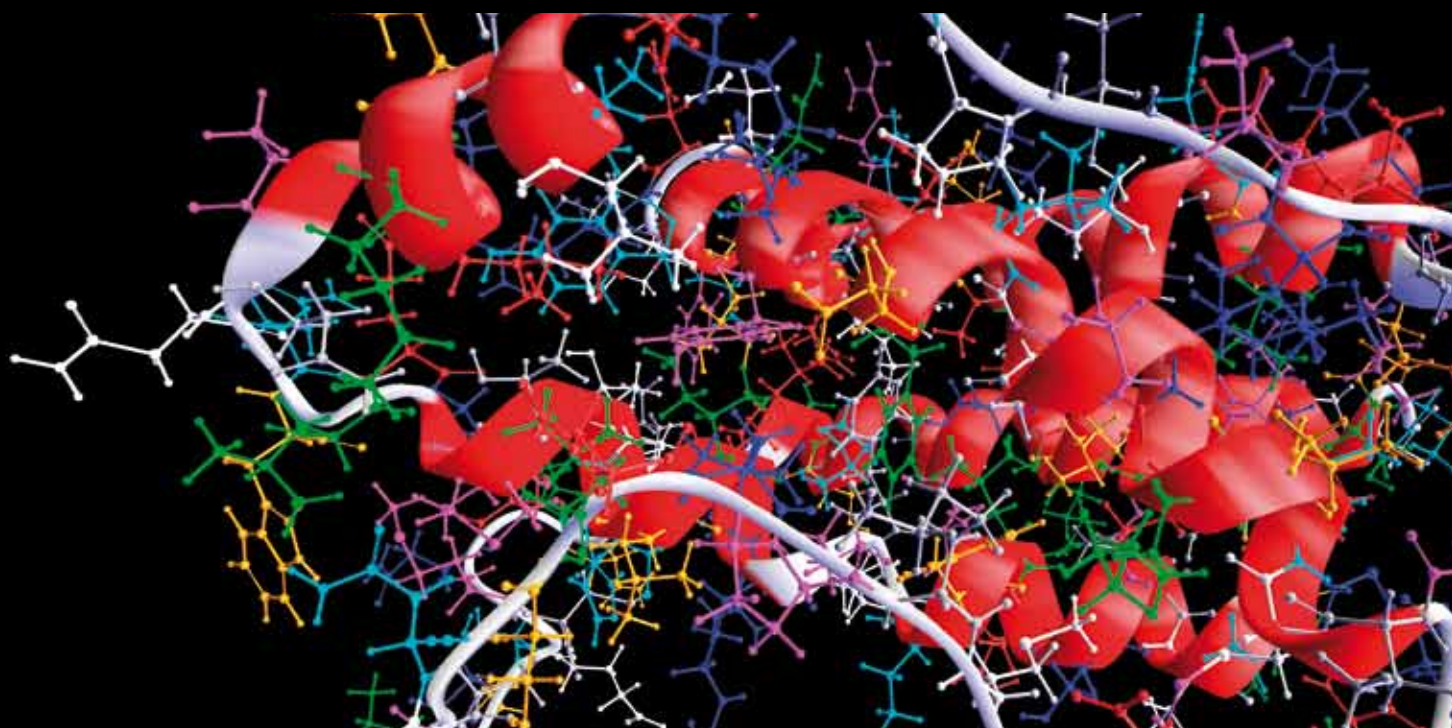# Integrative Approaches in Computational Biomedical Imaging

Guest Editors: Huafeng Liu, Pengcheng Shi, and Yunmei Chen

Integrative Approaches in Computational Biomedical Imaging

Integrative Approaches in Computational Biomedical Imaging

Guest Editors: Huafeng Liu, Pengcheng Shi, and Yunmei Chen

Editorial Board

Contents

Editorial

Integrative Approaches in Computational Biomedical Imaging

Huafeng Liu,[1] Pengcheng Shi,[2] and Yunmei Chen[3]

[1] *State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China*
[2] *B. Thomas Golisano College of Computing and Information Sciences, Rochester Institute of Technology, Rochester, NY 14623, USA*
[3] *Department of Mathematics, University of Florida, 458 Little Hall, Gainesville, FL 32611-8105, USA*

Correspondence should be addressed to Huafeng Liu, liuhf@zju.edu.cn

Increasingly wider availability of biomedical imaging modalities, such as X-ray computed tomography (CT), magnetic resonance imaging (MRI), ultrasonic imaging, positron emission tomography (PET), single-photon emission computed tomography (SPECT), and optical imaging, have led to significant progresses in both biomedical research and clinical practice. Advancements in imaging technology and image informatics have facilitated and enabled quantitative understanding of the biological structures, functional organisms, and pathological mechanisms.

This special issue presents 14 latest research contributions in computational biomedical imaging. The focuses have been on the computational principles, strategies, methods, and algorithms that create, extract, and understand imaging data of biomedical significance.

The first group of papers has showcased the development and applications of rigorous optimization techniques in biomedical imaging. The work based on dual unscented Kalman filtering provides a reliable and efficient platform to understand the BOLD signals from fMRI. In the other paper, the robust discrete particle swarm optimization algorithm is adopted to solve the branch-cut phase unwrapping problem of MRI data. Without the need for a regularization penalty term, a semigreed method for bioluminescence tomography reconstruction is fast and stable for small animal imaging.

The second group of papers has addressed image segmentation and tissue characterization problems, ever popular among image analysis researchers. The work of L. Li et al. applies machine learning techniques to characterize lymph node metastasis in gastric cancer from gemstone spectral imaging with much improved outcomes, while G. V. Sanchez. Ferrero et al. make new advances in ultrasonic speckle characterization using generalized Gama distribution and generalized Gama mixture model.

In recent years, physically meaningful modeling and simulation have become integral components for model-constrained imaging and model-based image understanding. Thus, the third group of papers has continued the stride in this direction. The propagation of myocardial electrical activation is simulated using a monodomain model, without the need of explicit mesh constraints. In similar spirit, one work performs forward modeling of 3D optical molecular imaging with simplified spherical harmonics approximation using extended finite element method.

The final group of papers has highlighted some new imaging modalities and their potential use in clinical practice. The further development of ultrasound-based quantitative imaging of Young's modulus with high accuracy is presented in this issue. And the novel Diffusion Kurtosis Imaging shows that more detailed classification of brain tissues is clinical significance.

We appreciate the many high quality submissions to this special issue from leading researchers in computational biomedical imaging. Because of space limitation, we can only accept a small number of most exciting works. We sincerely hope that the remaining interesting research outcomes will be soon published in other venues.

Huafeng Liu
Pengcheng Shi
Yunmei Chen

Research Article

Numerical Study for Blood Flow in Pulmonary Arteries after Repair of Tetralogy of Fallot

Ming-Jyh Chern,[1] Ming-Ting Wu,[2,3] and Sheau-Wei Her[1]

[1] *Department of Mechanical Engineering, National Taiwan University of Science and Technology, Taipei 10607, Taiwan*
[2] *Department of Radiology, Kaohsiung Veterans General Hospital, Kaohsiung 81362, Taiwan*
[3] *Faculty of Medicine, School of Medicine, National Yang-Ming University, Taipei 112, Taiwan*

Correspondence should be addressed to Ming-Ting Wu, wu.mingting@gmail.com

Pulmonary regurgitation (PR) is a common phenomenon in pulmonary arteries in patients after repair of tetralogy of Fallot (TOF). The regurgitation fraction of left pulmonary artery (LPA) is usually greater than right pulmonary artery (RPA) according to clinic data. It may be related to blood flow in pulmonary arteries. Therefore, understanding hemodynamics in pulmonary arteries helps to comprehend the reason. The aim of this study is to use 3D reconstructed pulmonary artery models from magnetic resonance imaging (MRI) and to use numerical approaches for simulation of flow variations in pulmonary arteries after repair of TOF. From the numerical results, the blood flow is influenced by the bifurcation angles and geometry of pulmonary artery. The regurgitation happens first in LPA after repair of TOF due to the small angle between LPA and main pulmonary artery (MPA). The recirculation region which obstructs forward blood flow to the left lung is found in LPA during acceleration of systole. We also analyze the pressure distribution; the extreme pressure variations are in dilation area of MPA. Numerical data including regurgitation in MPA, LPA, and RPA are compared with phase contrast MR measured data. Good agreements are found between numerical results and measured data.

1. Introduction

Tetralogy of Fallot (TOF) is the most common congenital heart disease, accounting for 10% of all common congenital heart diseases. TOF is composed of (1) ventricular septal defect, (2) pulmonary stenosis, (3) right ventricular hypertrophy, and (4) overriding aorta. Surgerical correction is the fundamental treatment. However, the surgical reconstruction of TOF usually complicates with pulmonary regurgitation (PR) phenomena. PR is caused by incomplete close and deformity of the pulmonary valve after surgical reconstruction of the right ventricular outflow track in repaired TOF. It causes complex blood flow in pulmonary arteries, which is related to right ventricular dysfunction. Phase contrast magnetic resonance (PC-MR) technology is an established noninvasive technique to estimate blood flow in the vessels and has been used to investigate the PR phenomena in repaired TOF patients. According to Kang et al.'s [1] PC-MR measurements of repaired TOF, it was found that the regurgitation fraction (RF, i.e., backward flow volume/forward flow volume) of PR in left pulmonary artery (LPA) is higher than right pulmonary artery (RPA) in repaired TOF. The reason causing this discrepancy of RF is still not clear. We hypothesize that the surgically created structure-fluid dynamical interaction is one of the possible reasons. In this study, we utilize numerical modeling to investigate the relation between the structure, fluid dynamic, and PR in repaired TOF.

Ho et al. [2] reported many complications in 21 TOF patients after surgery according to their cardiac MR measurements, for example, severe pulmonary regurgitation (PR), right ventricular dilatation, right ventricular outflow obstruction, ventricular septal defect patch leakage, and arrhythmias. Kang et al. [1] discussed the difference of PR between left and right pulmonary arteries from PC-MRI measurements for 22 patients after repaired TOF surgery. They found that the regurgitant fraction in LPA is greater than that in RPA and MPA in most of repaired TOF

patients. Meanwhile, Kang et al. [1] also indicated that the relationship between the regurgitant fraction and fraction of regurgitant flow duration in MPA and RPA is linear. Wu et al. [3] investigated the effect of pulmonary regurgitation on perfusion of the lungs in 43 patients after repaired TOF using PC-MRI. They calculated forward flow volume, backward flow volume, net flow volume, and regurgitation fraction in the left and right pulmonary arteries. Comparison of the perfusion between the left lung (L%) (L% = left lung/left + right lung) was provided in their PC-MRI measurements. In terms of the comparison, they also found that the PR phnomenon of LPA is higher than RPA in repaired TOF patients.

Helbing et al. [4] explored the abnormalities in right ventricular (RV) diastolic function with 19 children after repair of TOF and 12 healthy volunteers by PC-MR technique. They found that right ventricular end-diastolic volume (RVEDV) of repaired TOF patients is larger than a healthy person's. Furthermore, it was revealed that right ventricular ejection fraction of repaired TOF is lower than healthy children. Frigiola et al. [5] demonstrated that RV systolic function in 124 patients with surgically treated TOF is affected by PR. In all patients, right ventricular systolic function of repaired TOF patients is poorer than healthy children. RV isovolumic acceleration is inversely proportional to PR in Frigiola et al.'s results. van den Berg et al. [6] indicated that some risk factors including RV abnormal dilation, long QRS duration, and severe PR are associated with RV dysfunction in terms of PC-MRI measurements and electrocardiogram (ECG) of 59 patients after repair of TOF. Grothoff et al. [7] described that PR is an important factor influencing prolongation of QRS duration in ECG measurements for 67 patients after repair of TOF. They indicated that PR phenomena usually accompany QRS prolongation.

Pulmonary valve replacement (PVR) has been shown to improve ventricular function, stabilize QRS duration, and reduce atrial and ventricular arrhythmias in the following studies. For instance, Vliegen et al. [8] confirmed that PVR can reduce regurgitation and shunting defects in 26 patients after repaired TOF using MRI. Their results showed that the mean preoperative PR was 46 ± 10%. 80% patients had no residual PR after PVR. They also found that right ventricular end-diastolic volume decreased from 305 ± 87 to 210 ± 62 mL and RV end-systolic volume also decreased from 181 ± 67 to 121 ± 58 mL. van Huysduynen et al. [9] showed that PVR can reduce QRS duration from 150 to 140 ms according to their cardiac MR and ECG measurements for 26 patients after repair of TOF. Sung et al. [10] used particle flow visualization experiment to investigate the effects of varying degree of pulmonary valvular stenosis in transparent glass models. Porcine pulmonary arteries and valve were considered to fabricate those glass models. Dilatation existed in MPA and LPA of the glass models. They found that significant secondary flows appear in MPA due to its dilatation. They also observed that strength of the secondary flows in the LPA and RPA increased as the degree of valvular stenosis increased. Their results proved that the geometry of MPA with dilatation and pulmonary valvular stenosis have remarkable effects on the pulmonary artery hemodynamics.

The variation of blood flow with pulmonary regurgitation in pulmonary arteries is an important factor for evaluation of TOF defects. Computational fluid dynamics (CFD) is a useful and noninvasive method to calculate blood flow in arteries. In recent years, CFD is commonly applied to simulate cardiovascular complex flow. It provides numerical solutions for accurate assessment of congenital heart disease treatments. In addition, applications of CFD on simulations of blood flow in pulmonary arteries can be found in several technical papers. For example, Tang et al. [11] used CFD to compare flow phenomena of two various pulmonary artery models, spiral and Lecompte, (nonspiral) which are usually used in arterial switch operation. They explained that the spiral method is better than the Lecompte method in terms of numerical results which reveals that the spiral method has more uniform velocity distribution, wall shear stress, and less power loss ratio than the Lecompte model. Corno and Mickaily-Huber [12] compared two different pulmonary artery models, circular and elliptical models, of distal conduit anastomosis by CFD. Their study proved that the anastomosis cross-sectional area has a great impact on blood flow distribution in pulmonary artery. They suggested that the elliptical anastomosis may be useful to improve the morbidity and degree of distal stenosis at clinical applications. Chern et al. [13] established *in vitro* pulmonary artery models after repair of TOF and observed the flow patterns with varying PR in pulmonary arteries after repair of TOF patients by CFD simulations. They also confirmed that the PR fraction in LPA is higher than in RPA and in LPA. Also, they found that vortices move toward the pulmonary valve during regurgitation in numerical results. The vortex motion may damage the pulmonary valve.

Recently, many studies applied medical images to reconstruct 3D arterial peripheral models for undertaking accurate CFD simulations. Redaelli et al. [14] detailed reconstruction of 3D vascular numerical models from magnetic resonance images and performed CFD simulations using the reconstructed vascular models. Moreover, Tang et al. [15] combined MRI and CFD to analyze blood flow in normal pulmonary arteries under resting and exercise states. They reported wall shear stress (WSS), oscillatory shear index (OSI), and the variation of energy efficiency due to exercise in pulmonary arteries. They found the low mean WSS regions in the distal pulmonary arteries no matter in rest or exercise conditions. High OSI values occurred in those low mean WSS regions and branching locations where swirling flow and separation flow were observed. They discovered the energy efficiency average decrease of 10% between rest and exercise conditions. They concluded that this approach is useful for investigation of the disease development and applications of surgical planning. Das et al. [16] calculated the total energy of the MPA to assess the RV performance between after repaired TOF and normal pulmonary arteries using MRI and CFD. Their results indicated that the repaired TOF RV has lower stroke work than the normal one. They also computed the net energy transferred at the MPA that in normal case has higher net energy than repaired TOF case.

CFD has been applied to surgical planning for pulmonary arteries as well (see [17, 18]). For example, Hsia et al. [19] studied the influence of various inferior vena cava connected to pulmonary arteries on blood flow using CFD simulations. They considered four basic TCPC geometries which were reconstructed from angiographic measurements. They demonstrated that the anastomosis in which extracardiac conduit is connected with left pulmonary artery has the lowest energy loss among four various TCPC cases. They also compared energy dissipation in five graft extracardiac conduits with various diameters (10~30 mm) by CFD. They found that the least energy dissipation occurs in the conduit of diameter 20 mm. They confirmed that the geometry of the surgically created pathway in the TCPC is very important in terms of energy dissipation. Sun et al. [20] investigated the influence of antegrade pulmonary blood flow on bidirectional cavopulmonary anastomosis using CFD. They considered four mean flow rates (0.5 L/min, 1 L/min, 1.5 L/min, and 2 L/min) of MPA. They found that the flow ratio of LPA/RPA increased when the amount of antegrade pulmonary blood flow increased. Increasing antegrade pulmonary blood flow may cause significant different blood flows to two lungs. They also observed that blood flow into LPA from MPA is larger than into RPA because the angle between MPA and LPA is larger than between MPA and RPA.

Furthermore, Pekkan et al. [21] showed that the flow patterns at the normal fetal aortic arch and pulmonary artery, which were obtained by CFD and *in vitro* experiment methods, were similar to each other. They found the swirling flow at the pulmonary artery during the deceleration phase of a cardiac cycle both in the numerical results and in the experimental flow visualization. Wang et al. [22] utilized CFD technique and flow visualization experiments to compare two types of TCPC including intraatrial connection and extracardiac inferior vena-cava- (IVC-) to-MPA connection. Their results showed that more complex unsteady flow structures occur in intra-atrial connection model in experimental and computational results.

According to those studies mentioned previously, we know the PR always happens in a repaired TOF patient and the blood flow in LPA is less than in RPA from MRI statistical analysis. It is found that the blood flow distribution highly depends on the geometry of vascular vessels. Nevertheless, variation of blood flow in real pulmonary arteries after repair of TOF is not clear. Meanwhile, numerical prediction of blood flow becomes popular in investigating the relationship between hemodynamics and cardiovascular diseases as mentioned in previous paragraphs. Hence, the aim of this study is to observe the blood flow distribution and to discuss the effect of geometry on PR in pulmonary arteries by CFD simulation coupled with 3D reconstructed MRI pulmonary artery models. Furthermore, to find the reason to cause higher PR in LPA than in RPA, numerical simulations are performed to provide distribution of pressure, mass flow rate, regurgitation fraction, and blood flow streamlines in pulmonary arteries after repair of TOF.

TABLE 1: Parameters of pulmonary arteries.

Case	Normal	Case 1	Case 2	Case 3
L_{RPA} (cm)	5.23	3.99	4.93	5.27
L_{LPA} (cm)	2.71	3.60	2.36	2.39
L_{MPA} (cm)	3.85	5.31	4.17	3.14
d_{RPA} (cm)	1.39	1.40	1.13	1.21
d_{LPA} (cm)	1.36	1.30	1.09	1.68
d_{LPA}/d_{LPA}	0.98	0.92	0.96	1.39
θ_1 (degree)	125°	134°	106°	136°
θ_2 (degree)	112°	52°	83°	70°
θ_1/θ_2	1.12	2.58	1.28	1.94
A_{Inlet} (cm^2)	3.48	4.58	4.89	5.83

2. Mathematical Formulae and Numerical Model

In this study, the branching pulmonary artery geometry obtained by MRI is utilized to establish a numerical model in computational fluid dynamics. The pulmonary artery is a Y shape bifurcate vessel including LPA and RPA. Various pulmonary arteries after repair of TOF are measured by MRI and considered to observe regurgitation phenomena. First of all, MRI measurements of pulmonary arteries after repair of TOF are reconstructed as CAD models. Subsequently, mesh generation is undertaken in those reconstructed CAD models. Blood flows are then simulated using a flow solver in those numerical models of pulmonary arteries. Details are demonstrated in the following sections.

2.1. MRI and Grid Generation. In order to establish a realistic 3D numerical model of pulmonary artery, MRI measurements are utilized to obtain peripheries of pulmonary arteries after repair of TOF and a healthy person. Regions of the branching pulmonary arteries are remarked in slice planes of MR images based on DICOM format. Every slice thickness of Cases 1–3 was 3.6 mm, 4.4 mm, and 4.4 mm, respectively. The normal one was 4 mm. Those 2D remarked regions are connected together be a complete 3D CAD pulmonary artery model. Those features of four pulmonary models are listed in Table 1 and shown in Figure 1. We found that the MPA dilation phenomenon and an acute angle between LPA and MPA appear in all reconstructed CAD models of pulmonary arteries after repair of TOF.

Subsequently, computational grids are generated inside those reconstructed CAD models using ICEM-CFD software. Since peripheries of the pulmonary arteries are too complicated to generate structured grids, unstructured grids are used in the numerical models. Figure 2 presents the procedure to reconstruct CAD models from MRI measurements and mesh generation.

2.2. Governing Equations. Blood flow obeys mass and momentum conservation. In this study, few assumptions are used to simplify equations for mass and momentum conservation. Mathematical formulae of mass and momentum are denoted as:
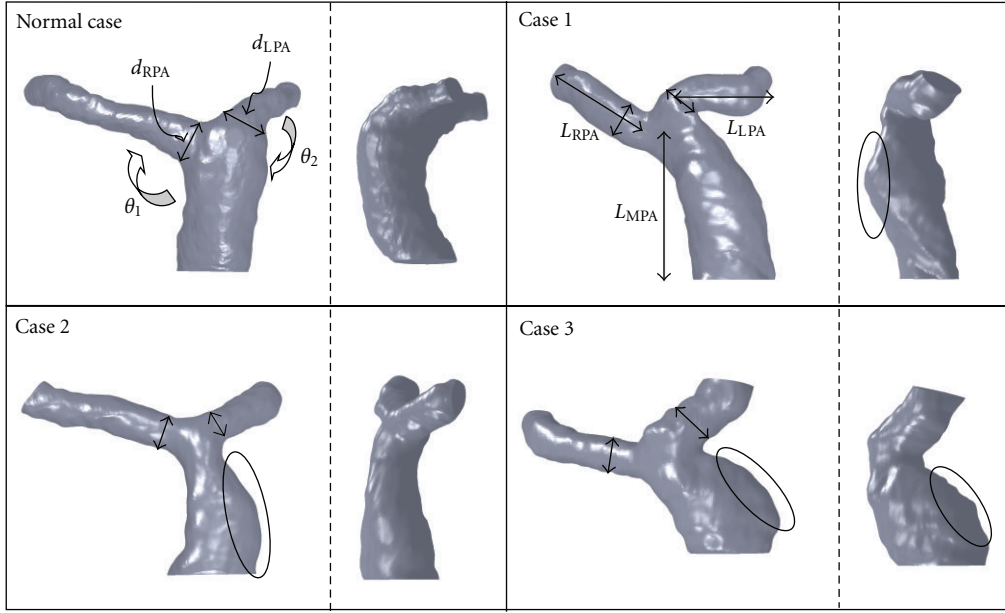
Figure 1

continuity equation:

$$\nabla \cdot \mathbf{u} = 0, \tag{1}$$

and momentum equation (Navier-Stokes equation)

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot (\mathbf{u}\mathbf{u}) = -\frac{1}{\rho}\nabla P + \nabla \cdot (\nu \nabla \mathbf{u}), \tag{2}$$

where $\mathbf{u}$ represents a velocity vector of flow, P is pressure, t is time, ρ is density of blood, and ν is kinematic viscosity of blood. The density of blood, ρ, is set as $1060\,\mathrm{kg\cdot m^{-3}}$. The dynamic viscosity (μ) of blood is set as a constant as $0.004\,\mathrm{kg\cdot m^{-1}\cdot s^{-1}}$. Although blood is non-Newtonian, viscosity of blood varies with respect to shear rate and vessel diameter. Given that shear rate is greater than $100\,\mathrm{s^{-1}}$, the blood viscosity can be regarded as a constant like a Newtonian fluid in accordance with Pedley [23] and Berger and Jou [24]. Because a pulmonary artery is a large vessel, the non-Newtonian effect is not strong. Hence, blood in pulmonary artery is considered as Newtonian fluid for simplifying equations. According to Singh et al. [25], the wall of vessels is regarded as a rigid tube. That is, compliance of pulmonary artery is not considered.

2.3. Boundary Conditions. Regurgitation occurs in the pulsatile blood flow of pulmonary arteries after repair of TOF. Regurgitation fraction, RF, is defined as the ratio of backward blood volume to forward blood volume in a cardiac cycle. A health person's RF is usually without regurgitation or very low. High RF indicates that abnormal blood flow behavior occurs in pulmonary arteries. That is, low RF means that most blood flows into lungs to exchange oxygen. RF is defined as

$$\mathrm{RF} = \frac{\text{Backward volume in a cardiac cycle, } Q_b}{\text{Forward volume in a cardiac cycle, } Q_f}, \tag{3}$$

Table 2: The regurgitation fraction (RF).

Case	RF		
	MPA	LPA	RPA
1	0.337	0.449	0.252
2	0.164	0.186	0.149
3	0.288	0.304	0.260

where Q_b and Q_f refer to backward and forward blood flow volumes in a cardiac cycle, respectively.

Figure 3 shows boundaries of the established model. To solve those equations, appropriate boundary conditions are required. Blood flow velocity profiles of individual patients were measured by phase contract MRI at the inlet boundary of the MPA entrance. The PC-MR image matrix was 256 pixels × 256 pixels. The image voxel sizes of Cases 1–3 were 0.141×0.141, 0.078×0.078, and 0.082×0.082, respectively. The normal case was 0.0625×0.0625. The measured nonuniform flow is provided for the inlet condition. The interpolation method is used to map the inlet velocity profile to inlet grid points. Also, RFs of Cases 1–3 are 0.337, 0.164, and 0.288 as shown in Table 2, respectively.

The atmospheric pressure is imposed at the exits of the RPA and LPA. The wall of vessels is regarded as a rigid object. Nonslip boundary conditions are imposed in vessel wall. That is, velocity is zero in the wall. For the initial conditions of the pulmonary artery, velocity is set as zero.

2.4. Parameters Setting. The software, CFD-ACE+, is based on a finite volume method that is used to calculate the blood flow in a pulmonary artery. The finite volume method has been applied to solving the Navier-Stokes equations in many engineering applications. For example, Georgios and Ioannis [26] used the finite volume method in computation of radiative heat transfer. Kabinejadian and Ghista [27] employed the
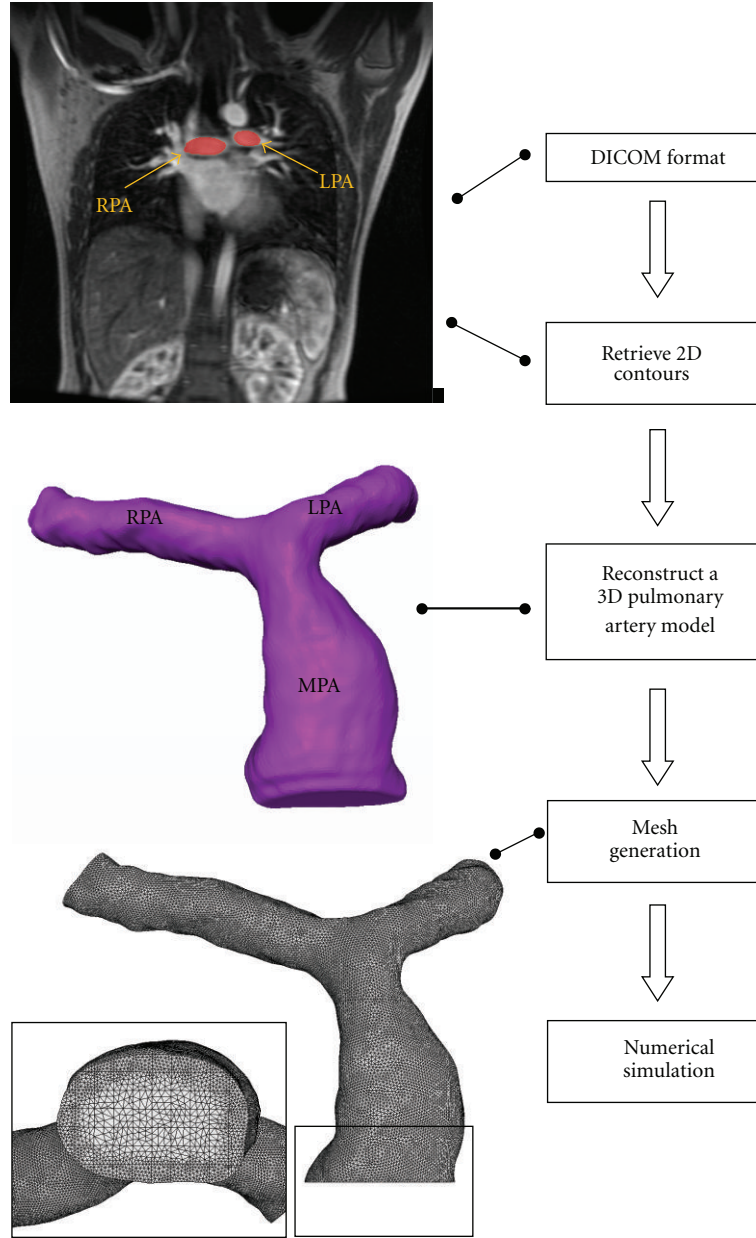
FIGURE 2: Flow chart of the reconstructed pulmonary artery model.

method to solve the NavierStokes equations for simulations of blood flow in a coronary arterial bypass graft. Details of the finite volume method for solving the Navier-Stokes equations can be found in Versteeg and Malalasekera [28]. Computational grids mentioned in Section 2.1 are utilized to discretize the Navier-Stokes equations. As a result, a set of simultaneous algebraic equations for velocity and pressure are obtained and solved implicitly. To observe complex blood flow patterns in pulmonary artery, transient simulations are undertaken. The second-order Crank Nicholson scheme is employed to undertake the time marching procedure. The time step size ranges from 0.01 to 0.015 s. Four cardiac cycles are conducted in each numerical simulation. The total simulation time is different in each case because the duration of a cardiac cycle is different in each person. The

mass residual tolerance for numerical solutions is set to 10^{-4} at each time step. The SIMPLEC scheme proposed by van Doormaal and Raithby [29] is used to obtain pressure solutions. The Window-based PC cluster is utilized to perform numerical simulations in this study. The average total calculation time for each case is about 1~2 months.

2.5. Grid Independence. To ensure that the numerical solutions do not vary with the grids, the grid-independence tests must be performed. For example, Figure 4 shows that the velocity profiles obtained by various computational cells are close to each other in Case 1 when the number of the cells exceeds 1110784. Therefore, we adopt 1110784 cells to conduct the following numerical simulations in Case 1. Moreover, we chose the 1168920 and 1035407 cells for
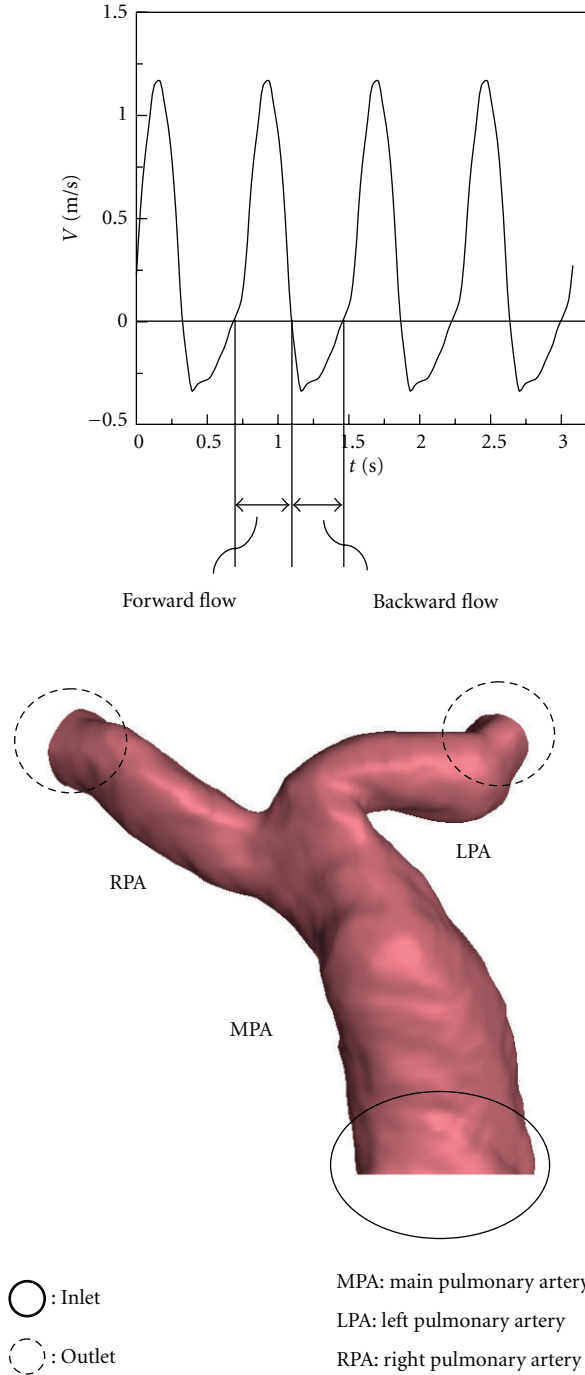
FIGURE 3: Boundary conditions.

Cases 2 and 3 according to results of grid-independence tests, respectively. 1079490 cells are used in the healthy case.

3. Results and Discussion

3.1. Influences of Bifurcation Angles and Geometry of Pulmonary Artery on Blood Flow. Pulmonary regurgitation is a common phenomenon in patients after repair of TOF, but its effect on pulmonary arteries is not clear. The established

numerical pulmonary artery models include one healthy and three pulmonary arteries after repair of TOF. Table 2 shows that Cases 1–3 have smaller angles between LPA and MPA than a healthy one. θ_2 indicating the angle between LPA and MPA is 52°, 83°, and 70° in Cases 1–3, respectively. Nevertheless, θ_2 in a normal pulmonary branch is 112°. On the other hand, angles between RPA and MPA in Cases 1–3 are larger than the normal one. θ_1 indicating the angle between RPA and MPA in Cases 1–3 are 134°, 106°, and 136°, respectively. The angle of the normal one is 125°.
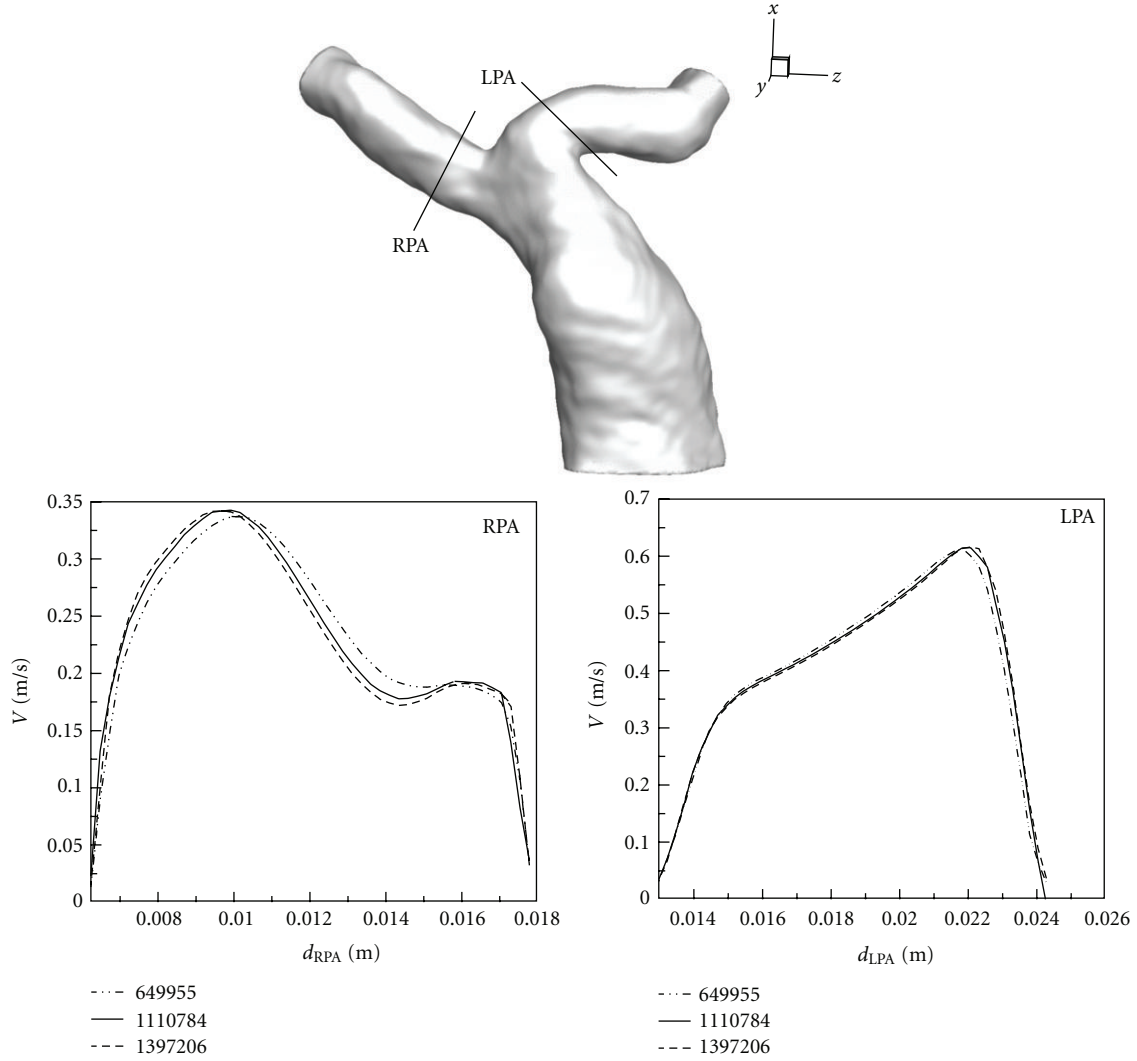
FIGURE 4: Grid-independence test.

Figure 5 describes streamlines of the healthy case in a central vertical sectional plane within a cardiac cycle. There is no regurgitation in healthy pulmonary arteries. Figure 5(a) shows the flow pattern in the acceleration of systole. A smooth forward flow pattern is found in this stage. Subsequently, two recirculations occur in LPA and RPA due to the deceleration of the forward flow in Figure 5(b). Those two vortices grow and stay in entrances of LPA and RPA in diastole of a cardiac cycle as shown in Figures 5(c) and 5(d).

Figure 6 is utilized to compare flow patterns of the healthy one and Cases 1–3 in acceleration of systole and maximum in a cardiac cycle. It is found that a recirculation occurs in LPA in Cases 1–3 but not in the healthy one. It is because θ_2 of Cases 1–3 is smaller than the healthy one. As a result, separation happens in LPA even though the forward flow in MPA is accelerated. The recirculation in LPA reduces the blood flow volumetric rate to the left lung. Figure 7 (n-3), (1-3), (2-3), and (3-3) present streamlines of the healthy one and Cases 1–3 in deceleration of systole in a cardiac cycle. The recirculation bubble in Cases 1–3 grows. As mentioned in the previous paragraph and shown in Figure 7 (n-3), there

is also a recirculation appearing in the healthy one during deceleration of systole. At the end of systole, very complex flow patterns including several vortices are found in Cases 1–3 as shown in Figure 7 (1-4), (2-4), and (3-4). Subsequently, Figure 8 reveals flow patterns of the healthy one and Cases 1–3 in diastole of a cardiac cycle. Regurgitation happens in MPA of Cases 1–3. Obvious dilations of Cases 2 and 3 are found in MPA. During regurgitation in MPA, a strong vortex is formed in dilation of MPA of Cases 2 and 3. Nevertheless, there are two vortices appearing in MPA of Case 1. No vortex is found in MPA of the healthy one.

Since vortices are found in MPA of Cases 1–3 in regurgitation, it is interesting to investigate the flow structure in specified cross-sections. Figure 9 shows flow patterns of the healthy one and Cases 1–3 in specified cross-sections of MPA, LPA, and RPA in diastole. According to Perry and Steiner's [30] definitions of larger vortex in 3D flow, a stable node is found in the cross-section a-a′ of the healthy MPA. Stable focuses appear in cross-sections b-b′ and c-c′ of healthy RPA and LPA. The stable node in healthy does not occur in Cases 1–3. One or more focuses are found in

(a)

(b)

(c)

(d)
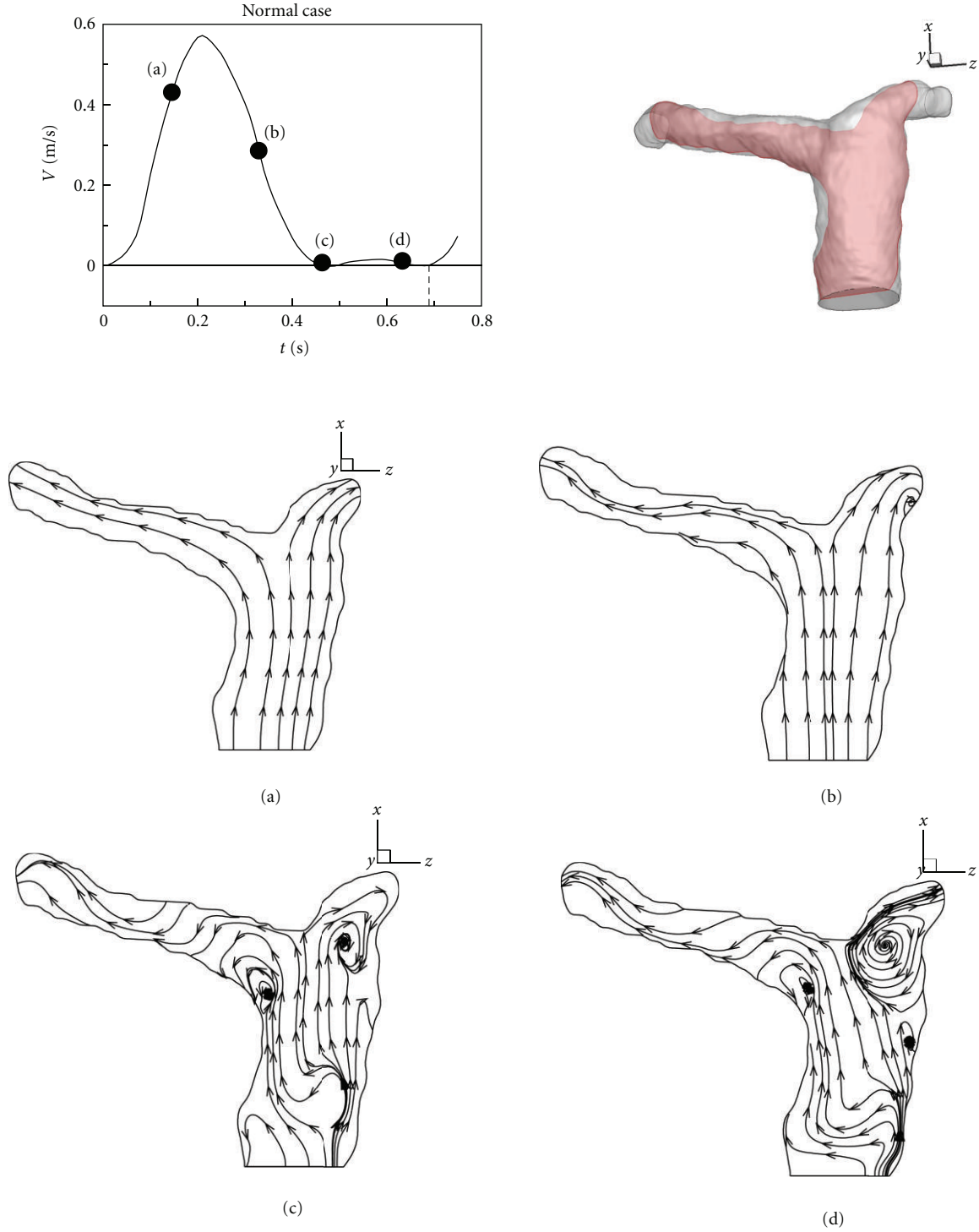
FIGURE 5: Evolution of streamlines in healthy pulmonary arteries within a cardiac cycle.

MPA of Cases 1–3. An unstable bifurcation line, a saddle point, and a stable focus happen in the cross-section d-d′ of MPA of Case 1. A stable focus and another unstable focus are observed in the cross-section i-i′ of MPA of Case 2. In Case 3, there are a saddle point and a stable focus presenting in the cross-section q-q′ of MPA. In addition, the focus appearing in the healthy RPA is found in Cases 1–3. It is stable in the healthy one and Cases 1 and 3, but it is unstable in Case 2.

The focus in healthy LPA does not appear in all Cases 1–3. The flow pattern in LPA of Case 1 shows a very strong and smooth and reversed flow. In Cases 2 and 3, the focus shrinks because of the strong reversed flow.

In order to know whether regurgitation happens first in LPA or not, Figure 10 exhibits the time of inception of regurgitation in RPA and LPA in Cases 1–3. Figures 10(a), 10(c), and 10(e) present that regurgitation first occurs in LPA
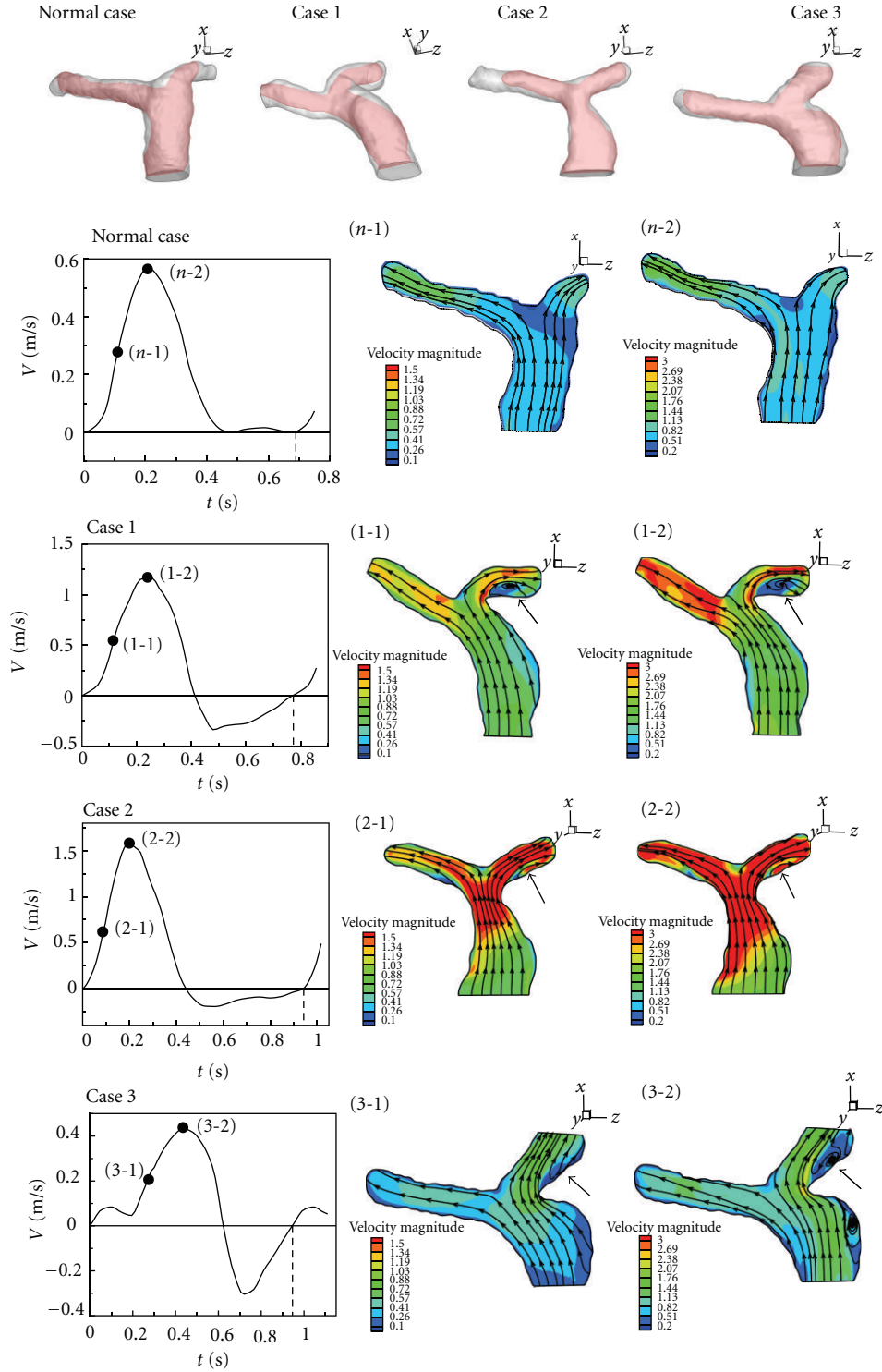
FIGURE 6: Flow patterns and velocity distribution in midacceleration of systole and maximum in a cardiac cycle.

of Cases 1–3. Also, an obvious recirculation appears in LPA of Cases 1–3. In terms of Figures 10(b), 10(d), and 10(f), it turns out that the time of inception of regurgitation in RPA is nearly 0.03 seconds behind that in LPA of Cases 1–3.

According to Figures 6, 7, and 8, the complex flow patterns always happen in regurgitation of MPA of Cases 1–3. 3-D strong vortex motion is found in this stage as shown

in Figures 7 (1-4), (2-4), and (3-4) and 8 (1-6), (2-6), and (3-6) in the ends of systole and diastole because the direction of blood flow in MPA changes. The 3D vortex motion in MPA may cause extraload in the right ventricle since the healthy case does not have regurgitation and the vortex motion.

In clinical cases, there is no dilation in MPA before repair of TOF. The dilation often occurs in MPA after repair
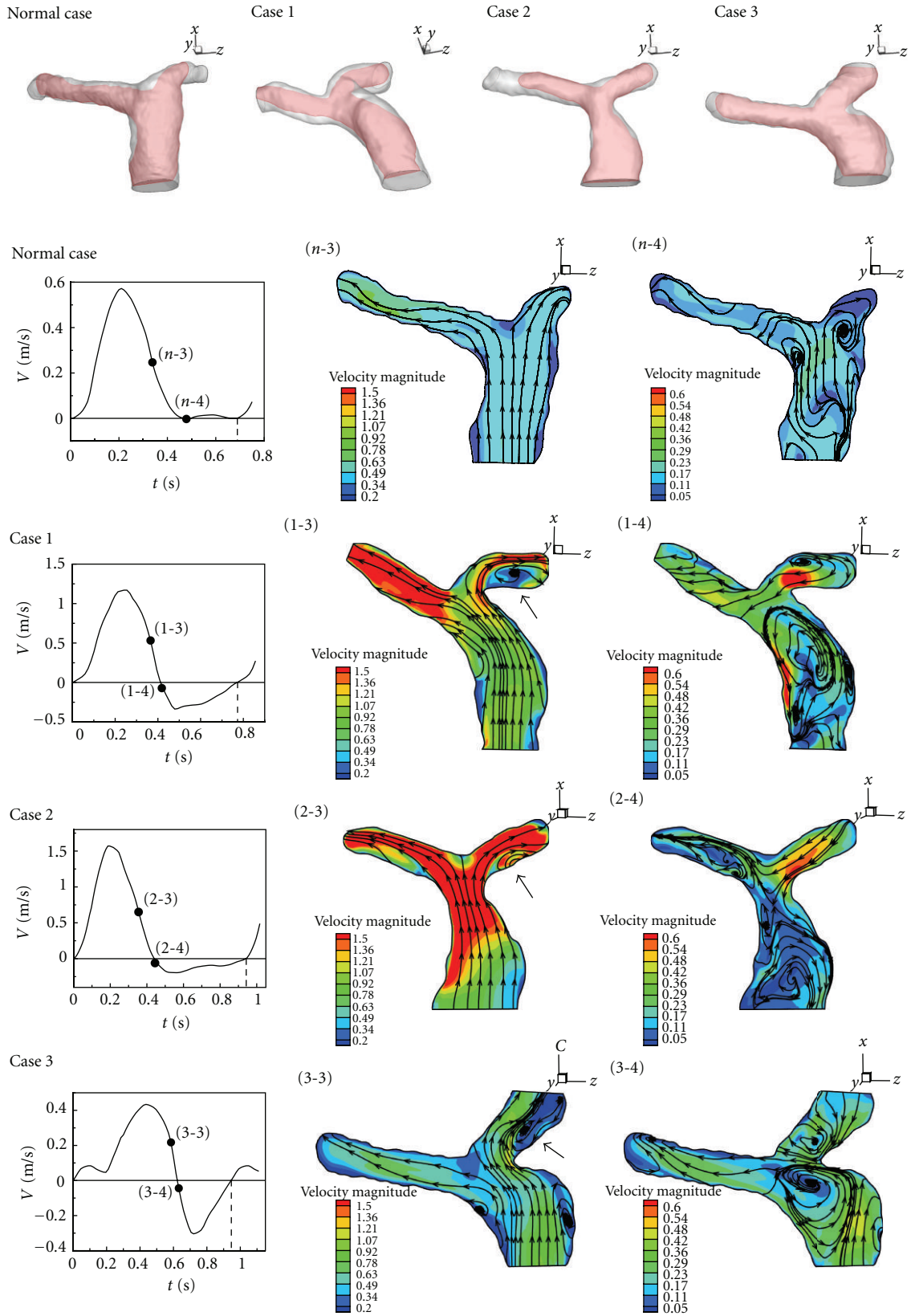
FIGURE 7: Flow patterns and velocity distribution in deceleration and end of systole in a cardiac cycle.
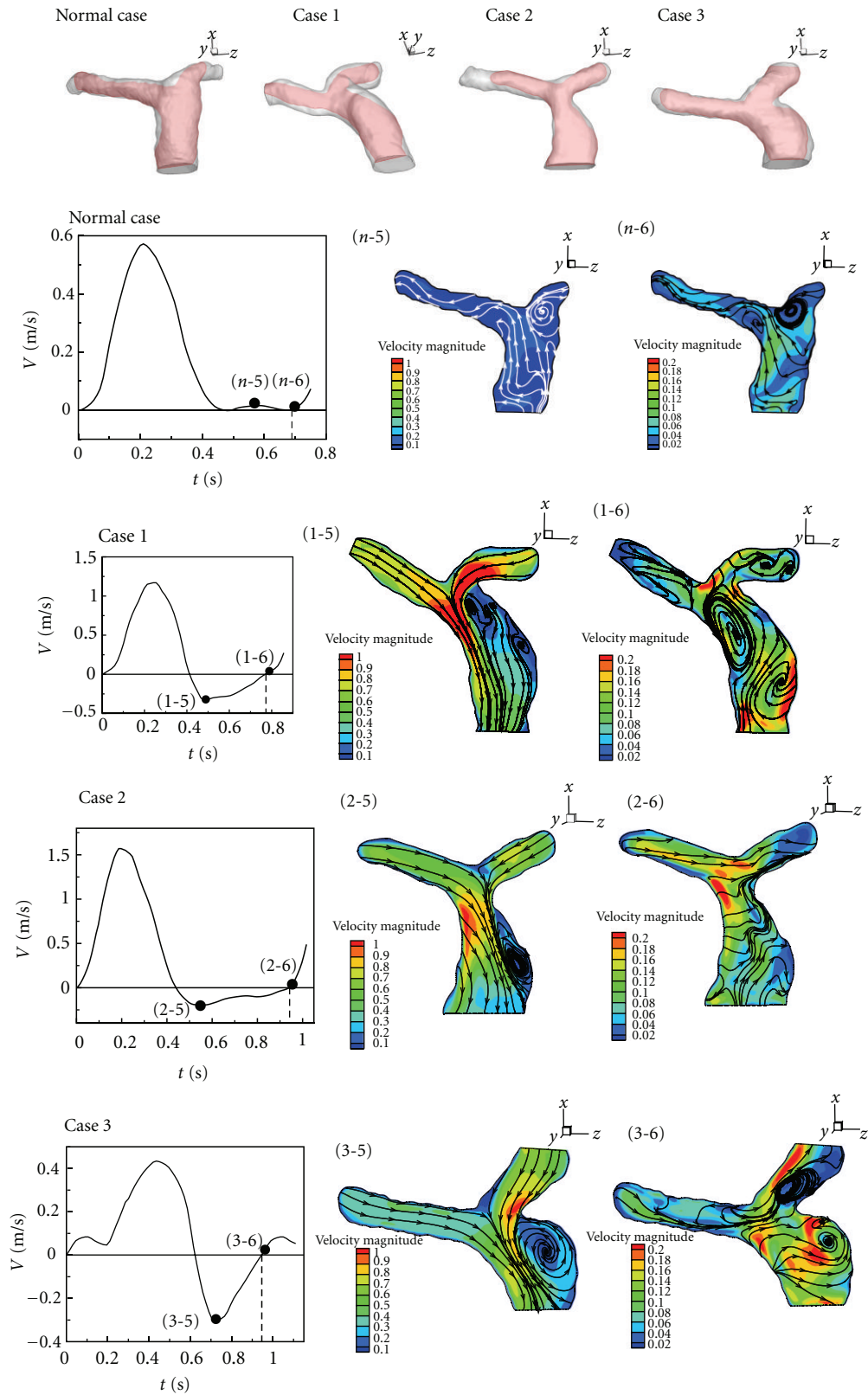
FIGURE 8: Flow patterns and velocity distribution in minimum of profile and end of diastole in a cardia cycle.
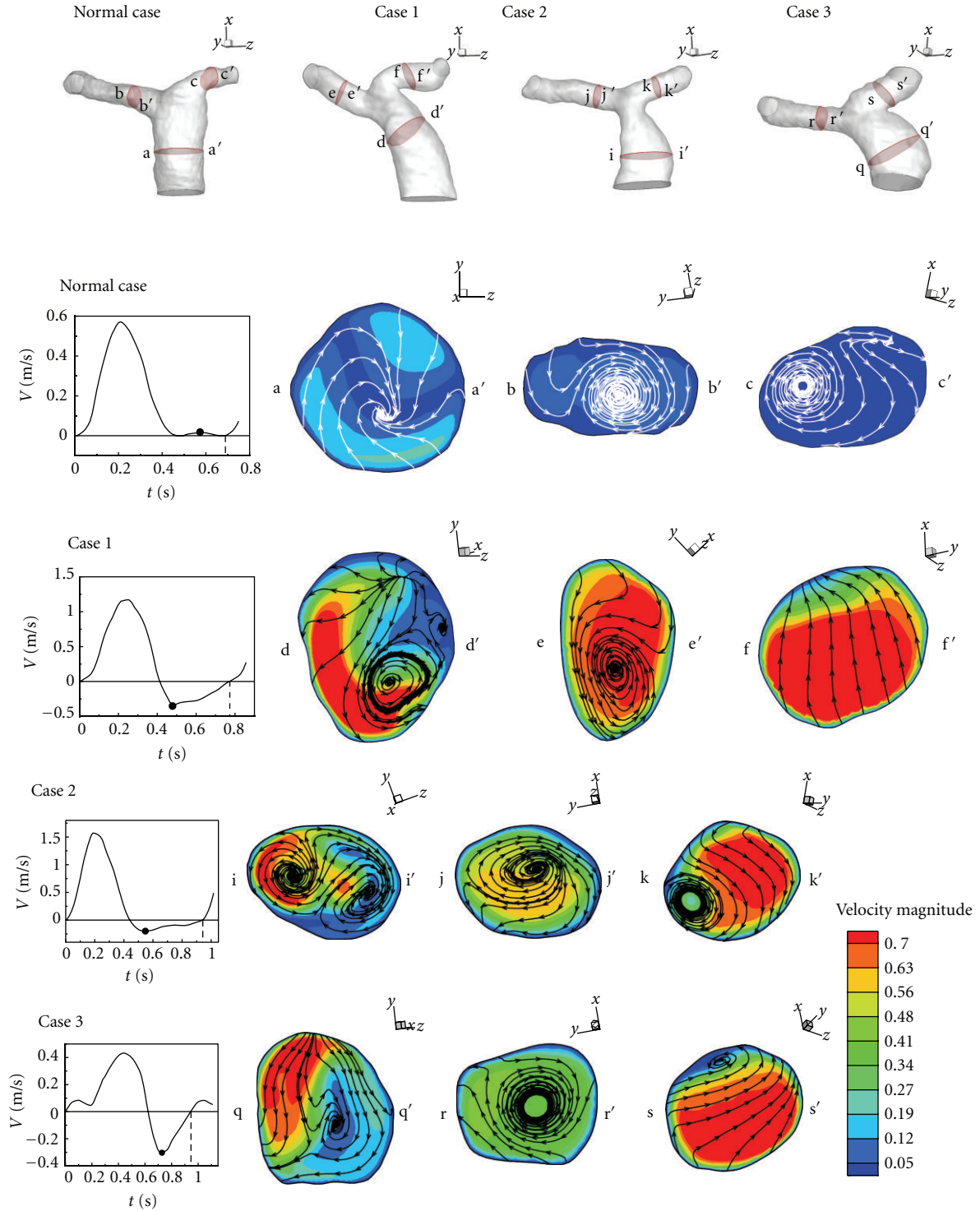
FIGURE 9: Flow patterns and velocity distribution in minimum of a cardiac cycle.

of TOF. In terms of numerical results, it is found that 3-D strong vortex motion appears in dilation of MPA in regurgitation. Therefore, the dilation causes abnormal flow distribution in MPA during regurgitation period in a cardiac cycle. Apparently, the dilation in MPA is an important factor to influence variation of blood flow in diastole.

3.2. Analysis of Pressure Distribution.
In our models, the dilation occurs in MPA of all repaired TOF cases. However, the reason causing dilation is not clear. Therefore, it is necessary to investigate the dilated area. This section gives the analysis of the pressure distribution in pulmonary arteries in a cardiac cycle. Figure 11 shows pressure distributions of
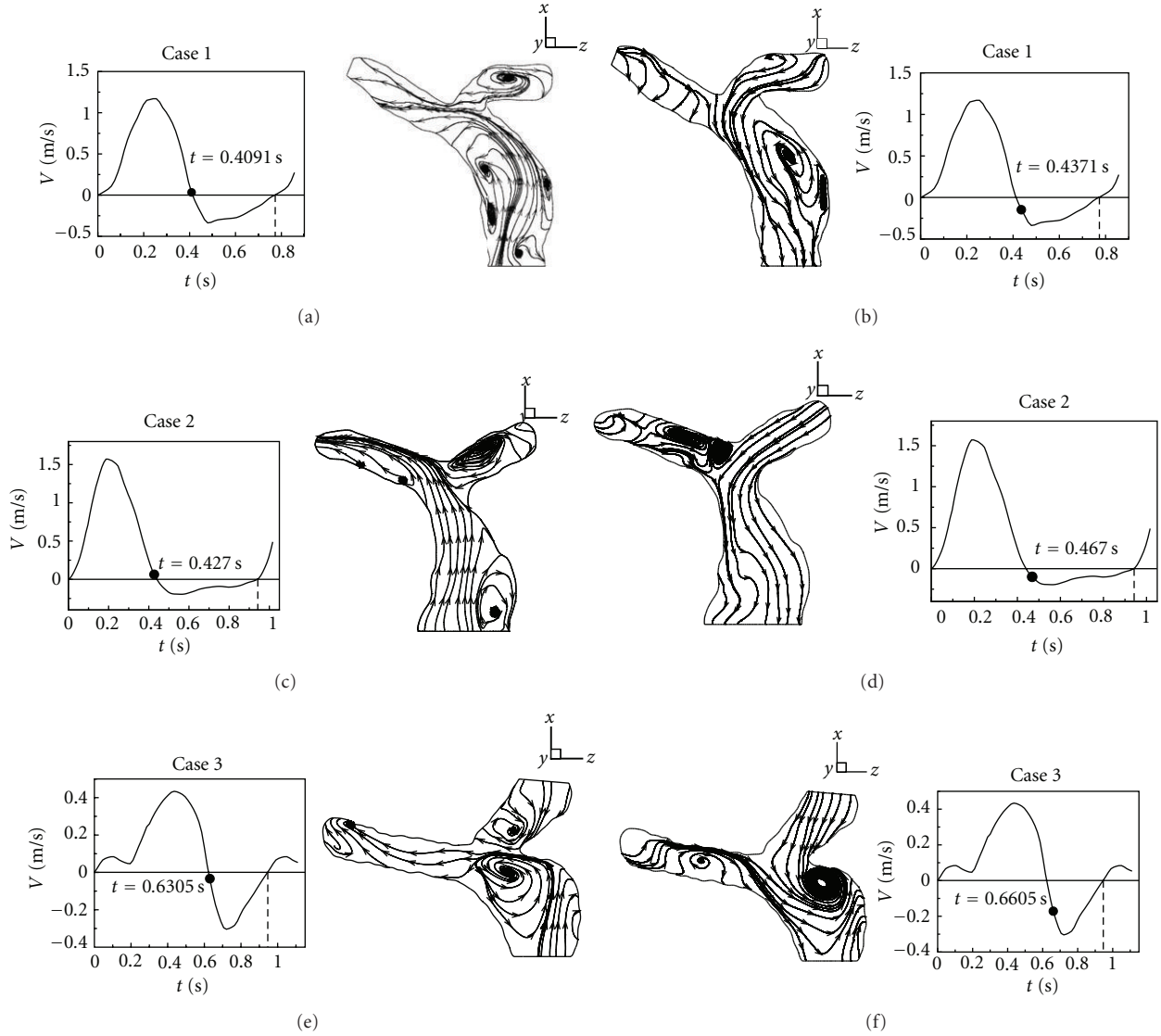
Figure 10: First time of regurgitation happening in LPA and RPA.

the healthy case and Cases 1–3 in the maximum of a cardiac cycle. A high pressure region is found in the bifurcation area for all cases. In addition, another high pressure region is observed in the dilation of MPA of Cases 1–3. Pressures distributions of Cases 1 and 2 are very nonuniform. It should be noticed that there is a stenosis between the bifurcation and MPA in Case 2, so the high pressure in the dilation of Case 2 is extreme and higher than other cases. Figure 12 displays pressure distributions of the healthy one and Cases 1–3 in diastole. A low pressure region is discovered in the dilation area of MPA in Cases 1–3. It is caused by the 3-D vortex motion as shown in Figure 8. According to Figures 11 and 12, the pressure varies extremely in the dilation of MPA in a cardiac cycle. For example, the pressure change in the dilation of Case 2 is around $12860 \, \text{N/m}^2$. Nevertheless, the maximum pressure change in the healthy one is around $2100 \, \text{N/m}^2$. The large pressure variation in MPA would affect the vessels wall in MPA after repaired TOF seriously.

In terms of Figures 11 and 12, large pressure change may cause obvious deformation of MPA. As we mentioned, the 3-D vortex motion due to regurgitation in MPA plays an important role in the low pressure region. The situation could lead to deterioration of dilation in the longer term.

Figure 13 presents time histories of inlet pressure of MPA of Cases 1–3 and the healthy one in a cardiac cycle. The solid curve presents the inlet pressure of the healthy one. The peak value of pressure is around $2000 \, \text{N/m}^2$ in systole. Very low negative pressure occurs in diastole. Dashed curves present pressure variations of Cases 1–3. It is found that the peak values of pressure of Cases 1 and 2 are higher than the healthy one in systole. Moreover, negative pressure happens in Cases 1 and 2 in the diastole since regurgitation happens. It is found that the extreme prssure variation exists at the inlet of MPA of Cases 1 and 2. The cardiac cycle of Case 3 is longer than other cases, so the systole period of Case 3 is longer than other cases as well. Due to the larger inlet cross-section of
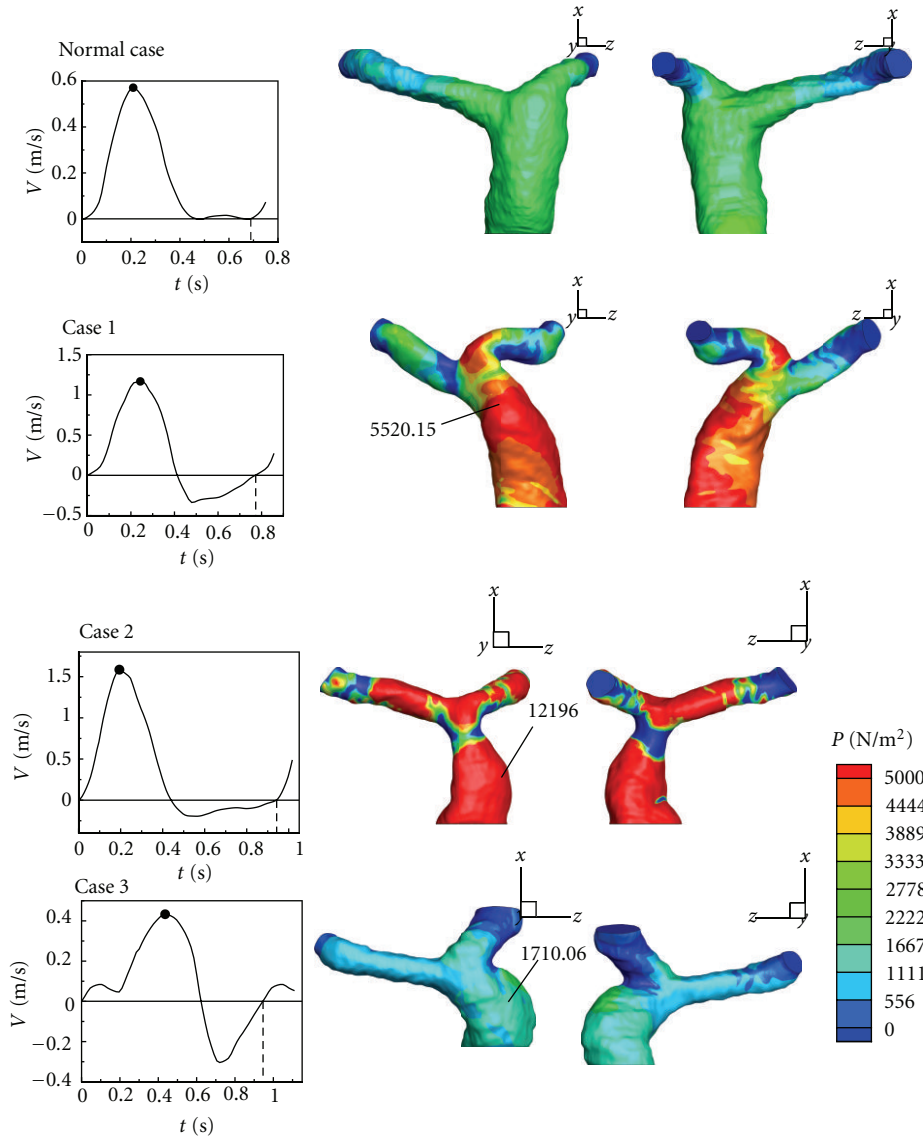
FIGURE 11: Comparing the distribution of pressure in pulmonary arteries in maximum of a cardiac cycle.

Case 3 than other cases as shown in Table 3, the peak value of inlet pressure of Case 3 is not so large as Cases 1 and 2. However, obvious negative pressure due to regurgitation exists in Case 3 in diastole. It is known that the damage of the semilunar valve is one of the complications in patients after repaired TOF. Large inlet pressure variations shown in Cases 1 and 2 may be one of the reasons to induce dysfunction of the pulmonary valve.

TABLE 3: Regurgitation volume (cm³) in a cardiac cycle.

Case	Regurgitation volume			LPA/RPA
	MPA	LPA	RPA	
1	47	27	20	1.35
2	32	17	15	1.1
3	33	24	9	2.67

3.3. Mass Flow Rate and Regurgitation. The pulmonary artery geometry effects the mass flow rate distributions in MPA, LPA and RPA. Figure 14 shows time histories of mass flow rates in LPA and RPA of Cases 1–3 in a cardiac cycle. Solid and dashed curves represent mass flow rates in RPA and LPA, respectively. Red curves show the variations of mass flow rates in the healthy one. The mass flow rates in LPA and RPA are very close in the healthy one. There is no

regurgitation in LPA and RPA. Blue, black, and green curves represent mass flow rates in Cases 1–3. Serious regurgitation is found in RPA and LPA of Cases 1–3 in diastole. Also, it is found that regurgitation always appears in LPA first in three cases. The forward flow amount of LPA is less than RPA in Cases 1 and 2. The phenomenon does not happen in Case 3. It may be because the diameter of LPA is larger than that of RPA as shown in Table 3. Meanwhile, the amount of
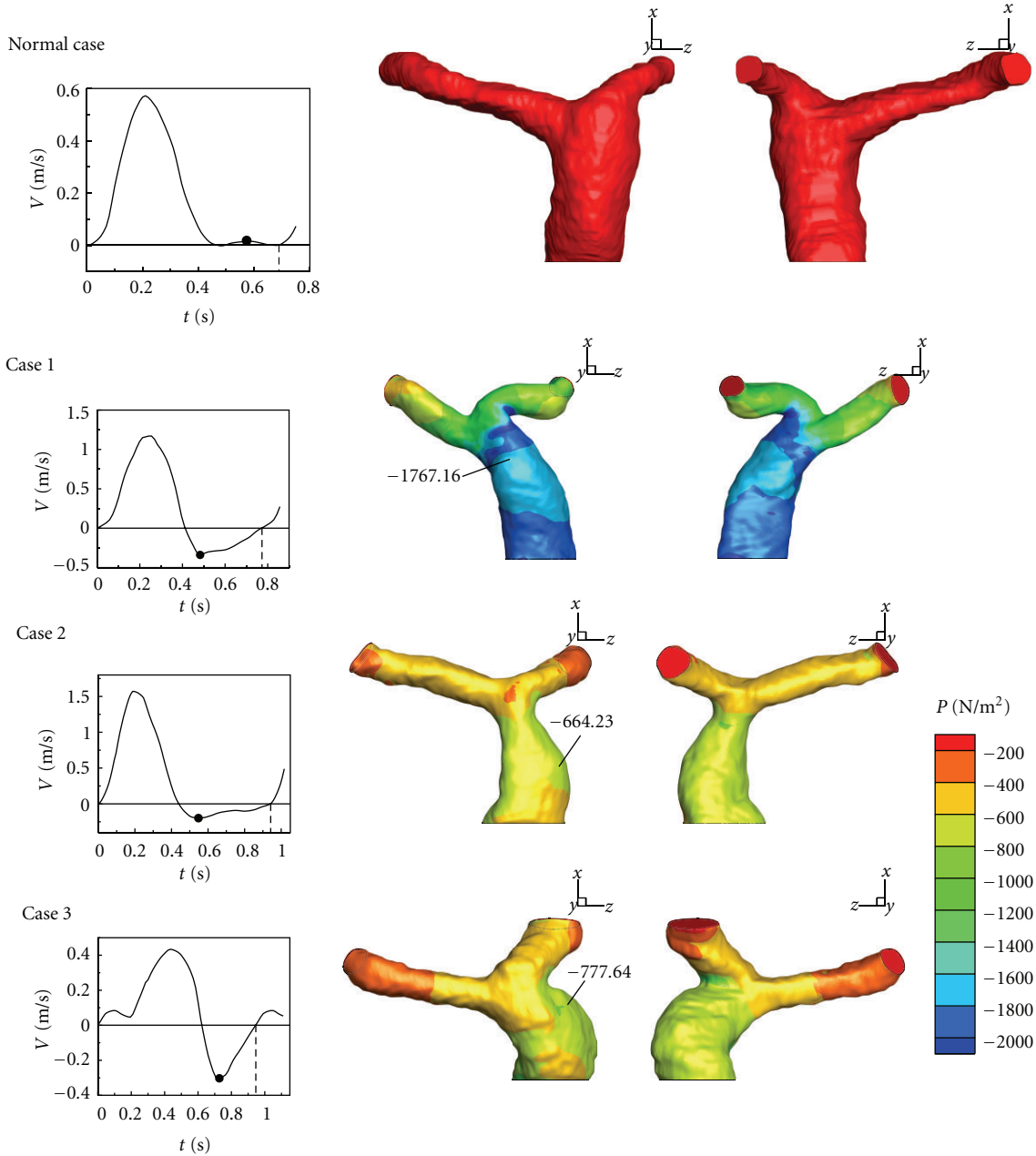
FIGURE 12: Comparing the distribution of pressure in pulmonary arteries in minimum of a cardia cycle.

reversed flow in LPA is larger than RPA in Cases 1–3. Table 3 presents regurgitation volumes of RPA and LPA of Cases 1–3 in a cardiac cycle. The ratio of regurgitation of LPA to RPA is always higher than 1. Diameters of RPA and LPA of Cases 1 and 2 are close to each other (see Table 2). Nevertheless, the ratio of θ_1 to θ_2 of Case 1 is larger than that of Case 2. As a result, the ratio of regurgitation of LPA to RPA in Case 1 is higher than Case 2.

3.4. Comparison between Numerical Results and Clinical PC-MR Measurement Data. Clinical PC MR measurement data from 34 patients after repair of TOF were provided by Kaohsiung Veterans General Hospital. Regurgitation in MPA, LPA, and RPA was determined in PC MR measurements. Subsequently, the numerical results are compared with those measurement data. Figure 15 shows that RF of MPA varies from 0.1 to 0.6. Measurement data and numerical results are denoted as black hollow and color symbols, respectively. The ratio of LPA RF to RPA RF decreases as RF of MPA increases in measurement data. It is found that numerical results are close to measurement data. Figure 16 shows the relationship between regurgitation of RPA and MPA. Essentially, the regurgitation of RPA is proportional to that of MPA in measurement data. The numerical results also show the trend in Figure 16. Figure 17 reveals the relationship between LPA RF and MPA RF. Measurement data are scattered
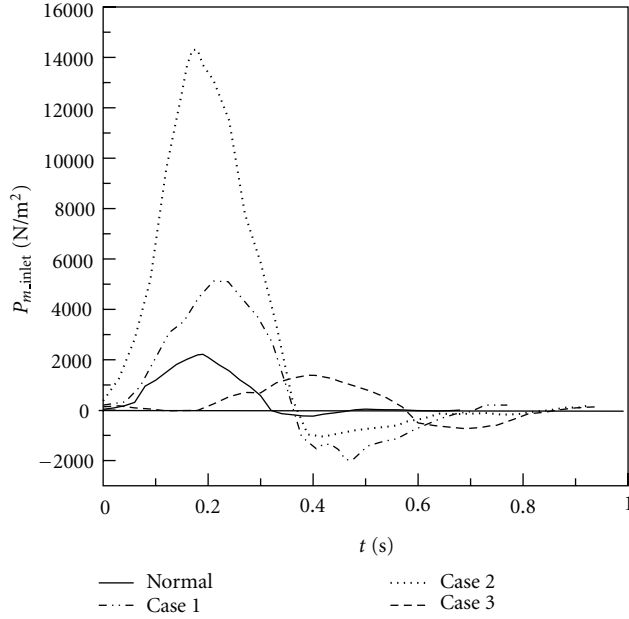
FIGURE 13: Time histories of inlet pressure of MPA in a cardiac cycle.
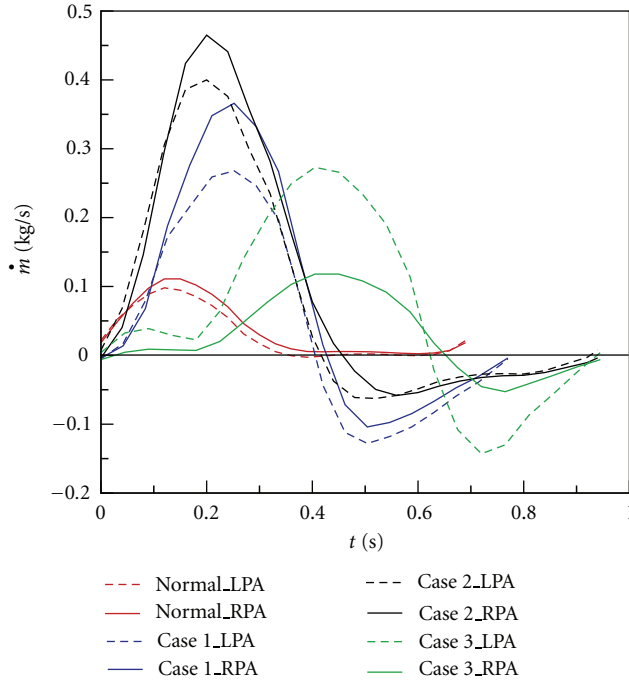
FIGURE 14: Time histories of mass flow rate in a cardiac cycle.

FIGURE 15: Comparing the L/R RF and MPA RF relationship between medical measurement and numerical computational values.
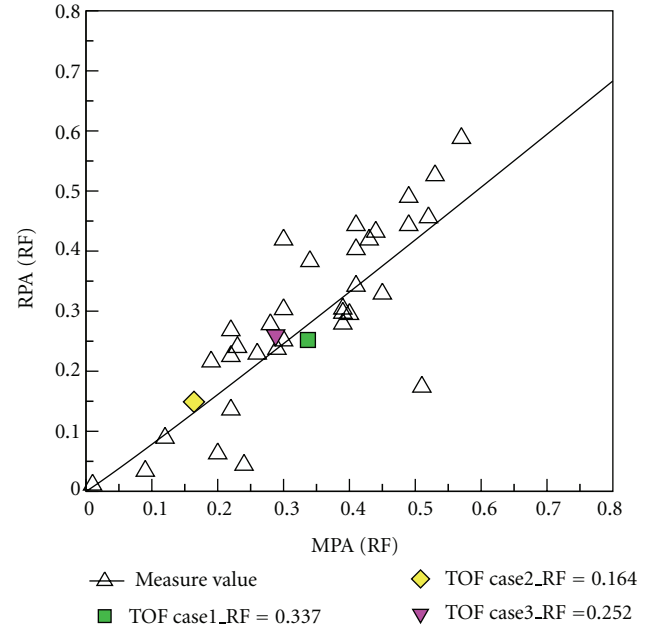
FIGURE 16: Comparing the RPA RF and MPA RF relationship between medical measurement and numerical computational values.

in the figure. Predicted results are within measurement data as shown in the figure. Consequently, the numerical computational values are acceptable based on RF analysis with clinical measurement values as shown in Figures 15, 16, and 17.

4. Conclusions

We utilize peripheric data of pulmonary arteries measured by MRI to reconstruct 3-D models of pulmonary arteries of healthy person and patients after repair of TOF. Those reconstructed models of three real pulmonary arteries are used to simulate blood flow motion. The degree of pulmonary regurgitation and geometry of pulmonary artery have significant effects on the blood flow variation.

Flow patterns are visualized by streamlines and velocity magnitude in numerical results. The recirculation region
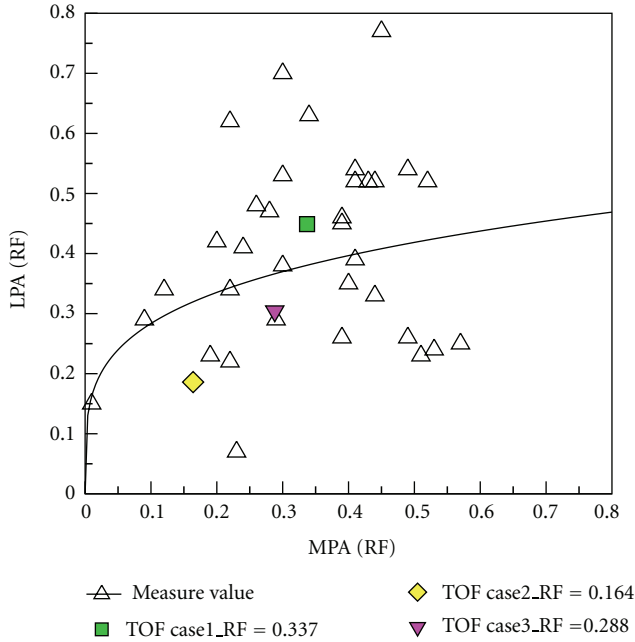
Figure 17: Comparing the LPA RF and MPA RF relationship between medical measurement and numerical computational values.

is found in LPA of all pulmonary arteries after repair of TOF in acceleration of systole. It obstructs the blood flow toward the left lung. The strong vortices are found in the dilation area of MPA in repaired pulmonary arteries during diastole of a cardiac cycle. Nevertheless, there is no such vortex in healthy MPA in diastole. In the ends of systole and diastole, 3-D complex vortices motion appears in repaired TOF pulmonary arteries. The blood flow of the health one is smoother than those. According to numerical results, it is found that the dilation causes abnormal flow distribution in MPA during regurgitation period in a cardiac cycle.

The comparison of the flow patterns and regurgitation proves that regurgitation happens first in LPA. The amount of regurgitation in LPA is larger than RPA in numerical results. The effect of predicted pressure distribution is also discussed. The dilation has extreme pressure change in a cardiac cycle. Nevertheless, the health MPA does not have such extreme pressure variation. The high pressure is one of the important factors to generate pulmonary arteries dilation.

We could understand the influence of pulmonary regurgitation and the blood flow patterns through this study. The results will be a useful reference for medical doctors before they perform operations for TOF patients.

Nomenclatures

Latin Symbols

$\dot{m}$: Mass flow rate, kg·s^{-1}
P: Pressure, Pa
Q_b: Backward blood flow volume in a cardiac cycle, m^3
Q_f: Forward blood flow volume in a cardiac cycle, m^3
RF: Regurgitation fraction
t: Time, s
$\mathbf{u}$: Velocity vector, m·s^{-1}.

Greek Symbols

θ: Bifurcation angle, degree
μ: Viscosity, kg·m^{-1} · s^{-1}
ν: Kinematic viscosity, m^2 · s^{-1}
ρ: Density, kg·m^{-3}.

Acknowledgments

References

[1] I. S. Kang, A. N. Redington, L. N. Benson et al., "Differential regurgitation in branch pulmonary arteries after repair of tetralogy of fallot: a phase-contrast cine magnetic resonance study," *Circulation*, vol. 107, no. 23, pp. 2938–2943, 2003.

[2] K. W. Ho, R. S. Tan, K. Y. Wong, T. H. Tan, S. Shankar, and J. L. Tan, "Late complications following tetralogy of fallot repair: the need for long-term follow-up," *Annals of the Academy of Medicine Singapore*, vol. 36, no. 11, pp. 947–953, 2007.

[3] M. T. Wu, Y. L. Huang, K. S. Hsieh et al., "Influence of pulmonary regurgitation inequality on differential perfusion of the lungs in tetralogy of Fallot after repair. A phase-contrast magnetic resonance imaging and perfusion scintigraphy study," *Journal of the American College of Cardiology*, vol. 49, no. 18, pp. 1880–1886, 2007.

[4] W. A. Helbing, R. A. Niezen, S. L. E. Cessie, R. J. van der Geest, J. Ottenkamp, and A. de Roos, "Right ventricular diastolic function in children with pulmonary regurgitation after repair of tetralogy of Fallot: volumetric evaluation by magnetic resonance velocity mapping," *Journal of the American College of Cardiology*, vol. 28, no. 7, pp. 1827–1835, 1997.

[5] A. Frigiola, A. N. Redington, S. Cullen, and M. Vogel, "Pulmonary regurgitation is an important determinant of right ventricular contractile dysfunction in patients with surgically repaired tetralogy of Fallot," *Circulation*, vol. 110, no. 11, supplement 1, pp. II153–II157, 2004.

[6] J. van den Berg, W. C. Hop, J. L. M. Strengers et al., "Clinical condition at mid-to-late follow-up after transatrial-transpulmonary repair of tetralogy of Fallot," *Journal of Thoracic and Cardiovascular Surgery*, vol. 133, no. 2, pp. 470–477, 2007.

[7] M. Grothoff, B. Spors, H. Abdul-Khaliq et al., "Pulmonary regurgitation is a powerful factor influencing QRS duration in patients after surgical repair of tetralogy of Fallot: a magnetic resonance imaging (MRI) study," *Clinical Research in Cardiology*, vol. 95, no. 12, pp. 643–649, 2006.

[8] H. W. Vliegen, A. van Straten, A. de Roos et al., "Magnetic resonance imaging to assess the hemodynamic effects of pulmonary valve replacement in adults late after repair of tetralogy of Fallot," *Circulation*, vol. 106, no. 13, pp. 1703–1707, 2002.

[9] B. H. van Huysduynen, A. van Straten, C. A. Swenne et al., "Reduction of QRS duration after pulmonary valve

replacement in adult Fallot patients is related to reduction of right ventricular volume," *European Heart Journal*, vol. 26, no. 9, pp. 928–932, 2005.

[10] H. W. Sung, T. L. Hsu, C. H. Hsu, and J. C. Hsu, "Pulmonary artery hemodynamics with varying degrees of valvular stenosis: an in vitro study," *Journal of Biomechanics*, vol. 31, no. 12, pp. 1153–1161, 1998.

[11] T. Tang, I. S. Chiu, H. C. Chen, K. Y. Cheng, and S. J. Chen, "Comparison of pulmonary arterial flow phenomena in spiral and Lecompte models by computational fluid dynamics," *Journal of Thoracic and Cardiovascular Surgery*, vol. 122, no. 3, pp. 529–534, 2001.

[12] A. F. Corno and E. S. Mickaily-Huber, "Comparative computational fluid dynamic study of two distal Contegra conduit anastomoses," *Interactive cardiovascular and thoracic surgery*, vol. 7, no. 1, pp. 1–5, 2008.

[13] M. J. Chern, M. T. Wu, and H. L. Wang, "Numerical investigation of regurgitation phenomena in pulmonary arteries of tetralogy of Fallot patients after repair," *Journal of Biomechanics*, vol. 41, no. 14, pp. 3002–3009, 2008.

[14] A. Redaelli, G. Rizzo, S. Arrigoni et al., "An assisted automated procedure for vessel geometry reconstruction and hemodynamic simulations from clinical imaging," *Computerized Medical Imaging and Graphics*, vol. 26, no. 3, pp. 143–152, 2002.

[15] B. T. Tang, T. A. Fonte, F. P. Chan, P. S. Tsao, J. A. Feinstein, and C. A. Taylor, "Three-dimensional hemodynamics in the human pulmonary arteries under resting and exercise conditions," *Annals of Biomedical Engineering*, vol. 39, no. 1, pp. 347–358, 2011.

[16] A. Das, R. K. Banerjee, and W. M. Gottliebson, "Right ventricular inefficiency in repaired tetralogy of Fallot: proof of concept for energy calculations from cardiac MRI data," *Annals of Biomedical Engineering*, vol. 38, no. 12, pp. 3674–3687, 2010.

[17] K. K. Whitehead, K. Pekkan, H. D. Kitajima, S. M. Paridon, A. P. Yoganathan, and M. A. Fogel, "Nonlinear power loss during exercise in single-ventricle patients after the Fontan: insights from computational fluid dynamics," *Circulation*, vol. 116, no. 11, supplement, pp. I165–I171, 2007.

[18] K. Pekkan, H. D. Kitajima, D. de Zelicourt et al., "Total cavopulmonary connection flow with functional left pulmonary artery stenosis: angioplasty and fenestration in vitro," *Circulation*, vol. 112, no. 21, pp. 3264–3271, 2005.

[19] T. Y. Hsia, F. Migliavacca, S. Pittaccio et al., "Computational fluid dynamic study of flow optimization in realistic models of the total cavopulmonary connections," *Journal of Surgical Research*, vol. 116, no. 2, pp. 305–313, 2004.

[20] Q. Sun, D. Wan, J. Liu et al., "Influence of antegrade pulmonary blood flow on the hemodynamic performance of bidirectional cavopulmonary anastomosis: a numerical study," *Medical Engineering and Physics*, vol. 31, no. 2, pp. 227–233, 2009.

[21] K. Pekkan, L. P. Dasi, P. Nourparvar et al., "In vitro hemodynamic investigation of the embryonic aortic arch at late gestation," *Journal of Biomechanics*, vol. 41, no. 8, pp. 1697–1706, 2008.

[22] C. Wang, K. Pekkan, D. de Zélicourt et al., "Progress in the CFD modeling of flow instabilities in anatomical total cavopulmonary connections," *Annals of Biomedical Engineering*, vol. 35, no. 11, pp. 1840–1856, 2007.

[23] T. J. Pedley, *The Fluid Mechanics of Large Blood Vessels*, Cambridge University Press, Cambridge, UK, 1980.

[24] S. A. Berger and L. D. Jou, "Flows in stenotic vessels," *Annual Review of Fluid Mechanics*, vol. 32, pp. 347–382, 2000.

[25] M. P. Singh, P. C. Sinha, and M. Aggarwal, "Flow in the entrance of the aorta," *Journal of Fluid Mechanics*, vol. 87, no. 1, pp. 97–120, 1978.

[26] N. L. Georgios and K. N. Ioannis, "Using the finite volume method and hybrid unstructured meshes to compute radiative heat transfer in 3-D geometries," *Numerical Heat Transfer, Part B*, vol. 62, no. 5, pp. 289–314, 2012.

[27] F. Kabinejadian and D. N. Ghista, "Compliant model of a coupled sequential coronary arterial bypass graft: effects of vessel wall elasticity and non-Newtonian rheology on blood flow regime and hemodynamic parameters distribution," *Medical Engineering & Physics*, vol. 34, no. 7, pp. 860–872, 2012.

[28] H. K. Versteeg and W. Malalasekera, *An Introduction to Computational Fluid Dynamics: The Finite Volume Method*, Addison Wesley Longman, Essex, UK, 1995.

[29] J. P. van Doormaal and G. D. Raithby, "Enhancements of the SIMPLE method for predicting incompressible fluid flow," *Numerical heat transfer*, vol. 7, no. 2, pp. 147–163, 1984.

[30] A. E. Perry and T. R. Steiner, "Large-scal vortex structures in turbulent wakes behind bluff bodies. Part 1. Vortex formation process," *Journal of Fluid Mechanics*, vol. 174, pp. 233–270, 1987.

Research Article

A Generalized Gamma Mixture Model for Ultrasonic Tissue Characterization

Gonzalo Vegas-Sanchez-Ferrero, Santiago Aja-Fernandez, Cesar Palencia, and Marcos Martin-Fernandez

Laboratorio de Procesado de Imagen, ETSI Telecomunicación Edificio de las Nuevas Tecnologías,
Campus Miguel Delibes s/n, Universidad de Valladolid, 47011 Valladolid, Spain

Correspondence should be addressed to Gonzalo Vegas-Sanchez-Ferrero, gvegsan@lpi.tel.uva.es

Several statistical models have been proposed in the literature to describe the behavior of speckles. Among them, the Nakagami distribution has proven to very accurately characterize the speckle behavior in tissues. However, it fails when describing the heavier tails caused by the impulsive response of a speckle. The Generalized Gamma (GG) distribution (which also generalizes the Nakagami distribution) was proposed to overcome these limitations. Despite the advantages of the distribution in terms of goodness of fitting, its main drawback is the lack of a closed-form maximum likelihood (ML) estimates. Thus, the calculation of its parameters becomes difficult and not attractive. In this work, we propose (1) a simple but robust methodology to estimate the ML parameters of GG distributions and (2) a Generalized Gama Mixture Model (GGMM). These mixture models are of great value in ultrasound imaging when the received signal is characterized by a different nature of tissues. We show that a better speckle characterization is achieved when using GG and GGMM rather than other state-of-the-art distributions and mixture models. Results showed the better performance of the GG distribution in characterizing the speckle of blood and myocardial tissue in ultrasonic images.

1. Introduction

Among the noninvasive imaging modalities, probably, the most widespread are the ultrasound imaging. The main reason of its success is that it provides a low-cost way to help diagnosing and can be used for many medical applications. However, ultrasonic (US) images are characterized by the presence of a peculiar granular pattern: the so-called *speckle*.

This term was adopted from the field of laser optics [1] in the early 1960s due to the similarity of the patterns between laser optics and ultrasonics. Although the nature of the speckle in US images stems from a different phenomenon, there still share some similarities. Both patterns come from the random interference of many coherent wave components reflected from different microscopic elements. In the case of US, the volume, the number of effective scatterers, and the acquisition process contribute to the formation of a speckle [2].

The analysis of backscattered echo from tissues needs a proper description of the ultrasonic signals. For this purpose, and due to the random nature of the speckle, several statistical models have been proposed in the literature. This characterization can be used either for segmentation [3], classification [4] purposes or for filtering the speckle itself [5–8]. The latter usually considers the speckle as an undesired consequence, since it degrades resolution and adds spatial noise to the image. Thus, filtering is commonly applied as a preprocessing step for further segmentation of regions of interest or to extract relevant measures for physiological analysis.

The statistical description of US signals provide an important information of the backscattered echo from tissues. The parameters of the statistical models allow identifying the features of tissues and provides important descriptors for classification. Some of the filtering algorithms relay on a Bayesian approach where an accurate statistical model

becomes necessary. As a consequence, modeling the amplitude statistics of US signals has been a very active area.

Several statistical models have been proposed in the last decades. Probably the most wellknown is the Rayleigh model, which is a one-parameter distribution which describes the so-called fully formed (or developed) speckle. This probabilistic distribution describes the behavior of a speckle when a high number of effective scatterers are present in the resolution cell. However, real images show a deviation from this model, this non-Rayleigh behavior can be due to a small number of scatterers in the resolution cell or when there are some dominant components in the cell. The most commonly accepted distributions that try to model non-Rayleigh distributions are the Rice (fully resolved speckle), K (partially formed speckle), and Homodyned K (partially resolved speckle).

Although, those models are based on physical assumptions of the backscattering process, some other distributions have proven to provide a good performance on real images. This is the case of Gamma [9–11] and Nakagami [12] distributions. The first is proposed as a two-parameter distribution that describes the result of interpolated/filtered fully formed speckle [9] and also has shown good results in empirical tests among other distributions [10, 11]. The Nakagami distribution proposed by Shankar for the case US characterization [12] is also a two-parameter distribution which generalizes the Rayleigh distribution. This distribution was adopted from the models proposed to describe the statistics of the returned echo radar.

The capability of the Nakagami distribution to model the backscattering from tissues for fully resolved and fully formed speckle made it become the most commonly accepted model for tissue characterization. However, the tails of the probabilistic density functions of Nakagami, K, Rayleigh, or Gamma do not show the impulsive response of speckle which originate heavier tails. In order to describe this impulsive response, a generalized Nakagami distribution was proposed by Shankar in [13]. This is a three-parameter model which has shown a better behavior than the Nakagami or Rayleigh, an expected result since it is a generalization of the other models. However, the generalized Nakagami distribution does not have closed-form maximum likelihood estimates (MLE) and, thus, it makes their use difficult. Note that, though Shankar in [13] said that the MLE can be obtained, the equations used were based on the results from Stacy and Mihram [14], which were calculated by the methods of moments and they also expressed the difficulties of obtaining an MLE: "Closed expressions for solutions to the maximum likelihood equations are highly unlikely." It is important to note that the results of [14] were obtained for the estimation of the Generalized Gamma (GG) distribution which is essentially the same as the generalized Nakagami distribution but with another parametrization.

The different nature of tissues is reflected in a different response of the speckle. Hence, a mixture model has shown to be a natural strategy for statistically describing the features of tissues. This approach has been previously used for segmentation purposes in the case of Nakagami mixture models (NMMs) by Destrempes et al. in [3], for classification with Rayleigh Mixture Models by Seabra et al. in [4], and for filtering with a mixture of Gamma and Gaussian Mixture Models in [8, 9]. All these approaches make use of the Expectation-Maximization (EM) [15] algorithm to calculate the parameters that better fit the empirical probability distribution function (PDF). This method is particularly useful when the MLE exists since it maximizes the expected value of the log-likelihood function with respect to the condition of the belonging to each tissue class for a given data.

The EM algorithm cannot be easily applied for the calculation of a Generalized Gamma Mixture Model (GGMM) without an MLE. However, some interesting results have been recently published on the calculation of the MLE of the Generalized Gamma which permit efficient computation of the GGMM.

The aim of this work is to revitalize the use of the Generalized Gamma distribution (also called, Generalized Nakagami Distribution) for tissue characterization. For this purpose, we present two main contributions: first, we propose a simple methodology to calculate the ML estimate which offers robust results comparing to the methods in the literature [14, 16, 17]. Second, two different methods were proposed for the calculation of the GGMM parameters. Both were developed by applying the EM method in the derivation of the proposed ML method. Results when comparing both methods to the GMM and NMM in real images showed the better fitting of the GGMM. The GGMM provides a posterior probability of belonging to each tissue class which can be of help for further filtering, segmenting, or classifying methods.

The rest of the paper is structured as follows. In Section 2.1, we introduce the distributions most commonly used for characterizing speckle of ultrasonic images. There, the GG distribution is motivated as a suitable generalization of the Gamma and Nakagami distributions which fail in describing the impulsive response of speckle. Then, in Section 2.2, we analyze the methods proposed in the literature for estimating the parameters of the GG distribution and a simple but robust method is proposed (Section 2.2.4). One of the advantages of this method is that it can be easily extended to estimate the parameters of a GGMM by means of the EM algorithm. Section 2.3 is devoted to the extension of the ML method to obtain the parameters of the GGMM where two algorithms are proposed. The performance of the ML estimate derived in Section 2.2.4 is compared to other state-of-the-art methods in Section 3.1 for synthetic data and for real cases in Section 3.2. The performance of the GGMM is analyzed in Section 3.3, where the GGMM is compared to NMM and GMM. Finally, we propose some applications for the GGMM in Section 3.4. In Section 4, we conclude.

2. Materials and Methods

2.1. Statistical Models for Describing the Nature of Speckle. The formation of US images begins with the emission of a pulse packet which travels through the tissue. The backscattering produced by the scatterers in the resolution cell contribute to the change of the pulse shape according to the characteristics of the media, that is, the number of scatterers as well as their size [4, 9, 12].

The contribution of the backscattered echo, $s(t)$, can be treated as a random walk due to the random location of the scatterers in the resolution cell [12]:

$$s(t) = \sum_{n=1}^{N} \alpha_n \cos(\omega_0 t + \phi_n), \qquad (1)$$

where ω_0 is the mean frequency of excitation and N is the number of effective scatterers in the resolution cell. The phases, ϕ_n, are usually modeled as uniformly distributed in $[0, 2\pi]$ and the amplitude is usually considered to be Normal distributed.

The fully formed speckle model assumes a high number of scatterers, so the *Central Limit Theorem* applies and the backscattered echo can be expressed as

$$s(t) = X \cos(\omega_0 t) + Y \sin(\omega_0 t), \qquad (2)$$

where X and Y are zero mean identically distributed Gaussian distributions.

Then, the envelop of the backscattered signal echo, $R = \sqrt{X^2 + Y^2}$ is Rayleigh distributed [1, 18]:

$$f_R(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} u(r), \qquad (3)$$

where $u(\cdot)$ is the Heaviside step function defined as

$$u(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0. \end{cases} \qquad (4)$$

Under the assumption of a high number of effective scatterers but with the presence of resolvable structures in the resolution cell (specular component, C), X and Y become nonzero Gaussian distributions. The envelop does no longer follow a Rayleigh distribution but a Rician one [18]:

$$f_R(r) = \frac{r}{\sigma^2} e^{-(r^2+C^2)/2\sigma^2} I_0\left(\frac{rC}{\sigma^2}\right) u(r), \qquad (5)$$

where $I_0(\cdot)$ is the modified Bessel function of first kind.

When the number of scatterers decreases and the Central Limit Theorem cannot be applied, more complicated distributions are proposed to model the distribution of the envelope. Concretely, the K distribution models the case when the number of scatterers is a random variable itself, which is modeled as a Poisson whose local mean is Gamma distributed, this is equivalent to consider σ as gamma distributed [2]:

$$f_R(r \mid \sigma) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2} u(r),$$

$$f_\sigma(\sigma) = \frac{1}{2b^2} \frac{1}{\Gamma(\nu+1)} \left(\frac{\sigma}{2b^2}\right) e^{-\sigma/2b^2} u(\sigma), \qquad (6)$$

so, the PDF of R is

$$f_R(r) = \int f_R(r \mid \sigma) f_\sigma(\sigma) d\sigma$$

$$= \frac{2}{b\Gamma(\nu+1)} \left(\frac{\sigma}{2b^2}\right)^{\nu+1} K_\nu\left(\frac{r}{b}\right) u(r), \qquad (7)$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind.

A generalization of the previous models appears when a specular component is considered and the number of scatterers, N, follows a negative binomial distribution. This is the case of the homodyned-K distribution [19]:

$$f_R(r) = r\left(\int \frac{x}{1 + x^2\sigma^2/2\nu} J_0(xC) J_0(xr) dx\right) u(r). \qquad (8)$$

This PDF has no closed expression and this limits its use.

On a completely different approach, Shankar in [12] proposed a Nakagami distribution as a "simpler universal model for tissue characterization." Unlike the previously reviewed models, the Nakagami is not based on physical arguments or on a bottom-up modeling of the scattering process. However, it has empirically shown a better performance than the Rayleigh and Rice distributions.

The Nakagami PDF is as follows:

$$f_R(r) = \frac{2m^m r^{2m-1}}{\Gamma(m)(2\Omega)^m} e^{-(m/2\Omega)r^2} u(r). \qquad (9)$$

This distribution offers good properties to describe the backscattered echo: the Rayleigh distribution is a particular case of the Nakagami ($m = 1$) and, additionally, when $m > 1$ is similar to the Rice distribution. However, this distribution has some limitations. The Nakagami model cannot fit the heavier tails of the empirical PDFs due to the impulsive nature of scatterers [13].

In order to describe the impulsive response of scatterers, Shankar proposed in [13] a generalized Nakagami distribution which is essentially the same as a Generalized Gamma distribution [14]. However, this distribution presents some difficulties in the estimation of its parameters, since there are no closed equations for the ML estimates.

In the next section, we describe some methods that have been used in the literature with special attention to methods that provide an ML estimate of the GG parameters. Additionally, we propose a simple method to calculate the ML estimates of the parameters. The results obtained in the derivation of this ML method provide the foundations for the development of the Generalized Gamma Mixture Model, which is the main contribution of this work.

2.2. Estimation of Parameters of the Generalized Gamma

2.2.1. Moments Method. This method was proposed by Stacy in [14]. For the derivation of the method, the following parametrization was adopted:

$$f(x \mid a, \nu, p) = p \frac{x^{p\nu-1}}{a^{p\nu}\Gamma(\nu)} e^{-(x/a)^p} u(x), \qquad (10)$$

where the parameters (a, ν, p) are all positive.

This is the definition of the GG distribution hereafter. For a given $p > 0$, all moments $E\{X^r\}$ exist.

Now, let Z be the random variable (RV) defined as

$$Z = \log\left(\frac{X}{a}\right)^p = p(\log(X) - \log(a)). \qquad (11)$$

For this RV, the central moments, $\mu_r(\cdot)$, of rth order are

$$\mu_r(Z) = p^r \mu_r(\log X). \tag{12}$$

Additionally, it is easy to show that, given a RV, X, which follows a GG distribution ($X \sim GG(a, \nu, p)$), the following properties hold:

$$kX \sim GG(ka, \nu, p), \quad k > 0$$

$$X^m \sim GG\left(a^m, \nu, \frac{p}{m}\right), \quad m \neq 0. \tag{13}$$

So, a new RV Z can be defined as $Z = \log(X/a)^p$ where $(X/a)^p$ follows a Gamma distribution of parameter ν. Hence, the log-transformed distribution of the Gamma RV is the following:

$$f_Z(z) = f_{GG}(e^z \mid 1, \nu, 1)e^z = \frac{1}{\Gamma(\nu)} \exp(\nu z - \exp(z)), \tag{14}$$

where $z \in \mathbb{R}$.

The moment generating function of Z is easily calculated as $E\{e^{tZ}\} = (t+\nu)/\nu$. Where $E\{\cdot\}$ is the expectation operator. So, the rth moment of Z is the following:

$$E\{Z^r\} = \frac{\Gamma^{(r)}(\nu)}{\Gamma(\nu)} = \Psi^{(r)}(\nu), \tag{15}$$

where $\Psi^{(r)}$ is the *polygamma* function defined as

$$\Psi^{(r)}(x) = \frac{d^{m+1}}{dx} \log \Gamma(x). \tag{16}$$

Finally, the three first central moments are defined as:

$$pE\{\log X - \log a\} = \Psi^{(0)},$$

$$p^2 \mu_2(\log X) = \Psi^{(1)}, \tag{17}$$

$$p^3 \mu_3(\log X) = \Psi^{(2)}.$$

These equations can be used to estimate the parameters of the $GG(a, \nu, p)$; $\hat{a}, \hat{\nu}, \hat{p}$, by approximating the moments by means of the sample moments:

$$\hat{a} = \exp\left(\overline{y} - \Psi^{(0)}(\hat{\nu})\right),$$

$$\hat{p} = -\text{sign}\left(g_y\right)\frac{\sqrt{\Psi^{(1)}(\hat{\nu})}}{S_y}, \tag{18}$$

$$-\left|g_y\right| = \frac{\Psi^{(2)}(\hat{\nu})}{(\Psi^{(1)}(\hat{\nu}))^{3/2}},$$

where $\overline{y} = (1/N) \sum_{i=1}^N \log x_i$, with $\{x_i\}_{i=1}^N$ the set of samples of X; S_y^2 is the sample variance of $\{y_i\}_{i=1}^N = \{\log x_i\}_{i=1}^N$, and g_y its sample skewness.

The estimates are derived by means of calculating the value $\hat{\nu}$ from the last equation of (18). So, a numerical calculation needs to be performed. In the original article [14], Stacy and Mihran provided a graph representing $\Psi^{(2)}(\nu)/(\Psi^{(1)}(\nu))^{3/2}$ for a range $\nu \in [0.1, 5]$.

This method, though provides a quite straight-forward calculation of the parameters, can provide estimates which are outside the parameter space. Yet, it is highly sensitive to the number of samples.

2.2.2. Heuristic Approaches.

In order to avoid the problems associated to the moments method, some heuristic methods have been proposed in the literature. As examples, Gomes et al. [16] proposed an iterative method which evaluates the best performance of the χ^2 goodness-of-fit test for a fixed p (see the parametrization of (10)). The parameters of the transformed samples $Y = X^p$, which are Gamma distributed, were calculated by the moments method. At the end of the loop, the set of parameters with least P value is chosen.

This method presents some shortcomings. First, the parameters of the Gamma distributed data were calculated by the moments method, so the problems associated to the moments method are not circumvent. However, even if a good estimate is calculated, the χ^2 goodness-of-fit test depends on the calculation of the estimated PDF which strongly depends on the number of bins considered and the assumption of a sample with sufficient large size.

Other heuristic method is the one presented by Wingo in [20]. This method, based on the one proposed by Hager and Bain [21], tries to solve the maximum likelihood equations for the GG distribution. The log-likelihood, $\mathcal{LL}$, of a RV $X \sim GG(a, \nu, p)$ for the parametrization presented in (10) is

$$\mathcal{LL}(a, \nu, p \mid \mathbf{x}) = \log\left(\left(\frac{p}{a^{p\nu}\Gamma(\nu)}\right)^n \prod_{i=1}^n x_i^{p\nu-1} e^{-\sum_{i=1}^n (x_i/a)^p}\right)$$

$$= n \log(p) - np\nu \log(a) - n \log(\Gamma(\nu))$$

$$+ (p\nu - 1) \sum_{i=1}^n \log x_i - \sum_{i=1}^n \left(\frac{x_i}{a}\right)^p, \tag{19}$$

where $\mathbf{x} = \{x_i\}_{i=1}^n$ is the set of samples.

Now, calculating the derivatives with respect to the parameters and setting it equal to zero, one can obtain the ML equations:

$$a^p = \frac{1}{n\nu} \sum_{i=1}^n x_i^p,$$

$$p \sum_{i=1}^n \log\left(\frac{x_i}{a}\right) - n\Psi(\nu) = 0, \tag{20}$$

$$\frac{n}{p} + \nu \sum_{i=1}^n \log\left(\frac{x_i}{a}\right) - \sum_{i=1}^n \left(\frac{x_i}{a}\right)^p \log\left(\frac{x_i}{a}\right) = 0,$$

where $\Psi(x) \equiv \Psi^{(0)}(x) = \Gamma'(x)/\Gamma(x)$.

This system of equations can be reduced to a single non-linear equation with p as the single unknown:

$$-\Psi(\nu) + \frac{p}{n} \sum_{i=1}^n \log(x_i) - \log\left(\sum_{i=1}^n x_i^p\right) + \log(n\nu) = 0, \tag{21}$$

where

$$\nu = -\left(\frac{p}{n}\sum_{i=1}^{n}\log(x_i) - p\frac{\sum_{i=1}^{n}x_i^p\log(x_i)}{\sum_{i=1}^{n}x_i^p} \right)^{-1},$$

$$a = \left(\frac{1}{n\nu}\sum_{i=1}^{n}x_i^p \right)^{1/p}. \tag{22}$$

So, the problem is reduced to calculate p from (21). Some authors reported the difficulty of solving this equation with the conventional numerical methods such as Newton-Raphson [21] and conclude that the MLE may not exist.

In [20], the author faced the problem by analyzing the effect of inappropriate zero finding algorithms. So, an heuristic method for isolating roots of a general scalar nonlinear equation was proposed. This method makes use of the root-isolation technique proposed in [22], which uses only function values to isolate the roots in a compact interval of the real line.

Though this method can provide an ML estimate of the parameter by solving (21), it has to heuristically divide the intervals where p is searched and calculate whether a root is in it or not by means of the mean value and variance of the function in each of the intervals, so many evaluations of the function are required.

2.2.3. ML Approach.
A very interesting analysis was recently published by Noufaily and Jones in [17], where an iterative approach is proposed to solve the likelihood equations, (20), in a way that the individual equations are uniquely solvable. This result provides a very promising technique for calculating the MLE parameters of the GG.

In that work, the log-likelihood equations were calculated following the re-parametrization proposed in [23]. Concretely, for a RV X which is distributed, $X \sim GG(a, \nu, p)$, the new RV $Y = \log X$ is calculated, whose PDF is the following:

$$f_Y(y) = \frac{p}{\Gamma(\nu)}\frac{e^{y p \nu}}{a^{p\nu}}\exp\left(-\frac{e^{yp}}{a^p}\right)$$

$$= \frac{\nu^{\nu-1/2}}{\sigma\Gamma(\nu)}\exp\left(\sqrt{\nu}w - \nu e^{w/\sqrt{\nu}}\right), \tag{23}$$

where $y \in \mathbb{R}$, $w = (y - \mu)/\sigma$, $\sigma = 1/p\sqrt{\nu}$ and $\mu = \log(a) + (1/p)\log(\nu)$.

So, in the end, the following equations have to be solved:

$$\mu = \sigma\sqrt{\nu}\log S_0, \tag{24}$$

$$R(\sigma) \equiv \frac{S_0}{S_1} - \overline{Y} - \frac{\sigma}{\sqrt{\nu}} = 0, \tag{25}$$

$$T(\nu) \equiv \log(\nu) - \Psi(\nu) - \frac{L}{\sqrt{\nu}} = 0, \tag{26}$$

where $L = (\mu - \overline{Y})/\sigma$ and $S_j = (1/n)\sum_{i=1}^{n}y_i^j\exp(y_i/\sigma\sqrt{\nu})$.

The important result of [17] is the demonstration that both (25) and (26) are well behaved with unique solutions in σ and ν, respectively. So, an iterative method can be

developed to calculate $\hat{\nu}$ by (26) from an initial guest of the parameters and then $\hat{\sigma}$ by solving (25). Finally, $\hat{\mu}$ is calculated by replacing the previous estimates in (24). These estimates can be used to calculate a new L to compute the new log-likelihood function. By repeating these steps until a desired accuracy, the estimates are achieved [23].

This method provides a fundamental result about the behavior of the log-likelihood equations, and guarantees their solution. However, the method does not provide any proof concerning its convergence or the uniqueness of the ML. Yet, this method needs to solve two nonlinear equations by numerical techniques, whereas the method proposed by Wingo in [20], previously described, only needs to solve a linear equation.

2.2.4. The Proposed Approach.
We propose a simple but efficient method to calculate the ML estimates of the GG distribution. The main advantage of the method is that it can be easily implemented and has the same properties of the method of [17], that is, the equation to solve are well behaved with unique solution. Additionally, the method just needs the calculation of just one non-linear equation and, thus, the computing time is considerably reduced.

The method consists in transforming the RV, $X \sim GG(a, \nu, p)$ by the following transformation $Y = X^{p_0}$ where p_0 is a positive real number. So, the new PDF of Y is the following:

$$f_Y(y) = \frac{p}{p_0}\frac{y^{(p/p_0)\nu-1}}{a^{p\nu}\Gamma(\nu)}\exp\left(-\frac{y^{p/p_0}}{a^p}\right)u(y). \tag{27}$$

Note that this PDF follows a Gamma PDF when $p_0 = p$. Hence, a reasonable way to find the P value is to find the value of p_0 that maximizes the likelihood of the GG distribution and also maximizes the Gamma distributed RV $Y = X^{p_0}$.

In order to see if this method provides a proper solution, we first demonstrate that the ML estimate of the parameters of the new random variable Y also maximizes the likelihood of the GG distribution when $p_0 = p$.

First, we calculate the ML estimates of the parameters of (27) for $p_0 = p$, whose log-likelihood is the following:

$$\mathcal{L}\mathcal{L}_Y = -np\nu\log(a) - n\log(\Gamma(\nu))$$

$$+ (\nu - 1)\sum_{i=1}^{n}\log(y_i) - \sum_{i=1}^{n}\frac{y_i}{a^p}. \tag{28}$$

The maximum with respect to the parameter a is easily calculated by taking the derivative with respect to the a and setting it equal to zero:

$$a_0^p = \frac{1}{n\nu}\sum_{i=1}^{n}y_i. \tag{29}$$

Finally, (28) can be maximized with respect to ν by introducing the value of a_0:

$$\log(\nu) - \Psi(\nu) = \log\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right) - \frac{1}{n}\sum_{i=1}^{n}\log(y_i). \tag{30}$$

Now, by introducing a_0 in the log-likelihood function of the GG distribution, (19):

$$\mathcal{LL}_X = n \log(p) - n\nu \log\left(\frac{1}{n\nu} \sum_{i=1}^{n} y_i\right) - n \log(\Gamma(\nu)) \tag{31}$$

$$+ \left(\nu - \frac{1}{p}\right) \sum_{i=1}^{n} \log(y_i) - n\nu.$$

Now, by maximizing with respect to ν, we obtain the following equation:

$$\frac{\partial \mathcal{LL}_X}{\partial \nu} \equiv - n \log\left(\frac{1}{n} \sum_{i=1}^{n} y_i\right) + n \log(\nu) \tag{32}$$

$$+ \sum_{i=1}^{n} \log(y_i) - n\Psi(\nu) = 0,$$

and finally, reordering terms, we obtain the same equation for which ν_0 is also a solution.

This result guarantees that there exists always a solution for the ML estimate of the GG distribution $(\hat{a}, \hat{\nu}, \hat{p})$ and the parameters $\hat{a}$ and $\hat{\nu}$ are those obtained for the ML estimate for the transformed RV $Y = X^{\hat{p}}$. Hence, there is always a solution for the MLE for a GG.

Additionally, since the MLE of a Gamma distribution always exist for whatever positive y_i values ((30) is well behaved), the problem is reduced to finding the value p that maximizes $\mathcal{LL}_X$ among the ones that maximize $\mathcal{LL}_Y$.

The search method for p was implemented by the Nelder-Mead method [24] while the Brent's algorithm was applied for calculating ν [25].

This method does not demonstrate the uniqueness of p as did not any of the methods in the literature. However, in our experience, we agree with Noufaily and Jones [17] that the global maximum of the $\mathcal{LL}_X$ appears to be distant to any other local maximum.

The main advantage of the method here proposed is that it is easy to implement and only one non linear equation has to be solved, whereas the method of [17] needs to solve two non-linear equations in each iteration and [20] method needs several calculations of non-linear equations for each interval considered for the isolation root technique.

2.3. Generalized Gamma Mixture Model. An additional advantage of the proposed method for the calculation of the MLE parameters for the GG distribution is that it can be easily adapted for the calculation of the parameters of GG Mixture Models (GGMM).

There were some attempts in the literature to obtain the parameters of a GGMM. Concretely, in [26], they calculated the GGMM by means of the Nelder-Mead and Gradient descent methods for maximizing the log-likelihood. However, that method is strongly sensitive to the number of mixtures since it is just a direct optimization of the log-likelihood score equations of the mixture model.

In this section, we derive the GGMM by applying the Expectation-Maximization methodology [15] and combining them with the method used to calculate the MLE of the GG distribution.

Let $X = \{x_i\}$, $1 \geq i \geq N$ be a set of samples. These samples are considered to be independent and identically distributed (IID) RVs. Now, the GGMM considers that these samples result from the contributions of J distributions:

$$p(x_i \mid \Theta) = \sum_{j=1}^{J} \pi_j f_X\left(x_i \mid \Phi_j\right), \tag{33}$$

where Φ is a vector of the parameters of the GGMM $(\pi_1, \ldots, \pi_J, \Theta_1, \ldots, \Theta_J)$ and Θ_j are the parameters of the PDF (in our case the parameters of the GG, represented as a_i, ν_j, p_j).

The joint distribution of IID samples is given by

$$p(X \mid \Theta) = \prod_{i=1}^{N} p(x_i \mid \Theta). \tag{34}$$

The EM is applied here to maximize the log-likelihood function when some hidden discrete random variables $Z = \{Z_i\}$ are introduced into the model. These RVs take values in $\{1, \ldots, J\}$ and indicate the class for which each sample x_i belongs.

Now, defining $\Theta^{(n)}$ as an estimate of the parameters of the mixture in the nth iteration, the expectation step is performed by calculating the expected value of the log-likelihood $\mathcal{LL}(\Theta \mid X, Z)$:

$$\mathcal{Q}\left(\Theta \mid \Theta^{(n)}\right) = E_{Z \mid \Theta^{(n)}}\{\mathcal{LL}(\Theta \mid X, Z)\}. \tag{35}$$

In the maximization step, the new estimate $\Theta^{(n)}$ is obtained by maximizing the expectation of the log-likelihood function $\mathcal{Q}(\Theta \mid \Theta^{(n)})$. These steps are iterated until a stop criterion such as $\mathcal{Q}(\Theta \mid \Theta^{(n+1)}) - \mathcal{Q}(\Theta \mid \Theta^{(n)}) < \text{Tol}$ for some preestablished tolerance (Tol) is reached.

The application of the EM algorithm for estimating the parameters of mixture models has been applied for several distributions, see, for example, [15, 27]. However, to the best of our knowledge, this is the first time a mixture model is presented for GG distributions.

In order to derive the estimates of the parameters in each iteration, we first define the joint distribution of IID samples X and the hidden random variables, Z as

$$p(X, Z \mid \Theta) = \prod_{i=1}^{N} p(x_i, z_i \mid \Theta), \tag{36}$$

where $p(x_i, z_i \Theta) = p(x_i z_i, \Theta) p(z_i \Theta)$.

Now, the log-likelihood function can be defined in the following way:

$$\mathcal{LL}(\Theta \mid X, Z) = \log(p(X, Z \mid \Theta)) = \sum_{i=1}^{N} \log p(x_i, z_i \mid \Theta)$$

$$= \sum_{i=1}^{N} \log p(x_i \mid z_i, \Theta) + \sum_{i=1}^{N} \log p(z_i \mid \Theta)$$

$$= \sum_{i=1}^{N} \log f_X(x_i \mid z_i, \Theta) + \sum_{i=1}^{N} \log \pi_{z_i}. \tag{37}$$

The expectation of the log-likelihood function with respect to the hidden RVs when data $\{x_i\}$ and the previous estimate $\Theta^{(n)}$ are known as:

$$
\begin{aligned}
\mathcal{Q}\left(\Theta \mid \Theta^{(n)}\right) &= E_{Z\mid\Theta^{(n)}}\{\mathcal{LL}(\Theta \mid X, Z)\} \\
&= \sum_{i=1}^{N} E_{Z\mid\Theta^{(n)}, x_i}\{\log f_X(x_i \mid \Theta) + \log p(z_i \mid \Theta)\} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J} p\left(Z_i = j \mid x_i, \Theta^{(n)}\right) \\
&\quad \times \left(\log f_X(x_i \mid \Theta_{z_i}) + \log \pi_j\right),
\end{aligned}
$$
(38)

where $\pi_j = p(Z_i = j \mid \Theta)$ is the probability of x_i to belong to the class j.

The probability $p(Z_i = j \mid x_i, \Theta^{(n)})$ can be calculated by applying the Bayes theorem as

$$
p\left(Z_i = j \mid x_i, \Theta^{(n)}\right) = \frac{f_X\left(x_i \mid \Theta^{(n)}\right) p\left(Z_i = j \mid \Theta^{(n)}\right)}{p(x_i \mid \Theta^{(n)})}. \quad (39)
$$

Note that (37) is composed of two terms, so the maximization step can be done to each term independently. For the term depending on the π_j some constraints have to be considered since they must hold $\sum_{j=1}^{J} \pi_j = 1$. An optimization via Lagrange's multipliers can be done in a straightforward way and they guarantee a necessary condition for optimality in this problem. The new Lagrange function with λ as the Lagrange multiplier is the following:

$$
\Lambda(\pi, \lambda) = \sum_{i=1}^{N}\sum_{j=1}^{J} \gamma_{i,j} \log \pi_j + \lambda\left(\sum_{j=1}^{J} \pi_j - 1\right), \quad (40)
$$

where we introduced $\gamma_{i,j} = p(Z_i = j \mid x_i, \Theta^{(n)})$ to simplify notation.

Now, calculating the derivative with respect to each π_j and setting it equal to 0, the following expression is derived:

$$
\sum_{i=1}^{N} \gamma_{i,j} = -\lambda \hat{\pi}_j. \quad (41)
$$

By summing both terms of the equation over j, we obtain $\lambda = -\sum_{i=1}^{N}\sum_{j=1}^{J} \gamma_{i,j} = -N$ and the estimates for the parameters π_j that maximize the Lagrange function (and the likelihood function) are

$$
\hat{\pi}_j = \frac{1}{N}\sum_{i=1}^{N} \gamma_{i,j} = \frac{1}{N}\sum_{i=1}^{N} p\left(Z_i = j \mid x_i, \Theta^{(n)}\right). \quad (42)
$$

For the calculation of the maximum of (37) with respect to $\Theta_j = (a_j, \nu_j, p_j)$, we first calculate the derivative with respect to a_j:

$$
\frac{\partial}{\partial a_j}\left\{\sum_{i=1}^{N}\sum_{j=1}^{J} \gamma_{i,j} \log f_X\left(x_i \mid \Theta_j\right)\right\} = 0, \quad (43)
$$

where the log-likelihood of $p(x_i \mid \Theta_j)$ is the one described in (19) for one sample x_i:

$$
\begin{aligned}
\log f_X\left(x_i \mid \Theta_j\right) = {}& \log p - p\nu \log(a) - \log \Gamma(\nu) \\
& + (p\nu - 1)\log x_i - \left(\frac{x_i}{a}\right)^p.
\end{aligned}
$$
(44)

The result is

$$
\hat{a}^{p_j} = \frac{\sum_{i=1}^{N} \gamma_{i,j} x_i^{p_j}}{\nu_j \sum_{i=1}^{N} \gamma_{i,j}}. \quad (45)
$$

Now, plugging (45) into (37) and deriving with respect to ν_j and setting it equal to 0

$$
\frac{\partial}{\partial \nu_j}\left\{\sum_{i=1}^{N}\sum_{j=1}^{J} \gamma_{i,j} \log f_X\left(x_i \mid \frac{\sum_{i=1}^{N} \gamma_{i,j} x_i^{p_j}}{\nu_j \sum_{i=1}^{N} \gamma_{i,j}}, \nu_j, p_j\right)\right\} = 0. \quad (46)
$$

It results in the following equality:

$$
\log\left(\nu_j\right) - \Psi\left(\nu_j\right) = \log\left(\frac{\sum_{i=1}^{N} \gamma_{i,j} x_i^p}{\sum_{i=1}^{N} \gamma_{i,j}}\right) + \frac{\sum_{i=1}^{N} \gamma_{i,j} \log\left(x_i^{p_j}\right)}{\sum_{i=1}^{N} \gamma_{i,j}}. \quad (47)
$$

Note that (47) is essentially the same as (30), which is well behaved and always has a unique solution. Thus, this nonlinear equation can be solved by numerical methods in the same way as was performed in the MLE of the GG parameters. In our case, we also used the Brent's algorithm [25].

The interval where the Brent's algorithm is performed can be derived by means of the following property:

$$
\frac{1}{2\nu_j} < \log\left(\nu_j\right) - \Psi\left(\nu_j\right) < \frac{1}{\nu_j}. \quad (48)
$$

So, the desired value of $\hat{\nu}_j$ in the interval

$$
\frac{1}{2A} < \hat{\nu}_j < \frac{1}{A}, \quad (49)
$$

where

$$
A = \log\left(\frac{\sum_{i=1}^{N} \gamma_{i,j} x_i^{p_j}}{\sum_{i=1}^{N} \gamma_{i,j}}\right) + \frac{\sum_{i=1}^{N} \gamma_{i,j} \log\left(x_i^{p_j}\right)}{\sum_{i=1}^{N} \gamma_{i,j}}. \quad (50)
$$

This property can be found in [28] and was also used in [17] for the calculation of the ML estimates of the GG.

Now, the problem can be stated in the same way as was done for the ML estimate proposed in Section 2.2.4. We are interested in the parameter p_j which maximizes the likelihood for the component $j \in [1, J]$. So, for each p_j, (45) and (47) provide the estimate of a_j and ν_j, respectively. By applying the Nelder-Mead algorithm to maximize the log-likelihood function for each component j, as was done for the ML estimates in Section 2.2.4, one can obtain the desired ML estimates. We will refer to this method as the $GGMM_1$ method.

It is important to note that the parameter estimates can be also solved by extending the ML method of [17]. For

this purpose, the parametrization proposed by Lawless [23] can be applied to the mixture model as was explained in Section 2.2.3.

The log-likelihood equations to be solved are completely equivalent to (45) and (47) due to the invariance of the ML estimates to the transformation $Y = \log(X)$. However, Lawless' parametrization allows us to extend the results of [17] to the case of GGMM. For the sake of clarity, we rewrite the parametrization:

$$\sigma_j = \frac{1}{p_j\sqrt{\nu_j}},$$

$$\mu_j = \log(a_j) + \frac{1}{p_j}\log(\nu_j), \tag{51}$$

$$k_j = \nu_j.$$

With this parametrization, (45) becomes

$$\hat{\mu}_j = \sqrt{k_j}\sigma_j \log(\widetilde{S}_0), \tag{52}$$

where

$$\widetilde{S}_r = \frac{\sum_{i=1}^{N} \gamma_{i,j} y_i^r e^{(y_i/\sigma_j\sqrt{k_j})}}{\sum_{i=1}^{N} \gamma_{i,j}}. \tag{53}$$

So, in the case of the parameter σ_j which maximizes the log-likelihood of $Y = \log(X)$:

$$\frac{\partial}{\partial\sigma_j}\left\{\sum_{i=1}^{N}\sum_{j=1}^{J}\gamma_{i,j}\log f_Y\left(y_i \mid \hat{\mu}_j, \sigma_j, k_j\right)\right\} = 0. \tag{54}$$

It results in

$$\frac{\widetilde{S}_1}{\widetilde{S}_0} - \frac{\sigma_j}{\sqrt{k_j}} - \frac{\sum_{i=1}^{N}\gamma_{i,j}y_i}{\sum_{i=1}^{N}\gamma_{i,j}} = 0. \tag{55}$$

This equation is well behaved and all the theoretical demonstrations obtained in [17] still hold: it is monotone decreasing and, when $\lim\sigma_j \to 0$, the function takes the value

$$y_{\max} - \frac{\sum_{i=1}^{N}\gamma_{i,j}y_i}{\sum_{i=1}^{N}\gamma_{i,j}} > 0. \tag{56}$$

As a conclusion, (55) has always a positive solution for any μ_j and k_j. Additionally, due to the invariance of the ML estimates for the transformation $Y = \log(X)$, there always exist a p_j for any a_j and ν_j.

The solution is in the interval

$$0 < \hat{\sigma}_j < \sqrt{k_j}\left(y_{\max} - \frac{\sum_{i=1}^{N}\gamma_{i,j}y_i}{\sum_{i=1}^{N}\gamma_{i,j}}\right). \tag{57}$$

So, the value can be calculated by any numerical method. We used here also the Brent's algorithm.

So, finally, from an initial guess of p_j one can calculate $k_j \equiv \nu_j$ from (47) and then use it to calculate the estimate of σ_j from (55), in an iterative way until a desired tolerance is reached.

This methodology generalizes the proposed method of [17] for the case of GGMM and we will refer to it as the $GGMM_2$ method.

2.4. Implementation Generalized Gamma Mixture Model.

In this section, we detail the implementation of both of the proposed methods for the GGMM.

In Algorithm 1, the Nelder-Mead method [24] was used for the calculation of $p_j^{(n)}$ and the Brent's algorithm [25] for $\nu_j^{(n)}$ in the interval given in (49).

In the case of Algorithm 2, the Brent's algorithm [25] was used for calculating σ and k in the intervals of (57) and (49), respectively.

The computational complexity of the previous GGMM methods when compared to the calculation of a simple GG depends on the number of components, J, assumed by the model. In each iteration of the EM algorithm, the expected parameters of each component have to be calculated. So, if the time consumed to estimate a GG is T, the calculation of the expected parameters of the mixture is $J \times T$.

When other mixture models such as RMM, NMM, and GMM are considered, the complexity of the EM method is similar to the GGMM. Note that both the GMM and the NMM need to solve a non-linear equation similar to (47) so the consumed time of the solution is the same. The additional cost of calculating the GGMM parameters is due to the calculation of the estimate of the parameter $p_j^{(n)}$. If we define the time to solve (47) and (45) as T_1, and T_2 as the time consumed for solving $p_j^{(n)}$, the computational time for a simple GG (T_{GG}) and a GGMM of J components (T_{GGMM}) would be

$$T_{GG} = T_1 + T_2,$$

$$T_{GGMM} = J \cdot (T_1 + T_2). \tag{58}$$

The estimated times in a Matlab (R2011a) implementation running in an ASUS G53SW laptop (Intel Core i7 2630QM Processor, 2.2 GHz, 8 GB RAM) were $T_1 = 1.637$ ms and $T_2 = 0.2056$ s.

3. Results and Discussion

3.1. Performance of the ML Method.

In this section, we show the performance of the proposed methods for calculating the parameters of a GG distribution. For this purpose, we performed 200 synthetic experiments and tested the methods presented in Section 2.2. Concretely, we tested the method of Stacy and Mihram [14], Gomes et al. [16], Noufaily and Jones [17], and our proposed method of Section 2.2.4. We will refer to them as *Stacy*, *Gomes*, *Noufaily,* and *proposed* methods, respectively.

The synthetic data was calculated in the same way as was done in [17]: a set of gamma-distributed random samples are generated by means of the method proposed in [29] and the GG-distributed data are obtained by taking the $1/p$th power of the samples. The parameters of the GG distribution were also calculated from sets of parameters in a reasonable dynamic range. The scale parameter a was set to 1 in all the experiments since this parameter just affects to the scale of the data. Both, the p parameter and the ν parameter were obtained from random samples of a uniform RV in the interval $[0.3, 5]$.

$\{x\}_{i=1}^N \leftarrow$ Samples
$J \leftarrow$ Number of components
$\Theta^{(0)} \equiv \{\pi_j^{(0)}, a_j^{(0)}, \nu_j^{(0)}, p_j^{(0)}\}_{j=1}^J \leftarrow$ Initial guess of parameters for
each component
$\gamma_{i,j}^{(0)} \leftarrow p(Z_i = j \mid x_i, \Theta^{(0)})$
maxIter $\leftarrow$ Maximum number of iterations
Tol $\leftarrow$ Tolerance
Err $\leftarrow \infty$
$n \leftarrow 0$ Iterations counter
while err $>$ Tol and $n <$ maxIter **do**
$\quad n \leftarrow n+1$
$\quad$**for** $j = 1 \leftarrow J$ **do**

$$p_j^{(n)} \longleftarrow \underset{pj}{\arg\max} \left\{ \sum_{i=1}^N \gamma_{i,j}^{(n-1)} \log f_X \left(x_i \mid \frac{\sum_{i=1}^N \gamma_{i,j} x_i^{p_j}}{\nu_j(p_j) \sum_{i=1}^N \gamma_{i,j}} \right) \right\}$$

$$\nu_j^{(n)} \longleftarrow \log(\nu_j) - \Psi(\nu_j) = \log \left(\frac{\sum_{i=1}^N \gamma_{i,j} x_i^{p_j^{(n)}}}{\sum_{i=1}^N \gamma_{i,j}} \right)$$
$$+ \frac{\sum_{i=1}^N \gamma_{i,j} \log(x_i^{p_j^{(n)}})}{\sum_{i=1}^N \gamma_{i,j}}$$

$$a_j^{(n)} \longleftarrow \left(\frac{\sum_{i=1}^N \gamma_{i,j} x_i^{p_j^{(n)}}}{\nu_j^{(n)} \sum_{i=1}^N \gamma_{i,j}} \right)^{1/p_j^{(n)}}$$

$\quad\quad\pi_j^{(n)} \leftarrow (1/N) \sum_{i=1}^N \gamma_{i,j}^{(n-1)}$
$\quad\quad\gamma_{i,j}^{(n)} \leftarrow p(Z_i = j \mid x_i, \Theta^{(n)})$
$\quad$**end for**
$\quad$err $\leftarrow \|\Theta^{(n)} - \Theta^{(n-1)}\|/\|\Theta^{(n-1)}\|$ Evaluate the relative tolerance
end while
Return $\Theta^{(n)}$

ALGORITHM 1: Implementation of the GGMM$_1$ method.

We choose this interval since lower values than 0.3 make the distribution to take values that tend to infinity as ν get closer to 0. This is an unrealistic situation when real images are considered. Additionally, when p takes lower values, the tail becomes heavier and the shape of the distribution also becomes unrealistic. These effects are shown in Figure 1, where some examples of the PDFs of the GG distribution are depicted.

The number of iterations for the proposed method and for the method of [17] was set to 100 and the tolerance function to 10^{-8}. The number of bins where the χ^2-test was performed in the method of [16] was 150 and the number of samples per experiment was 10^4. The comparisons of the methods were performed by comparing the goodness-to-fit of each distribution by means of two different measures: Kullback-Leibler divergence (KL) and Kolmogorov-Smirnov (KS) statistic. The former is a nonsymmetric measure of the difference between two probability distributions defined as

$$D_{\text{KL}}(p_n, f_X) = \sum_{i=1}^N p_n(i) \log \frac{p_n(i)}{f_X(i)}, \tag{59}$$

where p_n is the empirical PDF estimate and f_X is the theoretical distribution (the GG distribution). For the empirical estimate of the PDF, the number of bins of the histogram was set to 150.

The Kolmogorov-Smirnov statistic is the uniform norm of the cumulative distribution function (CDF), defined as

$$D_{\text{KS}} = \sup \left| \widehat{F}(i) - F_X(i) \right|, \tag{60}$$

where $\widehat{F}$ is the empirical CDF of data and F_X the theoretical CDF. The KS measure was chosen since it does not depend on the PDF estimate and can be calculated with a few number of samples. Additionally, the Glivenko-Cantelli theorem states that if the samples are drawn from distribution FX, then D_{KS} converges to 0 almost surely [30].

In Figure 2, the results for both measures are depicted. It is clear that the moments method of Stacy gives poorer results than the other methods for both measures. This result was expected since the moments method depends on moments of third-order, so the variance of the estimates becomes higher. The rest of the methods performed well for both measures. In the case of the D_{KL}, they fit practically the same while, in the case of D_{KS}, there are some better results for the method of Noufaily and the proposed one. This is the effect of the approximation of the PDF for the χ^2 test performed by the method of Gomes: it calculates the best set of parameters for an approximation of the empirical PDF which

$\{x\}_{i=1}^{N} \leftarrow$ Samples

$\{y\}_{i=1}^{N} \leftarrow \{\log(x_i)\}_{i=1}^{N}$ Transformed samples

$J \leftarrow$ Number of components

$\Theta^{(0)} \equiv \{\pi_j^{(0)}, a_j^{(0)}, v_j^{(0)}, p_j^{(0)}\}_{j=1}^{J} \leftarrow$ Initial guess of parameters for
each component

$\gamma_{i,j}^{(0)} \leftarrow p(Z_i = j \mid x_i, \Theta^{(0)})$

maxIter $\leftarrow$ Maximum number of iterations

Tol $\leftarrow$ Tolerance

err $\leftarrow \infty$

$n \leftarrow 0$ Iterations counter

maxIterML $\leftarrow$ Maximum number of iterations for the ML algorithm

TolML $\leftarrow$ Tolerance for the ML

while err > Tol and n < maxIter **do**

   $n \leftarrow n + 1$

   for $j = 1 \rightarrow J$ **do**

      $(a_{\mathrm{aux}}, v_{\mathrm{aux}}, p_{\mathrm{aux}}) \leftarrow (a_j^{(n-1)}, v_j^{(n-1)}, p_j^{(n-1)})$

      errML $\leftarrow \infty$

      $m \leftarrow 0$ Iterations counter for ML

      $\mathcal{LL}_j^{(0)} \leftarrow$ Calculate the log-likelihood for the jth component

      while errML > TolML and m < maxIterML **do**

         $m \leftarrow m + 1$

         $\sigma = 1/(p_{\mathrm{aux}}\sqrt{v_{\mathrm{aux}}})$

$$k \longleftarrow \log(v_j) - \Psi(v_j) = \log\left(\frac{\sum_{i=1}^{N} \gamma_{i,j} x_i^{p_{\mathrm{aux}}}}{\sum_{i=1}^{N} \gamma_{i,j}}\right)$$
$$+ \frac{\sum_{i=1}^{N} \gamma_{i,j} \log(x_i^{p_{\mathrm{aux}}})}{\sum_{i=1}^{N} \gamma_{i,j}}$$

$$\widetilde{S}_0 \leftarrow \frac{\sum_{i=1}^{N} \gamma_{i,j} \exp(y_i/\sigma\sqrt{k})}{\sum_{i=1}^{N} \gamma_{i,j}}$$

$$\widetilde{S}_1 \leftarrow \frac{\sum_{i=1}^{N} \gamma_{i,j} y_i \exp(y_i/\sigma\sqrt{k})}{\sum_{i=1}^{N} \gamma_{i,j}}$$

$$\sigma \leftarrow \frac{\widetilde{S}_1}{\widetilde{S}_0} - \frac{\sigma}{\sqrt{k}} - \frac{\sum_{i=1}^{N} \gamma_{i,j} y_i}{\sum_{i=1}^{N} \gamma_{i,j}} = 0$$

         $v_{\mathrm{aux}} \leftarrow k$

         $p_{\mathrm{aux}} \leftarrow 1/(\sigma\sqrt{k})$

$$a_{\mathrm{aux}} \leftarrow \left(\frac{\sum_{i=1}^{N} \gamma_{i,j} x_i^{p_{\mathrm{aux}}}}{v_{\mathrm{aux}} \sum_{i=1}^{N} \gamma_{i,j}}\right)^{1/p_{\mathrm{aux}}}$$

         $\mathcal{LL}_j^{(m)} \leftarrow$ Calculate the log-likelihood for the jth component

         errML $\leftarrow \|\mathcal{LL}_j^{(m)} - \mathcal{LL}_j^{(m-1)}\|/\|\mathcal{LL}_j^{(m-1)}\|$

      end while

      $(a_j^{(n)}, v_j^{(n)}, p_j^{(n)}) \leftarrow (a_{\mathrm{aux}}, v_{\mathrm{aux}}, p_{\mathrm{aux}})$

      $\pi_j^{(n)} \leftarrow (1/N) \sum_{i=1}^{N} \gamma_{i,j}^{(n-1)}$

      $\gamma_{i,j}^{(n)} \leftarrow p(Z_i = j \mid x_i, \Theta^{(n)})$

   end for

   err $\leftarrow \|\Theta^{(n)} - \Theta^{(n-1)}\|/\|\Theta^{(n-1)}\|$ Evaluate the relative tolerance

end while

ALGORITHM 2: Implementation of the GGMM$_2$ method.

depends on the number of bins and the number of samples of the dataset. So, as the number of samples is reduced or the number of bins is reduced, the estimate becomes worse.

In order to see the effect of this, we also show in Figure 3 the relative error of the estimates for all the methods (the relative error of an estimate $\hat{\theta}$ is calculated as $\epsilon_{\mathrm{rel}} = \|\theta - \hat{\theta}\|/\theta$, while the absolute error is $\epsilon_{\mathrm{abs}} = \|\theta - \hat{\theta}\|$). In the figure, the whiskers show the dynamic range of the data which is not considered an outlier. So, though the method of Gomes provides good fitting, the variance of the estimates is higher than the method of Noufaily and the proposed one. At first sight, the results of Figure 3 demonstrate the better performance of the proposed method in terms of variance of the ML estimates with no appreciable bias in the estimates.

(a) GG($a = 1, \nu, p = 1$)

(b) GG($a = 1, \nu, p = 2$)

(c) GG($a = 1, \nu, p = 3.5$)

(d) GG($a = 1, \nu, p = 5$)
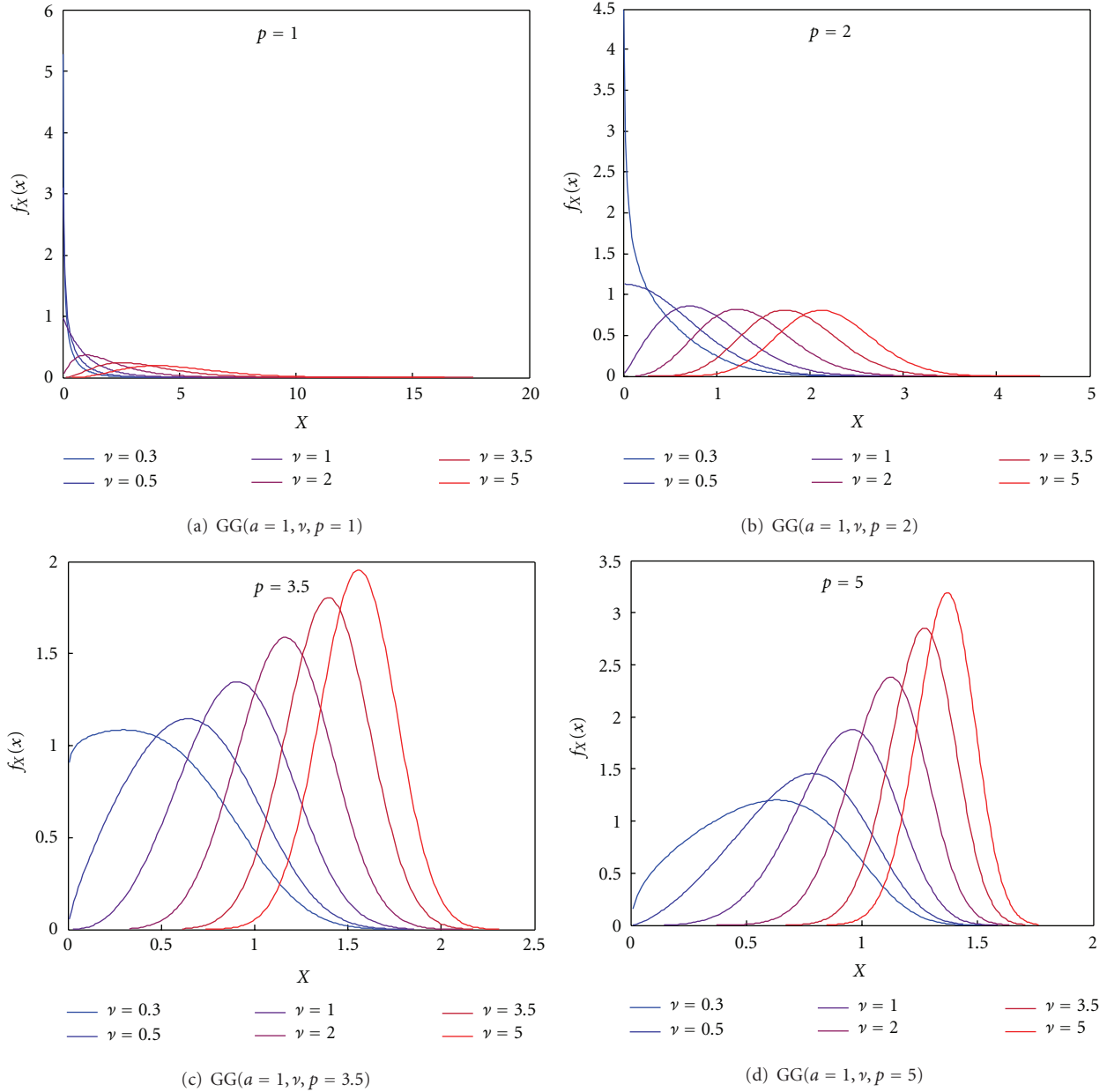
Figure 1: Some examples of the GG distribution for the parameters of the synthetic dataset.

An example of the fitting performance of the methods is shown in Figure 4 where the PDFs obtained with the methods are depicted as well as the absolute error and the relative error of the PDFs.

Following, we analyze the dependence of the estimates with the number of samples. The same experiment is repeated considering 500 samples. The results of both goodness-to-fit measures are shown in Figure 5, and the relative errors of the estimates are depicted in Figure 6. The performance for the D_{KL} measure is similar for all the methods. However, note that the value is considerably higher than the obtained for the case of 10^4 samples, this effect is caused by the difficulties of estimating the PDF with so few samples. Since the Gomes algorithm is based on the χ^2 test, it is

expected that its performance decreases and the variance of the parameter estimates increases. In the case of the D_{KS} measure, the performance of all methods is comparatively equal to the case of 10^4 samples but a higher variability is observed in the Gomes method due to the sensitivity to the number of samples.

The better performance of Noufaily and the proposed methods are seen in Figure 6 where the variability of the Noufaily method did not increase dramatically as the Gomes method did. The proposed method also presented a very low variance of the parameter estimates with no appreciable bias. In the light of these results, we can conclude that the proposed method is robust with respect to the number of samples and it does not introduce any appreciable bias in
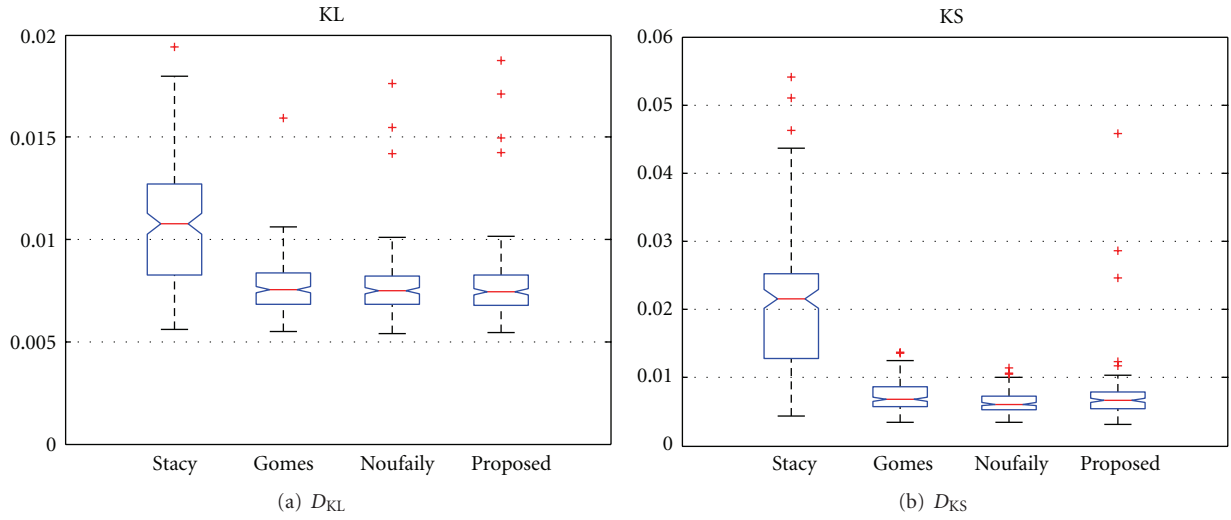
FIGURE 2: Results for D_{KL} and D_{KS} for 10^4 samples. Methods: Stacy and Mihram [14], Gomes et al. [16], Noufaily and Jones [17], and the proposed one of Section 2.2.4.
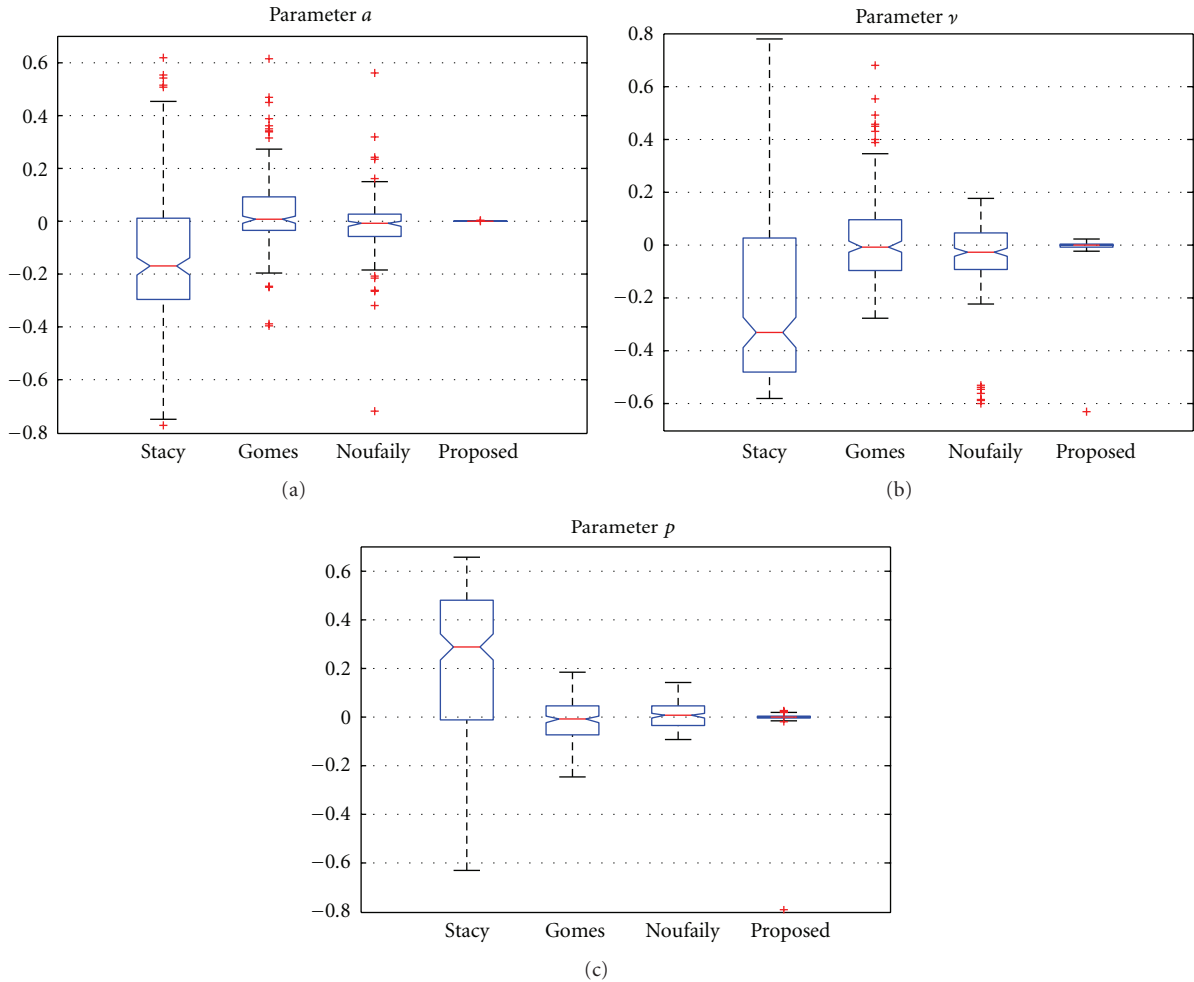
FIGURE 3: Results for the relative error of the estimates for 10^4 samples. Methods: Stacy and Mihram [14], Gomes et al. [16], Noufaily and Jones [17], and the proposed one of Section 2.2.4.
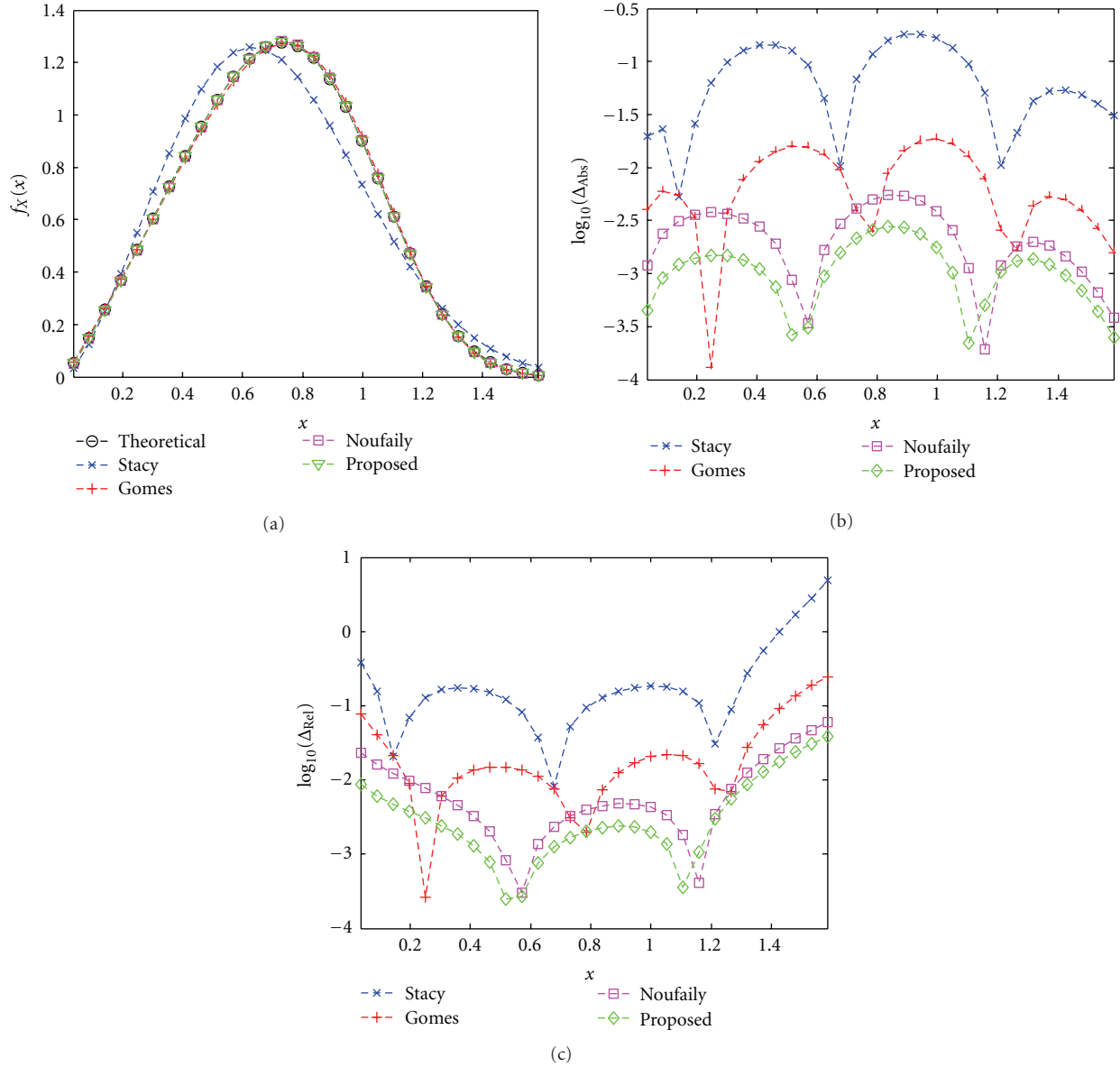
FIGURE 4: Example of the fitting performance for 10^4 samples. Methods: Stacy and Mihram [14], Gomes et al. [16], Noufaily and Jones [17], and the proposed one of Section 2.2.4. (a) Probability Density Functions, (b) Absolute error of the PDFs, and (c) Relative error of the PDFs.

the parameter estimates. The goodness-of-fit performance of both the Noufaily and the proposed method are similar, though the estimates are more accurate with the proposed method. This can be due to the better convergence of the Nelder-Mead method than the algorithm of the Noufaily method.

3.2. Tissue Characterization in Real US Images. In this section, we test the performance of the GG distribution for characterizing tissues of real images. For this purpose, we used a set of 518 real US images (584×145, 8 bits) obtained from 3 human subjects by means of a clinical machine GE Vivid 7 echographic system (GE Vingmed Ultrasound

AS, Horten, Norway). The images were acquired before the Cartesian rearrangement. The image collection was supervised by specialists Marta Sitges and Etelvino Silva (Hospital Clinic IDIBAPS Universitat de Barcelona, Spain). The subjects were volunteers for a study of the reconstruction process of ultrasonic images. The acquisition was done in the Hospital Clinic of Barcelona with its approval. The images were provided by Nicolas Duchateau (CISTIB-Universitat Pompeu Fabra, Ciber-BBN, Barcelona, Spain) and Bart Bijnens (Instituco Catalana de Recerca i Estudis Avan cats (ICREA), Spain).

In **Figure 7**, an example of a real US images is shown with its Cartesian rearrangement. The red contour is the

(a) D_{KL}

(b) D_{KS}
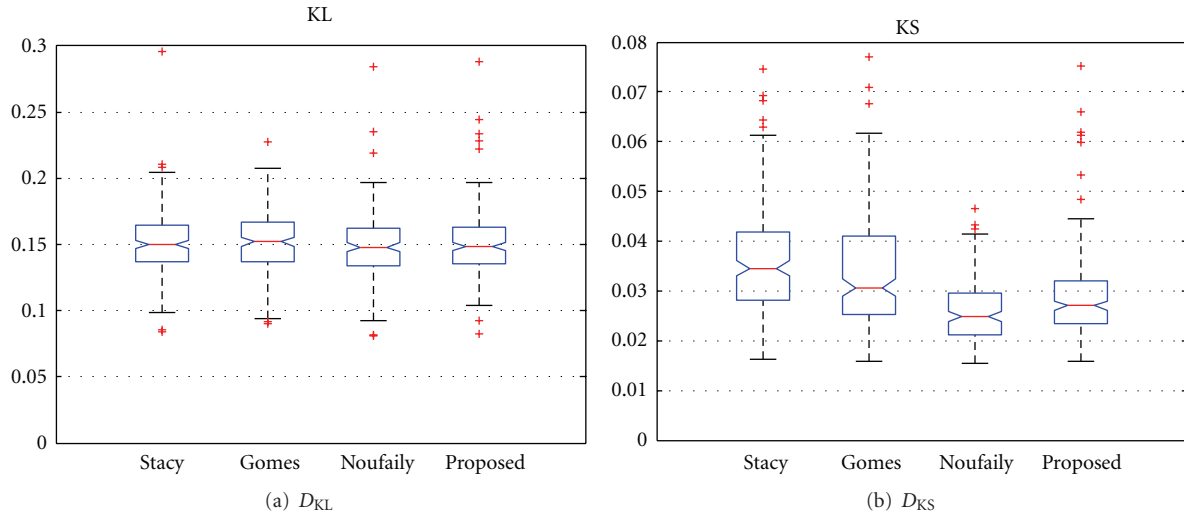
FIGURE 5: Results for D_{KL} and D_{KS} for 500 samples. Methods: Stacy and Mihram [14], Gomes et al. [16], Noufaily and Jones [17], and the proposed one of Section 2.2.4.

(a)

(b)

(c)
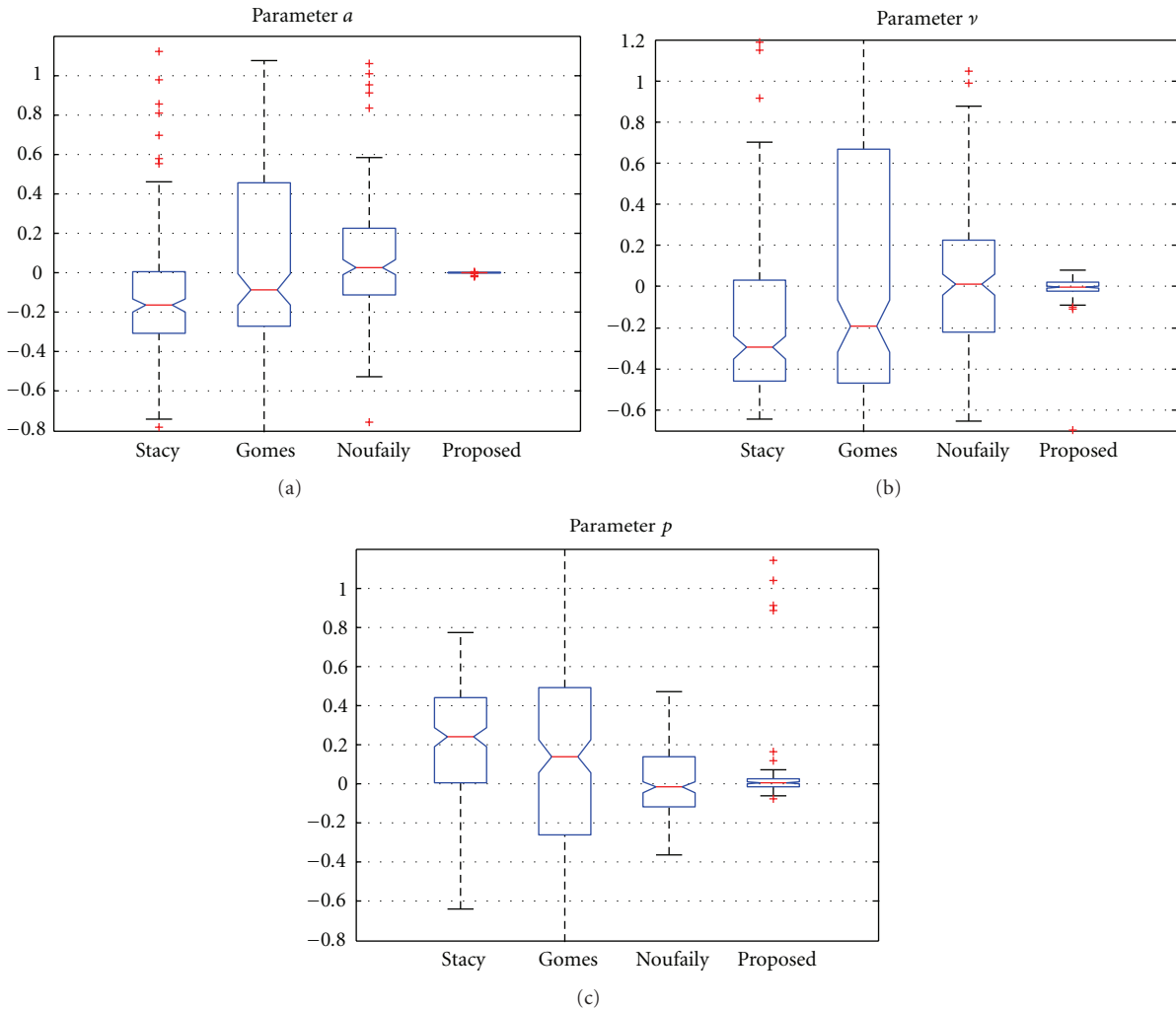
FIGURE 6: Results for the relative error of the estimates for 500 samples. Methods: Stacy and Mihram [14], Gomes et al. [16], Noufaily and Jones [17], and the proposed one of Section 2.2.4.

TABLE 1: Results of the t-test for blood.

Blood	P value	Hypothesis
Nakagami versus Gamma	$<10^{-15}$	H_1
Gamma versus GG	$3.38 \cdot 10^{-7}$	H_1
Nakagami versus GG	$<10^{-15}$	H_1

TABLE 2: Results of the t-test for myocardial tissue.

Myocardial tissue	P value	Hypothesis
Nakagami versus Gamma	$3.24 \cdot 10^{-4}$	H_1
Gamma versus GG	0.96	H_0
Nakagami versus GG	$2.74 \cdot 10^{-4}$	H_1

TABLE 3: Mean values for D_{KS}.

|	Nakagami	Gamma	Generalized Gamma
Blood	$5.5626 \cdot 10^{-2}$	$4.4970 \cdot 10^{-2}$	$\mathbf{4.2860 \cdot 10^{-2}}$
Myocardial	$5.7711 \cdot 10^{-2}$	$5.5665 \cdot 10^{-2}$	$\mathbf{5.5644 \cdot 10^{-2}}$

segmented areas of blood which are considered in the study, while the green contour is the segmented areas of tissue. The intersection of both regions was rejected in the study.

Additionally, the histogram of the image was depicted for the blood region as well as the fitted distributions most commonly used to characterize tissue. From the whole data set, a total number of 3185 regions were segmented for myocardial tissue while 1960 were segmented as blood. The sizes of regions vary depending on the tissue. However, it is high enough to provide a good estimate of the parameters. For instance, the segmented region of Figure 7 has 18250 samples for blood and 5529 for tissue.

In the case of Figure 7, the lower value of the histogram shown is 19 since the intensity values in the blood area were in the interval $[19, 156]$. The number of bins used for the representation of the histogram was set to 20 equally spaced in that interval.

The performance of the GG distributions was tested by estimating the PDFs for both tissue classes (myocardial tissue and blood) for the following distributions: Exponential, Rayleigh, Weibull, Normal, Nakagami, Gamma, and GG. The PDFs were compared by means of both the D_{KL} and the D_{KS} measures. The results of the comparison are depicted in Figure 8 where the better performance of the Gamma, Nakagami, and GG becomes clear. In order to see whether these measures are statistically significant, we carried out a Welch t-test for the Gamma, Nakagami, and GG distributions for the D_{KS} measures. This test was chosen since no equal variance should be assumed and the D_{KS} since it does not depend on the empirical PDF estimate but just on the samples. The assumed hypothesis H_0 is that "both distributions have the same mean," H_1 indicates that the null hypothesis can be rejected at a 5% of significance level.

The results are shown in Tables 1 and 2. Note that all the null hypothesis were rejected but just one: myocardial tissue. In that case, the difference of the mean value of the Gamma and the GG is not statistically significant. The mean values of the D_{KS} are represented in Table 3 where the lower mean value of the GG for both tissues can be appreciated. The results of the t-test of Tables 1 and 2, and the lower mean values of the D_{KS} evidence the better performance of the GG than the rest of the distributions, with the exception of the myocardial tissue, where a Gamma distribution offers the same performance.

3.3. Performance of the GGMM Methods. In this section, we test the performance of the proposed GGMM methods in three different scenarios. First, we test the necessity of using more than a simple GG for describing tissues with an increasing echolucent response of the effective scatterers. The case of a variation of the number of effective scatterers is also considered. This behavior can be found in structures with an increasing deterministic response that changes the speckle nature from fully formed speckle to fully resolved speckle. The variation of the number of effective scatterers can be found in structures which change their scattering cross-section.

In order to simulate B-mode US images, we followed the same methodology proposed in [9]. This method scans an image and records the data in a matrix which is corrupted by means of the speckle formation model of (1) where the tissue is modeled as a collection of scatters of size comparable to the wavelength. The speckle pattern is obtained by means of a random walk which does not assume any statistical distribution in order to avoid any bias of the results. The Cartesian arrangement is obtained by means of linear interpolation of the corrupted samples.

As a first example, we simulate an increasing echolucent tissue which varies its intensity from 0 to 255 from left to right. The sampling process and the resulting B-mode image are shown in Figures 9(a)-9(b). The number of samples were set to 50 angular samples and 100 radial samples, represented as red points. The amplitude of each scatterer is defined as a Normal distributed RV with zero mean and $\sigma = 8$. Note that, along with the variation of intensities from left to right, a specular component of the speckle will appear. The number of scatterers was set 20 in order to simulate fully formed speckle in regions with low echolucent response and fully resolved speckle in regions with high echolucent response. The resulting B-mode image is represented in Figure 9(b).

The fitted GG and GGMM with 2 components depicted in Figure 9(c) show that one simple GG fails to model the probabilistic behavior of a spatially variant echolucent tissue, while a GGMM with 2 components properly describes the echolucent variation.

As an additional experiment, in Figure 10, we represent the spatial variation of the number of effective scatterers. The simulation was performed with the same sampling parameters as was done in the previous experiment. In this case, the echolucid response was set to be homogeneous with no deterministic component. Thus, the nature of the speckle changes from fully formed speckle to partially formed speckle. The number of scatterers decreases from left to right from 256 to 1. The amplitude of each scatterer is defined as a Normal distributed RV with zero mean and $\sigma = 8$.

The speckle PDF in this case becomes more impulsive in areas with more effective scatterers (left part of Figure 10(a)),
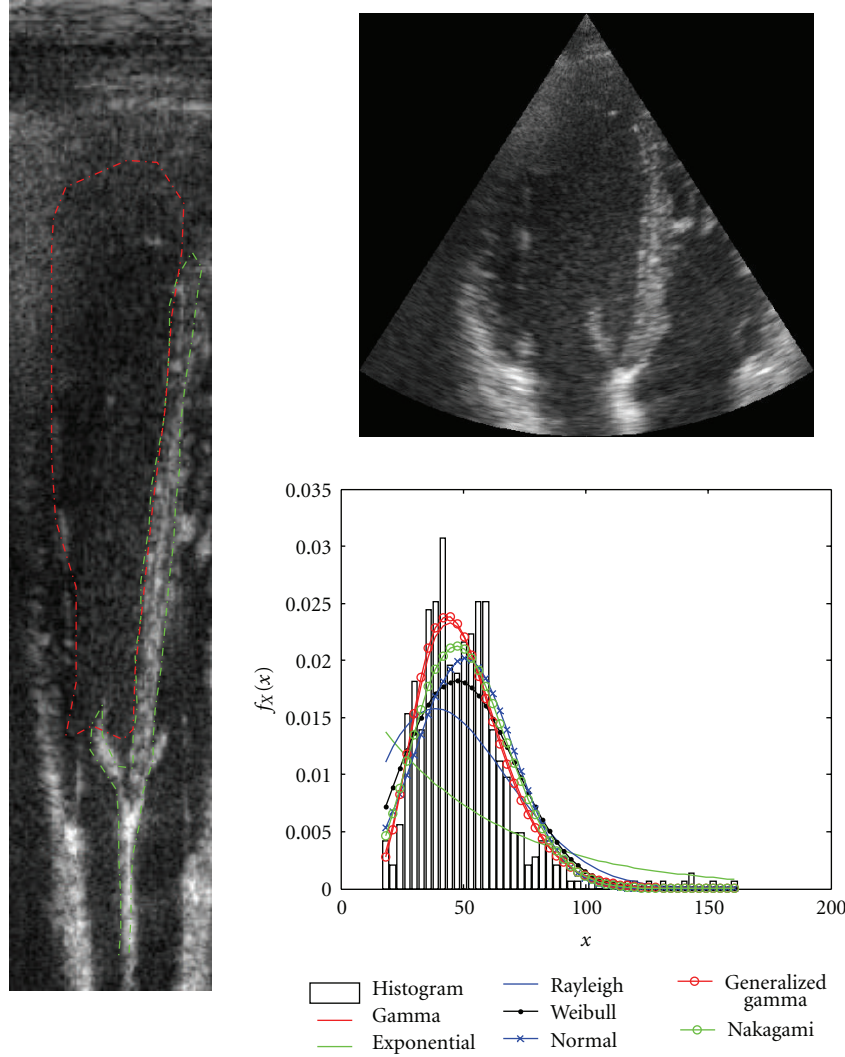
FIGURE 7: Example of an image of the data set. The red contour is the segmented areas of blood which are considered in the study, while the green contour is the segmented areas of tissue. The intersection of both regions was rejected in the study.

this behavior is observed in a lower decay of the tail. Both the simple GG and GGMM with 2 components were calculated from the data and are depicted in Figure 10(b). In that figure, the fitting of a simple GG clearly shows that one component does not suffice to describe a spatial variation of number of effective scatterers.

In the last synthetic experiment for testing the necessity of GGMM, we simulate an anatomic phantom of a kidney scan. For this purpose, we used the artificial kidney scan proposed by Jensen [31]. The image can be downloaded from the Field II website (http://field-ii.dk/). The sampling of the kidney and the resulting B-mode image are represented in Figure 11. In this case, a GGMM with 4 components was used to fit the PDF of the image. The probability of belonging to each component is represented in Figure 12 where the differentiation of tissues can be easily observed. In this case, a lower number of components fail to describe the kidney and the surrounding tissue which have a similar speckle response.

For testing the performance of the proposed GGMM methods with real data, we use the same data set used in the previous section. The number of components is set to two: blood and myocardial tissue. In order to compare the performance of the GGMM methods, we also fit a Gamma Mixture Model and a Nakagami Mixture model to the data [3, 32, 33]. Both the D_{KL} and the D_{KS} where calculated for the mixture models in each image. The number of iterations for each mixture model was set to 100 and the tolerance to 10^{-8}.

The lower values of D_{KL} and D_{KS} shown in Figure 13 evidences the better characterization of the GGMM when compared to the NMM or the GMM. These results were expected due to the results of the previous section. Again, the t-tests were performed to the D_{KS} measure of the data. All the mixtures were statistically different with the exception of the $GGMM_1$ and $GGMM_2$. In that case, a P value of 0.4906 was obtained. These results show once more that the GG

(a) D_{KL} for blood

(b) D_{KS} for blood

(c) D_{KL} for myocardial tissue

(d) D_{KS} for myocardial tissue
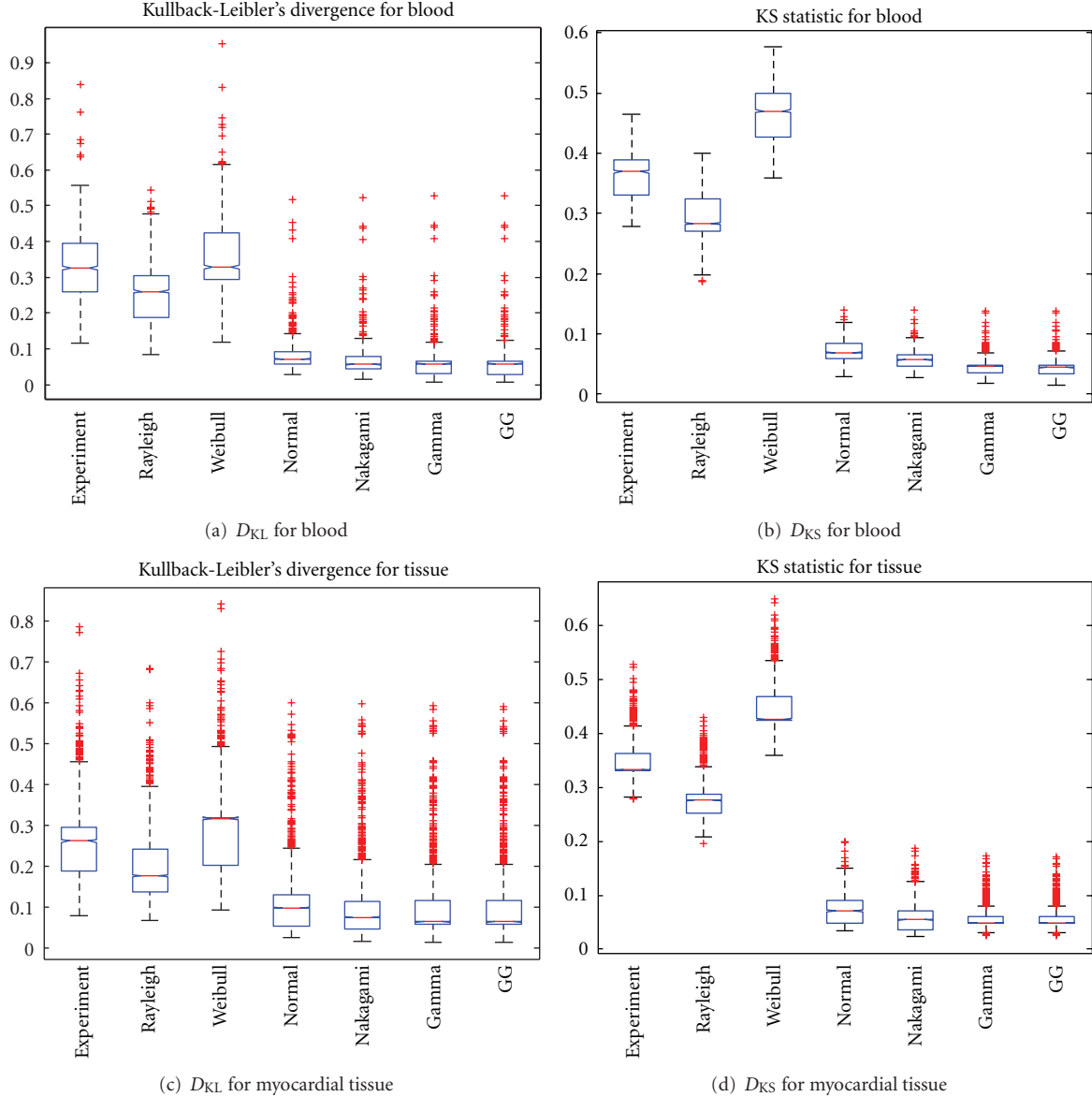
FIGURE 8: Results for the relative error of the estimates for 500 samples. Methods: Stacy Stacy and Mihram [14], Gomes et al. [16], Noufaily and Jones [17], and the proposed one of Section 2.2.4.

can characterize better than other commonly accepted distributions and the differences are significant.

3.4. Potential Applications of the GGMM. A proper characterization of the speckle by means of suitable distributions can be used to guide segmentation algorithms as the one in [3]. The parameters of the mixture model can be used as features for developing a classifier as was done in [4]. Furthermore, some filters use the probability of belonging to each tissue class. As an example of the performance of the GGMM, we show some results of the Probabilistic-Driven Oriented Speckle Reducing Anisotropic Diffusion (POSRAD) [8].

This last filter includes the probability of belonging each tissue class and adapts the diffusion tensor. Concretely, it calculates the structure tensor of the posterior probability and

detects the most probable edges of the image. This information is used to define the diffusion tensor which provides a better behavior in the boundaries of the image.

The structure tensor of the probability density function for each tissue class is calculated as:

$$T_j(x_i) = G_\sigma$$
$$* \left(\nabla_\sigma p(Z_i = j \mid x_i, \Theta) \cdot \nabla_\sigma p(Z_i = j \mid x_i, \Theta)^T \right),$$
(61)

where G_σ is a Gaussian kernel of standard deviation σ, and $\nabla_\sigma p(Z_i = j \mid x_i, \Theta)$ is the gradient of the probability density function for each tissue class filtered with a Gaussian kernel of standard deviation σ. Finally, let $\lambda_1^j \geq \lambda_2^j$ be the eigenvalues and $(\mathbf{v}_1^j, \mathbf{v}_2^j)$ their respective eigenvectors. The local

(a) Sampling

(b) Resulting B-mode image

(c) GG and GGMM fitted to data
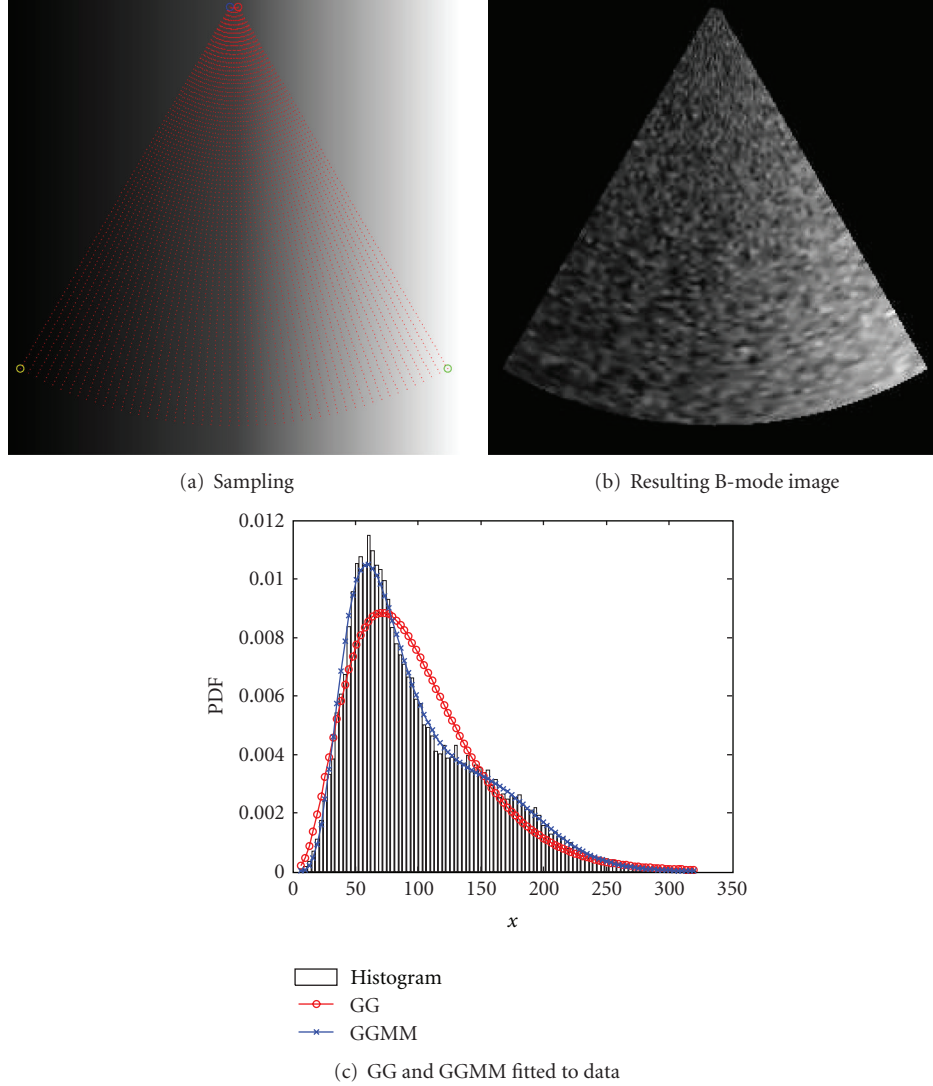
FIGURE 9: Simulation of spatial variant echolucent response of tissue. (a) Sampling of an increasing echolucent tissue. (b) Resulting B-mode image obtained by corrupting the samples by a random walk process of 20 scatterers per resolution cell, in order to simulate a fully formed speckle in regions with low echolucent response and fully resolved speckle in regions with high echolucent response. (c) Histogram of (b) and (a) GG and GGMM with 2 components.

orientation of the maximal variation of probability of the class $\mathcal{C}_j$ is given by $\mathbf{v}_1^j$, and the local orientation of the minimal variation is given by $\mathbf{v}_2^k$.

Let us consider the following diffusion equation:

$$u(x, 0) = u_0$$
$$\frac{\partial u}{\partial t} = \mathrm{div}(D\nabla u), \tag{62}$$

where the matrix D is the diffusion tensor which can be described by its eigenvectors $(\mathbf{v}_1, \mathbf{v}_2)$ and eigenvalues λ_1, λ_2.

Given a diffusion tensor, D, the diffusion of the intensity values of the image is performed in the direction of eigenvectors with different diffusion coefficients. For each eigenvector, its eigenvalue defines the diffusion coefficient and, thus, an anisotropic diffusion can be achieved.

As an example, when one eigenvalue is equal to 1 and the other one is 0, a complete anisotropic diffusion is obtained, since the intensity values diffuse in the direction of the eigenvector associated to the eigenvalue equal to 1. This would be the desired behavior of a filter in regions where structures must be preserved. When both eigenvalues are equal to 1, the diffusion process becomes isotropic and the intensity levels diffuse equally in all directions. This case would be the desired behavior for homogeneous regions where no structures must be preserved.

The POSRAD philosophy makes use of the structure tensors determined out of the probability maps to obtain the most probable structures. In that case, the diffusion filter should be anisotropic. When no probable structures are detected, the diffusion should be isotropic.

Since we have J structure tensors (each tissue class with probability density function), we choose the eigenbase of the

(a) Simulated B-mode image

(b) GG and GGMM fitted to data
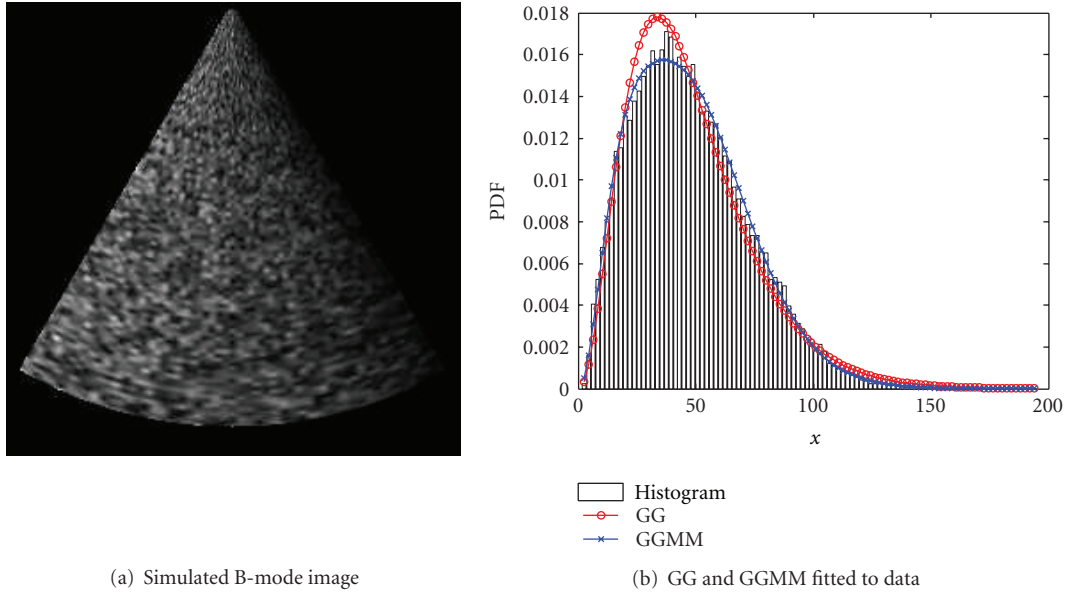
FIGURE 10: Simulation of spatial variant density of scatterers. The number of scatterers per resolution cell decreases from left to right in order to simulate fully formed speckle in regions with low density and partially resolved speckle in regions with high density.

(a) Anatomic phantom of a kidney

(b) Simulated B-mode image

(c) GG and GGMM fitted to data
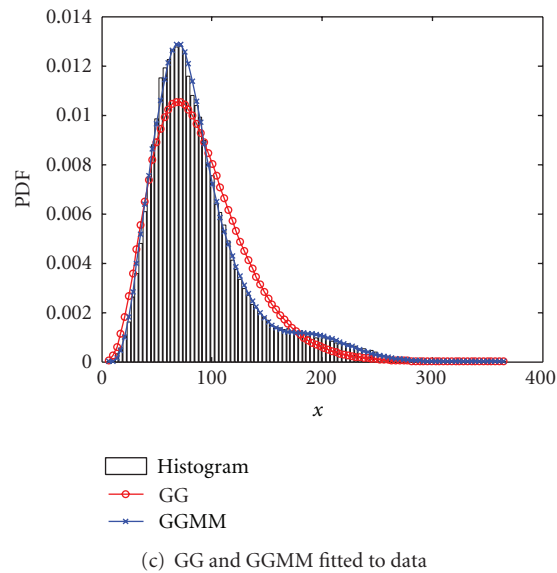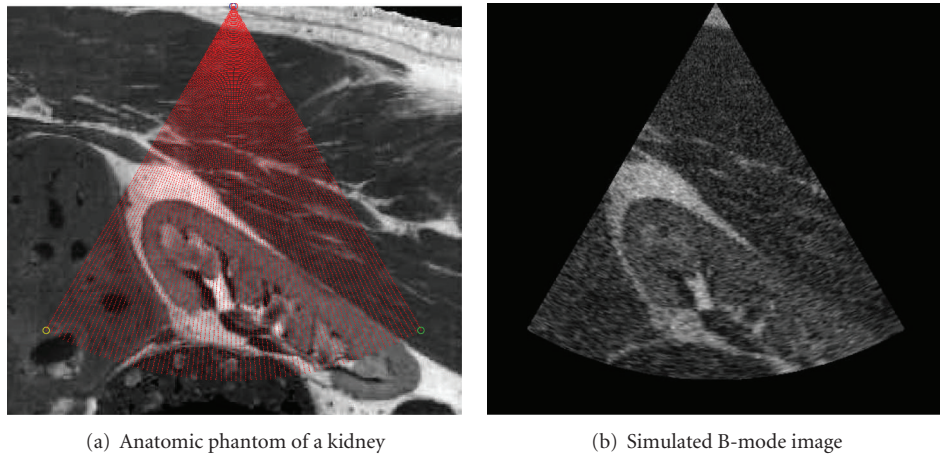
FIGURE 11: Simulation of an anatomic phantom of a kidney scan.

(a)

(b)

(c)

(d)
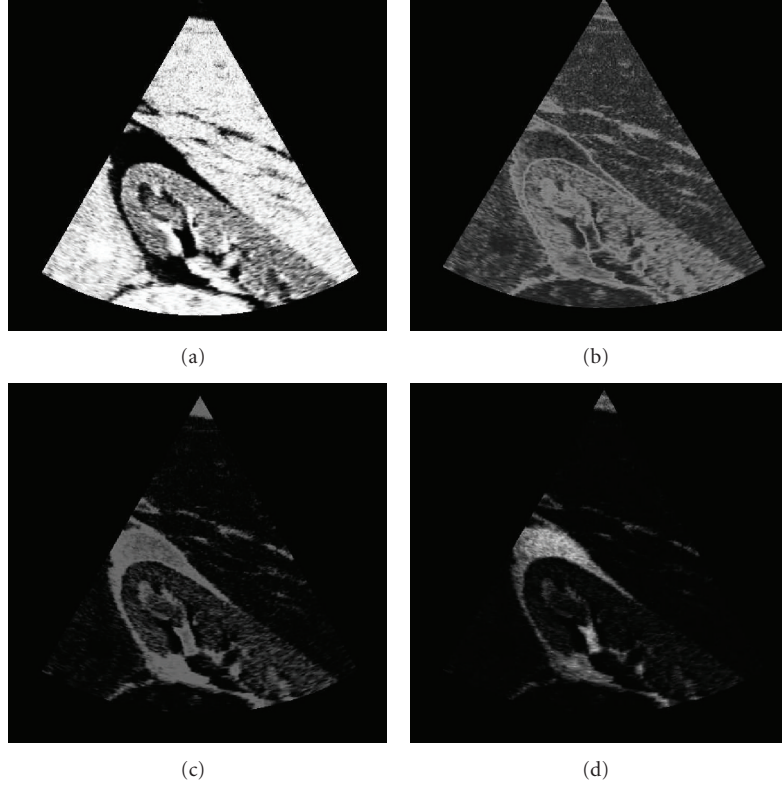
FIGURE 12: Probability of belonging to each component of the GGMM fitted to the image in Figure 11(b). The components are sorted in increasing mean value.

(a) D_{KL}

(b) D_{KS}
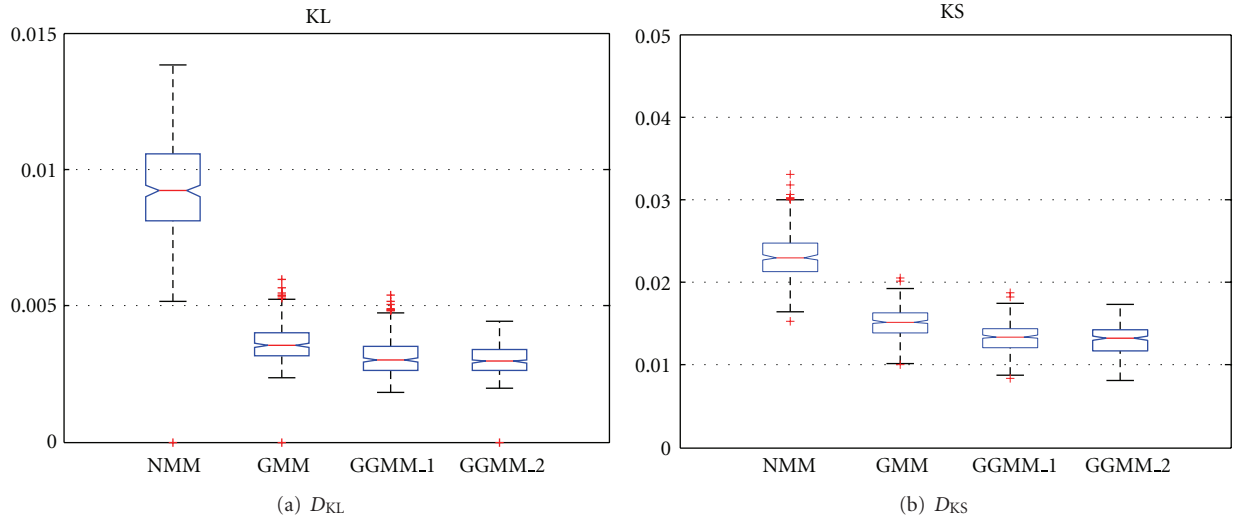
FIGURE 13: Results for D_{KL} and D_{KS} of the Mixture Models: GGMM$_1$, GGMM$_2$, GMM and NMM.

structure tensor with maximal λ_1^j: $\hat{j} = \arg\max_j(\lambda_1^j)$. This base gives the orientation of the maximal variation of probability among all the classes.

The interpretation of this choice is that we choose as boundary the one with the maximal gradient of the probability density function over all tissue classes. This way, the

most probable boundary is preserved in the filtering process. In the basis of $T_{\hat{j}}$, namely, $(\mathbf{e}_1, \mathbf{e}_2)$, the diffusion matrix D is defined as

$$D = E \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} E^T, \tag{63}$$

(a) $p(Z_i = 0 \mid x_i, \Theta)$

(b) $p(Z_i = 1 \mid x_i, \Theta)$

FIGURE 14: Probability of belonging to each tissue class, where the class 0 describes the blood and the class 1 describes the myocardial tissue.
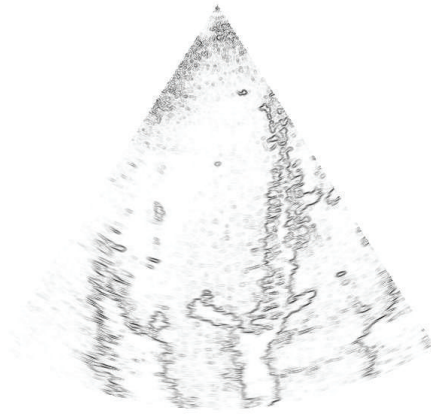
FIGURE 15: Anisotropic behavior of the filter. The most probably edges of the image are described by the lower values of $\lambda_1 \in [0, 1]$.

(a) Original image

(b) Filtered image
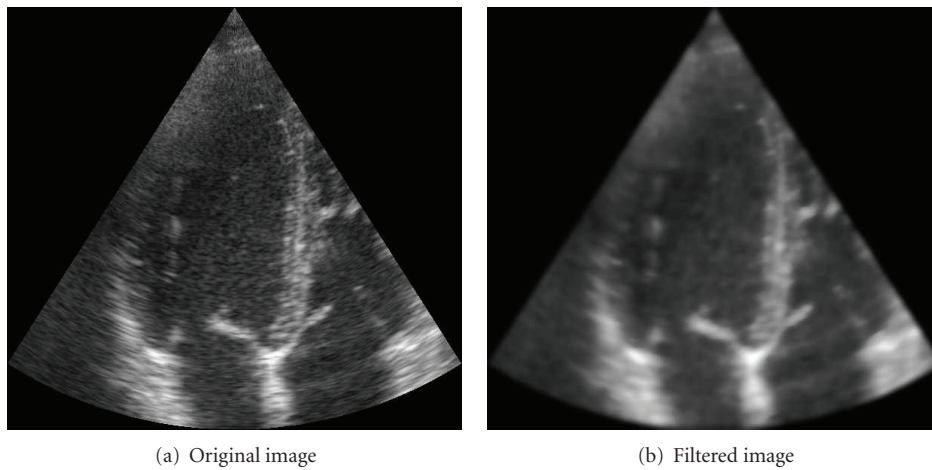
FIGURE 16: Results of the POSRAD filter. The anisotropic behavior of the filter is appreciated in the preserved details of the myocardial tissues.
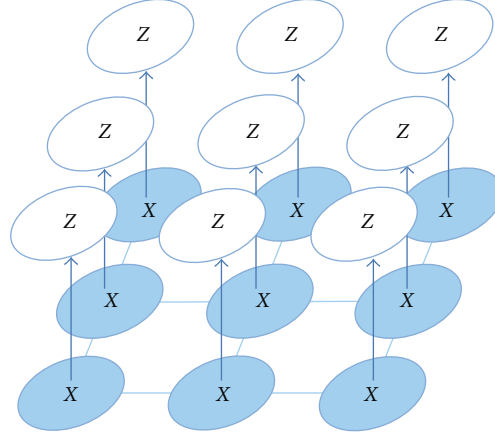
FIGURE 17: Undirected graph. The nodes represent a random variable X and the edges' relationships between nodes. Each random variable can be classified as a tissue class J.

where

$$\lambda_1 = 1 - \|\nabla_{\mathbf{e}_1,\sigma} p(Z_i = j \mid x_i, \Theta)\|_2$$
$$\lambda_2 = 1, \tag{64}$$

and $\| \cdot \|_2$ is the 2-norm, $\nabla_{\mathbf{e}_i,\sigma}$ is the directional derivative in the direction $\mathbf{e}_i$ filtered with a Gaussian kernel with a standard deviation σ, and E is the matrix whose columns are the eigenvectors $(\mathbf{e}_1, \mathbf{e}_2)$. This definition performs a diffusion filtering in the direction of the minimal variation of probability $(\mathbf{e}_2)$ while preserves the maximal variation of probability since $\|\nabla_{\mathbf{e}_1,\sigma} p(Z_i = j \mid x_i, \Theta)\|_2$ will have a value closed to 1. Note that the discrete approximations of $\|\nabla_{\mathbf{e}_1,\sigma} p(Z_i = j \mid x_i, \Theta)\|_2$ is bounded in $[0, 1]$, thus $\lambda_1 \in [0, 1]$.

In Figure 14, we show the probability of belonging to each tissue class, $p(Z_i = j \mid x_i, \Theta)$, provided by the GGMM method (the GGMM$_2$ was used for this example). All the figures of the example are represented in their Cartesian arrangement in order to ease visualization of fine structures. Note that the structures are clearly identified by each posterior probability of each tissue class and the filter can perform an efficient anisotropic diffusion. To see this, in Figure 15, we represent λ_1, which describes the anisotropic behavior of the filter. When $\lambda_1 = 1$, the filter acts like a conventional isotropic filter, whereas the pure anisotropic behavior is carried out when $\lambda_1 = 0$.

Finally, the resulting image after 40 iterations is depicted in Figure 16 in comparison to the original one.

As a final application of the GGMM, one can make use of the pixel-wise probability of belonging to each tissue class to obtain a spatially coherent probability by introducing an undirected graph where the nodes (each pixel of the image) represent a random variable and the edges of the graph represent the relationships between nodes as it is represented in Figure 17. The problem is reduced to find the labels for each node by taking into account the relationships between nodes of the local neighborhood (the Markov property is assumed). This problem, though is intractable in terms of direct probabilistic inference, can be solved by means of the Loopy Belief Propagation (LBP) algorithm introduced by Pear in [34]. This algorithm performs approximate inference of a graphical model. Although LBP does not guarantee to converge due to the presence of loops in the graph, however it has shown good experimental results and is commonly used [35].

In the end, the problem is faced as a discrete MRF where the labels, Z, are each tissue class and the nodes are the pixels of the image. The energy to be minimized by the LBP method can be defined as

$$V(Z) = \sum_{i=1}^{N} V_1(Z_i) + \sum_{k \in \eta(i)} V_2(Z_i, Z_k), \tag{65}$$

where $\eta(i)$ is the neighborhood of the ith node, $V_1(Z_i) = -\log p(Z_i = jx_i, \Theta)$, and

$$V_2(Z_i, Z_k) = - \sum_{k \in \eta(i)} \log p(Z_k = z_i \mid x_k, \Theta). \tag{66}$$

The output of the LBP is a belief of node i belongs to class $Z = j$. Thus, the probability with spatial coherence can be directly obtained from the outputs of the LBP algorithm. In Figure 18, the probability of each tissue class when the spatial coherence is introduced.

These coherent probability maps can be of help for classifying purposes or as prior information for segmentation algorithms. The valuable information that they provide can be seen in a simple experiment in which we consider the classification of two tissues (blood and myocardial tissue) and we compare the results with the k-means algorithm applied to the original image and a simple classifier consisting of assigning the class with maximum posterior probability. The results of this example are shown in Figure 19 where the identification of the myocardial tissue is clearly obtained by the posterior probability of the GGMM, whereas the k-means method cannot properly define a contour of each tissue.

4. Conclusions

Throughout this work we have analyzed the advantages of using a GG distribution for characterizing the speckle in

(a) $p(Z_i = 0 \mid x_i, \Theta)$ (b) $p(Z_i = 1 \mid x_i, \Theta)$
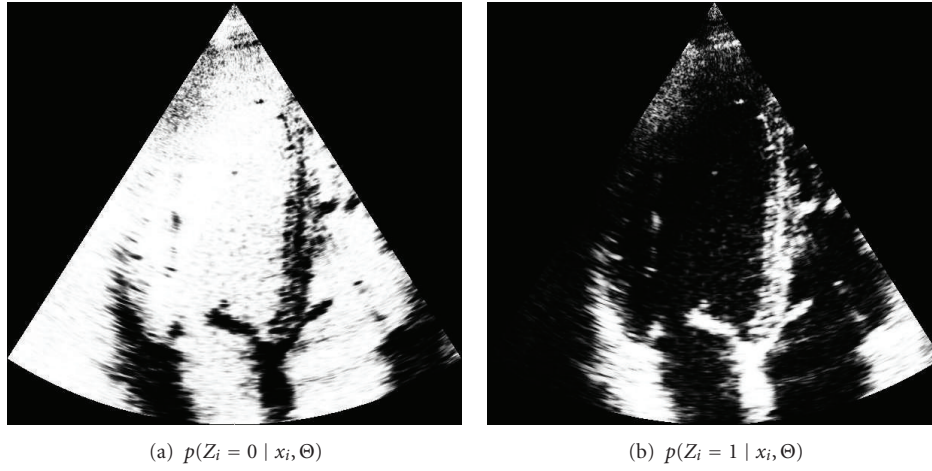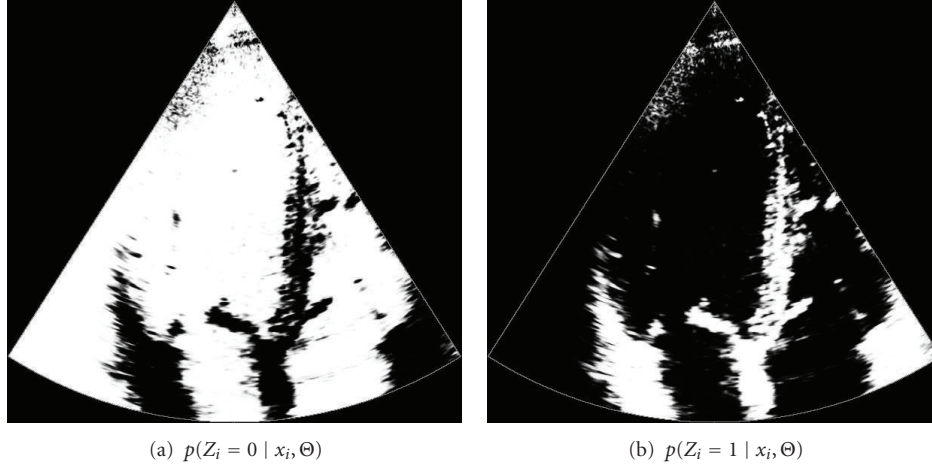
FIGURE 18: Probability of belonging to each tissue class after the LBP, where the class 0 describes the blood and the class 1 describes the myocardial tissue.


(a) Classification with k-means   (b) Classification with the GGMM probability maps after LBP
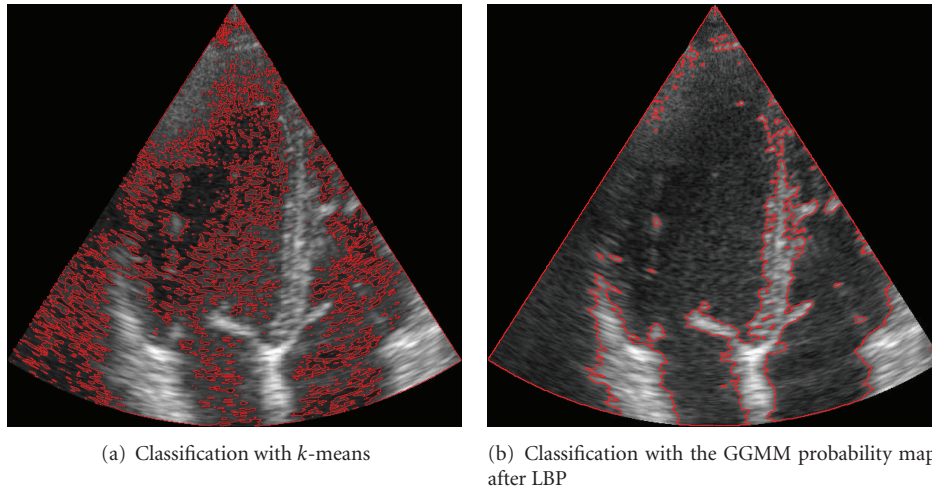
FIGURE 19: Simple example of the valuable information of the posterior probability obtained from the GGMM with spatial coherence.

ultrasound images. This distribution offers a suitable way to deal with the impulsive behavior of speckle which causes heavier tails in the distributions. Additionally, the GG is a natural generalization for many distributions commonly used to characterize the speckle: Rayleigh, Gamma, Nakagami, Weibull, Exponential, and Rician [13]. Thus, it has all the advantages of these distributions and avoids some of their generalization problems.

Although some approaches have used this distribution in the literature, the inconveniences of estimating its parameters make this option thorny and not attractive. The problem stems from the inaccurate estimate of the moments method proposed in [14] and the impossibility of obtaining a closed-form ML estimates. Some solutions have been recently proposed such as heuristic methods [16], which are strongly dependent on the number of samples, and iterative methods [17] which depend on the initial condition.

In this work, we have proposed a simple methodology to calculate the ML estimate which offers robust results comparing to the methods in the literature [14, 16, 17]. It is worth to mention that the fundamentals of the ML method of [17] and the proposed one are the same since both try to find the solution of three simultaneous non-linear equations. However, the different optimization technique makes the proposed method more robust. Additionally, the performance for describing speckle was tested in a set of 518 real US images of the heart, in which 3185 regions were manually segmented for myocardial tissue and 1960 for blood.

Results with t-tests applied to the D_{KS} goodness-of-fit measure demonstrated the better behavior of the GG in most of the cases and in those cases where there were no statistical difference, the other distribution is a particularization of the GG.

The formulation of the proposed method allows to generalize this methodology to a GGMM. These mixture models are of great value due to the different nature of the echogenic response of tissues in the received signal. Two different methods were proposed for the calculation of the GGMM

parameters GGMM$_1$ and GGMM$_2$. Both were developed by applying the EM method in the derivation of the proposed ML method, the optimization technique for GGMM$_1$ follows the same approach used for the proposed ML method. The GGMM$_2$ method makes use of the optimization technique proposed by [17]. Results when comparing both methods to the GMM and NMM in real images showed the better fitting of the GGMM. No statistical differences were detected between GGMM$_1$ and GGMM$_2$.

Through this paper we showed the better behavior of the GGMM methods when compared to the RMM, NMM, and GMM for the case of cardiac imaging. The potentials of mixture models have proven a good classification performance in intravascular ultrasonic images for RMM [4]. Additionally, the NMM showed good results for segmentation in carotid arteries [3]. In the case of filtering Cardiac imaging, the mixture models have also shown good results [9].

We think the GGMM methods here proposed can be used with good results in the aforementioned modalities since they generalize the RMM, NMM, and GMM in a natural way and allow to describe heavier tails of the PDFs that the RMM, NMM, and GMM fail to fit. Many other US modalities such as breast, liver, and kidney should be considered. We hope the proposed GGMM methods can encourage future research for tissue characterization in those different US modalities.

Finally, we want to recall that the potential applications of GGMM do not confine to those proposed in this paper. We hope the results of this work can revive the use of the GG distribution and its extension, the GGMM, in many other areas.

Acknowledgments

References

[1] J. W. Goodman, "Some fundamental properties of laser speckle," in *Laser Speckle and Related Phenomena*, vol. 9–75 of *Topics in Applied Physics*, chapter, pp. 1145–1150, Springer, Heidelberg, Germany, 1975.

[2] T. Eltoft, "Modeling the amplitude statistics of ultrasonic images," *IEEE Transactions on Medical Imaging*, vol. 25, no. 2, pp. 229–240, 2006.

[3] F. Destrempes, J. Meunier, M.-F. Giroux, G. Soulez, and G. Cloutier, "Segmentation in ultrasonic B-mode images of healthy carotid arteries using mixtures of Nakagami distributions and stochastic optimization," *IEEE Transactions on Medical Imaging*, vol. 28, no. 2, pp. 215–229, 2009.

[4] J. C. Seabra, F. Ciompi, O. Pujol, J. Mauri, P. Radeva, and J. Sanches, "Rayleigh mixture model for plaque characterization in intravascular ultrasound," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1314–1324, 2011.

[5] Y. Yu and S. T. Acton, "Speckle reducing anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1260–1270, 2002.

[6] S. Aja-Fernández and C. Alberola-López, "On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2694–2701, 2006.

[7] K. Krissian, C.-F. Westin, R. Kikinis, and K. G. Vosburgh, "Oriented speckle reducing anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 16, no. 5, pp. 1412–1424, 2007.

[8] G. Vegas-Sánchez-Ferrero, S. Aja-Fernández, M. Martín-Fernández, A. F. Frangi, and C. Palencia, "Probabilistic-driven oriented speckle reducing anisotropic diffusion with application to cardiac ultrasonic images," in *Proceedings of the 13th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part I*, pp. 518–525, 2010.

[9] G. Vegas-Sánchez-Ferrero, D. Martíin-Martíinez, S. Aja-Fernández, and C. Palencia, "On the influence of interpolation on probabilistic models for ultrasonic images," in *Proceedings of the 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '10)*, pp. 292–295, April 2010.

[10] Z. Tao, H. D. Tagare, and J. D. Beaty, "Evaluation of four probability distribution models for speckle in clinical cardiac ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 25, no. 11, pp. 1483–1491, 2006.

[11] M. M. Nillesen, R. G. P. Lopata, I. H. Gerrits, L. Kapusta, J. M. Thijssen, and C. L. de Korte, "Modeling envelope statistics of blood and myocardium for segmentation of echocardiographic images," *Ultrasound in Medicine and Biology*, vol. 34, no. 4, pp. 674–680, 2008.

[12] P. M. Shankar, "A general statistical model for ultrasonic backscattering from tissues," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 47, no. 3, pp. 727–736, 2000.

[13] P. M. Shankar, "Ultrasonic tissue characterization using a generalized Nakagami model," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 48, no. 6, pp. 1716–1720, 2001.

[14] E. W. Stacy and G. A. Mihram, "Parameter estimation for a generalized gamma distribution," *Technometrics*, vol. 7, no. 3, pp. 349–358, 1965.

[15] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[16] O. Gomes, C. Combes, and A. Dussauchoy, "Parameter estimation of the generalized gamma distribution," *Mathematics and Computers in Simulation*, vol. 79, no. 4, pp. 955–963, 2008.

[17] A. Noufaily and M. Jones, "On maximization of the likelihood for the generalized gamma distribution," *Computational Statistics*. In press.

[18] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans," *IEEE Transactions on Sonics and Ultrasonics*, vol. 30, no. 3, pp. 156–163, 1983.

[19] V. Dutt and J. F. Greenleaf, "Ultrasound echo envelope analysis using a homodyned K distribution signal model," *Ultrasonic Imaging*, vol. 16, no. 4, pp. 265–287, 1994.

[20] D. R. Wingo, "Computing maximum-likelihood parameter estimates of the generalized gamma distribution by numerical root isolation," *IEEE Transactions on Reliability*, vol. 36, no. 5, pp. 586–590, 1987.

[21] H. W. Hager and L. J. Bain, "Inferential procedures for the generalized gamma distribution," *Journal of the American Statistical Association*, vol. 65, no. 332, pp. 1601–1609, 1970.

[22] B. Jones, W. G. Waller, and A. Feldman, "Root isolation using function values," *BIT Numerical Mathematics*, vol. 18, no. 3, pp. 311–319, 1978.

[23] J. F. Lawless, "Inference in the generalized gamma and log gamma distributions," *Technometrics*, vol. 22, no. 3, pp. 409–419, 1980.

[24] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[25] R. P. Brent, *Algorithms for Minimization without Derivatives*, Prentice Hall, Upper Saddle River, NJ, USA, 1973.

[26] M. E.-S. Waheed, O. A. Mohamed, and M. E. Abd El-Aziz, "Mixture of generalized gamma density-based score function for fastica," *Mathematical Problems in Engineering*, vol. 2011, Article ID 150294, 14 pages, 2011.

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

[28] H. Alzer, "On some inequalities for the gamma and psi functions," *Mathematics of Computation*, vol. 66, no. 217, pp. 373–389, 1997.

[29] G. Marsaglia and W. W. Tsang, "A simple method for generating gamma variables," *ACM Transactions on Mathematical Software*, vol. 26, no. 3, pp. 363–372, 2000.

[30] R. M. Dudley, *Uniform Central Limit Theorems*, Cambridge University Press, Cambridge, UK, 1st edition, 1999.

[31] J. A. Jensen, "Field: a program for simulating ultrasound systems," in *Proceedings of the 10th Nordicbaltic Conference on Biomedical Imaging*, vol. 4, supplement 1, part 1, pp. 351–353, 1996.

[32] A. R. Webb, "Gamma mixture models for target recognition," *Pattern Recognition*, vol. 33, no. 12, pp. 2045–2054, 2000.

[33] K. Copsey and A. Webb, "Bayesian gamma mixture model approach to radar target recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1201–1217, 2003.

[34] J. Pearl, "Reverend Bayes on inference engines: a distributed hierarchical approach," in *Proceedings of the American Association of Artificial Intelligence National Conference on AI*, pp. 133–136, 1982.

[35] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.

Research Article

Extended Finite Element Method with Simplified Spherical Harmonics Approximation for the Forward Model of Optical Molecular Imaging

Wei Li, Huangjian Yi, Qitan Zhang, Duofang Chen, and Jimin Liang

Life Sciences Research Center, School of Life Sciences and Technology, Xidian University, Xi'an, Shaanxi 710071, China

Correspondence should be addressed to Jimin Liang, jiminliang@gmail.com

An extended finite element method (XFEM) for the forward model of 3D optical molecular imaging is developed with simplified spherical harmonics approximation (SP_N). In XFEM scheme of SP_N equations, the signed distance function is employed to accurately represent the internal tissue boundary, and then it is used to construct the enriched basis function of the finite element scheme. Therefore, the finite element calculation can be carried out without the time-consuming internal boundary mesh generation. Moreover, the required overly fine mesh conforming to the complex tissue boundary which leads to excess time cost can be avoided. XFEM conveniences its application to tissues with complex internal structure and improves the computational efficiency. Phantom and digital mouse experiments were carried out to validate the efficiency of the proposed method. Compared with standard finite element method and classical Monte Carlo (MC) method, the validation results show the merits and potential of the XFEM for optical imaging.

1. Introduction

Light propagation model in biological tissue is the foundation of optical imaging. An accurate forward model is important for location and quantification of target distribution in the fields of optical imaging modalities, such as diffusion optical tomography (DOT), fluorescence molecular tomography (FMT), bioluminescence tomography (BLT), and Cerenkov luminescence tomography (CLT) [1–5]. The propagation of the emission photons in tissue can be accurately represented by the radiative transfer equation (RTE) or Monte Carlo (MC) models, but they are extremely computationally expensive. Therefore, the commonly used mathematical model in optical imaging field is the diffusion approximation (DA) to RTE. However, the DA model can be used only in the highly scattering property of the biological tissue, and is not suitable for the real mouse with complex internal tissues. To reach a compromise between the accuracy and efficiency, simplified spherical harmonics approximation (SP_N) to RTE is employed due to its capacity in improving the solution in transport-like domains with high absorption and small geometries [6, 7].

Owing to the complex and curvilinear geometries associated with the biological tissues, the classical finite element methods (FEM) with SP_N approximation become necessary for optical imaging, especially for heterogeneous tissues [8–10]. In the FEM scheme, the region of heterogeneous tissue is divided into small tetrahedron elements. The linear functions of the tetrahedron element are employed in the standard finite element basis function which requires the homogeneity of tissue in one element. The fine triangle mesh between the two tissues is required to conform to the complex internal boundary for ensuring the calculation accuracy of FEM. However, the generation of the internal boundary mesh is a hard work and time consuming task. Moreover, the fine mesh yields the excess cost in finite element computation, especially with SP_N approximation in three dimensions. Fortunately, the extended finite element method inherits all the advantages of standard FEM and exempts the internal boundary mesh generation. Consequently, the extend-finite element method (XFEM) may deal with the problems of fine triangle mesh generation perfectly.

XFEM was first introduced in the literature [11], and it has many applications in the area of mechanics [11, 12].

To the best of our knowledge, it has not been used with SP_N approximation for optical imaging. In this study, we establish the mathematical framework of XFEM with SP_N approximation as the optical forward model. Specifically, a mesh without the internal boundary conformation is employed in the XFEM scheme. The distance relationship between the tetrahedron vertex and the real tissue boundary is reflected by a signed distance function which is used to construct the enriched basis function. Then the enriched function is added to the standard finite element basis function, therefore a standard approximation of FEM is thus "enriched" in the discretized region (usually cut by the internal tissue boundary) of interest. In the weak form of SP_N equations, Gaussian quadrature is employed to calculate the integrals and the linear system equations are established. As a result, the calculation of XFEM can be carried out without the time-consuming internal boundary mesh generation. Moreover, the required overly fine mesh conforming to the complex tissue boundary which leads to excess time cost can be avoided. XFEM conveniences its application to tissues with complex internal structure and improves the computational efficiency. Numerical experiments with a phantom and a digital heterogeneous mouse were carried out to evaluate the performance of the proposed method. The results were compared with that of standard FEM and MC method to demonstrate the efficiency of XFEM.

The paper is organized as follows. In Section 2, the detailed procedure of using XFEM for solving SP_N equations is derived. In Section 3, we evaluate the performance of the proposed method by comparing with the standard FEM and MC method in the experiments, and demonstrate the efficiency of the method. Conclusions and discussions are given in the last section.

2. Method

2.1. SP_N Approximations. The general form of the $(N + 1)/2$ SP_N equations and its $(N + 1)/2$ boundary conditions for optical imaging in three dimensions are [13]:

$$- \nabla \cdot C_{i,\nabla\varphi_i}(r)\nabla\varphi_i(r) + \sum_{j=1}^{(N+1)/2} C_{i,\varphi_i}(r)\varphi_i(r)$$

$$= C_{i,Q}(r)Q(r), \quad r \in \Omega, \ i,j \in \left[1, \frac{(N+1)}{2}\right].$$

$$\sum_{j=1}^{(N+1)/2} C^b_{i,\nabla\varphi_j}(r)\overrightarrow{n} \cdot \nabla\varphi_j(r)$$

$$= \sum_{j=1}^{(N+1)/2} C^b_{i,\varphi_j}(r)\varphi_j(r) + C_{i,S}S_i(r),$$

$$r \in \partial\Omega, i,j \in \left[1, \frac{(N+1)}{2}\right],$$

(1)

where $\varphi_i(r)$ is the SP_N composite moments of the radiances in RTE, and $C_{i,\nabla\varphi_i}(r)$, $C_{i,\varphi_i}(r)$, $C_{i,Q}(r)$, $C^b_{i,\nabla\varphi_j}(r)$, $C^b_{i,\varphi_j}(r)$,

and $C_{i,S}$ denote the coefficients which are related to $\nabla\varphi_i(r)$, $\varphi_i(r)$, Q, and S for SP_N equations at each point r in the region Ω or boundary $\partial\Omega$. These can be calculated by the absorption coefficient $\mu_a[\text{mm}^{-1}]$, scattering coefficient $\mu_s[\text{mm}^{-1}]$, anisotropy parameter g, and refractive index n. $Q(r)$ is the internal source, and it represents the bioluminescence source in BLT or the fluorophore in and FMT. $S_i(r)$ is the external source which represent the external laser source in DOT and FMT. $\overrightarrow{n}$ represents the unit normal vector outward to the boundary.

The exiting partial current $J^+[\text{nW/mm}^2]$ can be obtained from detector readings at the tissue boundary $\partial\Omega$ which can be acquired by CCD camera in practical application, and we have the general formulation:

$$J^+ = \sum_{j=1}^{(N+1)/2} C^J_{\nabla\varphi_j}(r)\overrightarrow{n} \cdot \nabla\varphi_j(r)$$

$$+ \sum_{j=1}^{(N+1)/2} C^J_{\varphi_j}(r)\varphi_j(r),$$

(2)

where $C^J_{\nabla\varphi_j}(r)$ and $C^J_{\varphi_j}(r)$ are the coefficients which can be calculated for SP_N equations. The detailed derivation of (1)–(2) refers to the literature [13].

2.2. Extended Finite Element Discretization. The SP_N equations and its boundary conditions can be solved using Galerkin finite element scheme, the weak form of SP_N equations can be written as follows:

$$\int_\Omega C_{i,\nabla\varphi_i}(r) \cdot \nabla\varphi_i(r) \cdot \nabla v(r)dr$$

$$+ \sum_{j=1}^{(N+1)/2} \int_\Omega C_{i,\varphi_j}(r)\varphi_j(r) \cdot v(r)dr$$

$$- \sum_{j=1}^{(N+1)/2} \int_{\partial\Omega} C_{i,\nabla f_{\varphi_j}}(r)\nabla f_{n\cdot\varphi_i}\big(\varphi_j(r)\big) \cdot v(r)dr$$

(3)

$$= \int_\Omega C_{i,Q}(r)Q(r) \cdot v(r)dr$$

$$+ \int_{\partial\Omega} C_{i,f_{S_i}}(r)f_{S_i}(S_i(r)) \cdot v(r)dr,$$

where $f_{n\cdot\varphi_i}(\varphi_j(r))$ and $C_{i,f_{S_i}}(r)$ are the coefficient matrices with respect to $\varphi_j(r)$ and $f_{S_i}(S_i(r))$. $v(r)$ is the test function, and it is the same as the standard linear basis function in elements. For finite element analysis in three dimensions, tetrahedral elements have become popular in numerical computation, because of their ability to describe complex geometries such as heterogeneous tissues. Thus the volumetric domain is discretized into small tetrahedral elements, and the elements need to conform to the internal tissue boundary as shown in Figures 1(a) and 1(b). When the internal boundary is not smooth, the size of the element must be small enough to ensure the accuracy and this may cause difficulty in mesh generation and lead to huge computation burden. In XFEM framework, mesh is generated as a region
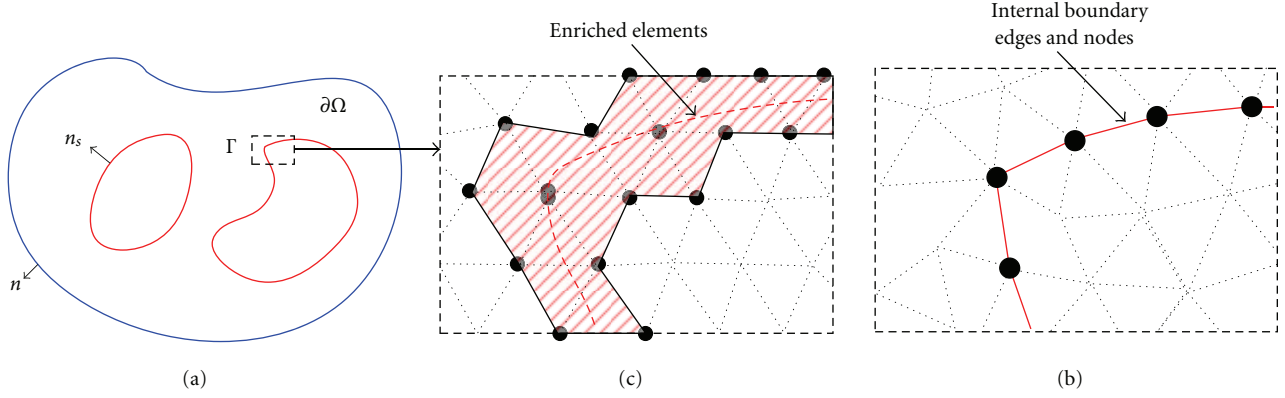
FIGURE 1: (a) Heterogeneous tissues. (b) Standard finite element mesh boundary conforms to the interface. (c) Enriched elements and nodes of XFEM "cut" by the interface.

consists many different tissues, and the elements "cut" by the actual tissue boundary are enriched as shown in Figures 1(a) and 1(c). Using the enriched basis functions constituted by the signed distance functions, the boundary can be determined.

2.3. Enriched Strategy. In XFEM scheme, the signed distance function is adopted to depict the internal boundary of different tissues. The definition of the signed distance function $N(r)$ is

$$N(r) = \min_{r^b \in \Gamma} \| r - r^b \| \operatorname{sign}\left(n_s \cdot \left(r - r^b \right) \right).$$

$$\operatorname{sign}(\xi) = \begin{cases} 1 & \text{if } \xi \geq 0, \\ -1 & \text{if } \xi < 0, \end{cases} \tag{4}$$

where r^b is the point at the internal boundary Γ, sign is the signed function, and n_s denotes the unit normal vector outward to the internal boundary as shown in Figure 1(a). Herein, the position relation between the discrete point and the continuous boundary is completely reflected by the function $N(r)$.

The enriched basis function $\psi_j(r)$ composed of the signed distance function $N(r)$ of the element at tetrahedron node j can be derived, and the continuous $N(r)$ is discretized with its value at the node. The enrichment functions introduce a discontinuity in the gradient of the radiances field $\varphi_i(r)$ or the distribution of optical parameters to Γ, thus the following integral of enriched function is more accurate than that of linear basis function:

$$\psi_j(r) = v_j(r)\big(|N(r)| - |N_j|\big).$$

$$N(r) = \sum_{k=1}^{4} v_k(r) N_k. \tag{5}$$

$$N_j = N(r_j),$$

where $v_j(r)$ is the linear basis function of the tetrahedron element. Using the enriched basis function $\psi_j(r)$, the

enriched approximation $\overline{\varphi}(r)$ can be written as the following form:

$$\varphi(r) \approx \sum_{i=1}^{N_c} \phi_i v_i(r).$$

$$\overline{\varphi}(r) \approx \varphi(r) + \sum_{j=1}^{N_e} a_j \psi_j(r) \triangleq \overline{v}(r)\overline{\varphi}, \tag{6}$$

where $\varphi(r)$ is the conventional approximation and N_c is the number of nodes. $\overline{\varphi}(r)$ is the extended approximation, a_j is the enriched degrees of freedom, and N_e is the number of the enriched nodes. It is clear that $\overline{\varphi}(r)$ includes $\varphi(r)$ as a special case, and each enriched element has eight degrees of freedom. Rewrite the extended linear basis functions and the discrete point value of $\overline{\varphi}(r)$ in matrix form, we obtain

$$\begin{aligned} \overline{v}(r) &= \left[\overline{v}_1(r), \overline{v}_2(r) \cdots \overline{v}_{N_c+N_e}(r) \right] \\ &= \left[v_1(r), v_2(r) \cdots v_{N_c}(r), \psi_1(r), \psi_2(r) \cdots \psi_{N_e}(r) \right]. \\ \overline{\varphi}_i &= \left[\phi_{i,1}, \phi_{i,2} \cdots \phi_{i,N_c}, a_{i,1}, a_{i,2} \cdots a_{i,N_e} \right]^T, \\ &\qquad i \in \left[1, \frac{(N+1)}{2} \right], \end{aligned} \tag{7}$$

where $\overline{v}(r)$ and $\overline{\varphi}_i$ have $N_c + N_e$ components for each variable of SP_N equations. This form can directly use to assemble the finite element matrix.

2.4. Integrals and System Matrix. For finite element discretization, the matrix form of the $(N+1)/2$ SP_N equations (3) can be written as follows:

$$\begin{aligned} &\sum_{\text{elem}} \begin{bmatrix} m_{1,\overline{\varphi}_1} & m_{1,\overline{\varphi}_2} & \cdots & m_{1,\overline{\varphi}_{(N+1)/2}} \\ m_{2,\overline{\varphi}_1} & m_{2,\overline{\varphi}_2} & \cdots & m_{2,\overline{\varphi}_{(N+1)/2}} \\ \vdots & \vdots & \vdots & \vdots \\ m_{(N+1)/2,\overline{\varphi}_1} & m_{(N+1)/2,\overline{\varphi}_2} & \cdots & m_{(N+1)/2,\overline{\varphi}_{(N+1)/2}} \end{bmatrix} \\ &\times \begin{bmatrix} \overline{\varphi}_1 \\ \overline{\varphi}_2 \\ \vdots \\ \overline{\varphi}_{(N+1)/2} \end{bmatrix} = \sum_{\text{elem}} \begin{bmatrix} \overline{b}_1 \\ \overline{b}_2 \\ \vdots \\ \overline{b}_{(N+1)/2} \end{bmatrix}, \end{aligned} \tag{8}$$
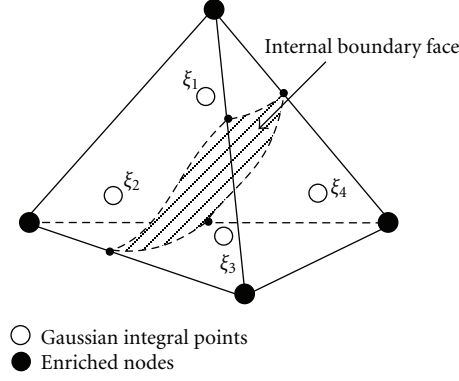
Figure 2: Gaussian quadrature in an enriched element.

where elem denotes all the elements in the region, for the $(N + 1)/2$ variables $\overline{\varphi}$ and $\overline{b}$, each of which has $N_c + N_e$ components in total according to (7). $m_{i,\overline{\varphi}_j}$ in the left coefficient matrix is the small element matrix and can be written in details:

$$m_{i,\overline{\varphi}_j}$$

$$= \begin{cases} \int_{\Omega_e} C_{i,\nabla\varphi_i}(r) \cdot \nabla\overline{v}_p \cdot \nabla\overline{v}_q dr + \int_{\Omega e} C_{i,\varphi_j}(r)\overline{v}_p \cdot \overline{v}_q dr \\ - \int_{\partial\Omega} C_{i,\nabla f_{\varphi_i}}(r) f_{n\cdot\varphi_i}(\overline{v}_p) \cdot \overline{v}_q dr, \quad i = j. \\ \int_{\Omega_e} C_{i,\varphi_j}(r)\overline{v}_p \cdot \overline{v}_q dr \\ - \int_{\partial\Omega} C_{i,\nabla f_{\varphi_j}}(r) f_{\overrightarrow{n}\cdot\varphi_i}(\overline{v}_p) \cdot \overline{v}_q dr, \quad i \neq j, \end{cases}$$

$$(9)$$

where $\overline{v}_p$ and $\overline{v}_q$ are the corresponding matrix elements of $\overline{v}(r)$ in (7) in the finite element Ω_e. p, q are the number marks of the four points in element Ω_e. Similarly, $\overline{b}_i$ in the right is

$$\overline{b}_i = \int_\Omega C_{i,Q}(r)\overline{Q}_p \cdot \overline{v}_q dr + \int_{\partial\Omega} C_{i,f_{S_i}}(r) f_{S_i}(S_i(r)) \cdot \overline{v}_q dr. \tag{10}$$

In finite element framework, for ease of calculation, the integrand is assumed continuous, and the Gaussian quadrature can be used. In this paper, on the one hand, exact integrals can be obtained for the standard linear elements. On the other hand, the second order Gaussian quadrature with four quadrature points is adopted for the remaining enriched elements as showed in Figure 2. The Gaussian quadrature formula in the standard element Ω_e is

$$\int_{\Omega_e} F(r)dr \approx \sum_{i=1}^{4} G_i F(\xi_i), \tag{11}$$

where $F(r)$ is the general integrand, G_i is the coefficient and all is 0.25, ξ_i is the Gaussian quadrature point as showed in Figure 2.

For the integrand $F(r)$, the integrals of enriched function can be calculated by the linear basis function $v_j(r)$, and $v_j(r)$ at the Gaussian points is a constant as follows:

$$\nabla\psi_j(r)|_{r=\xi_i}$$

$$= \nabla v_j(r) \cdot \left(|N(r)| - |N_j|\right)$$

$$+ v_j(r) \cdot \nabla\left(|N(r)| - |N_j|\right)|_{r=\xi_i}.$$

$$N(r)|_{r=\xi_i} = \sum_{k=1}^{4} v_k(\xi_i)N_k. \tag{12}$$

$$v_j(\xi_i) = 0.5854, \quad i = j.$$

$$v_j(\xi_i) = 0.1382, \quad i \neq j.$$

Incorporating (7) and (8), and assembling all the element matrixes (8), after using the Gaussian quadrature to calculate the enriched elements of (9), the linear system equations of SP$_N$ equations is established and can be rewritten in the matrix form:

$$M_{(N_c+N_e)*(N+1)/2,(N_c+N_e)*(N+1)/2}\overline{\Phi}_{(N_c+N_e)*(N+1)/2,1}$$

$$= \overline{B}_{(N_c+N_e)*(N+1)/2,1}, \tag{13}$$

where M is a matrix including $(N_c + N_e) * (N+1)/2$ rows and $(N_c + N_e) * (N+1)/2$ columns, $\overline{\Phi}$ is the unknowns including $(N_c + N_e) * (N+1)/2$ components, and $\overline{B}$ is the source term including $(N_c + N_e) * (N+1)/2$ components. Obtain the $\overline{\Phi}$ of (13) and instead of $\varphi(r)$ in (2). The linear relationship between the surface detector readings J^+ and the source S or Q is established.

The goal of forward problem is to get the relationship between the surface detector readings and the internal sources. To have the general and particular comparison, the correlation coefficient $\text{CORR}(J^+_{X/F}, J^+_{\text{Fine}})$ and the well known mean relative numerical error $\text{MRNE}_{\text{Fine}}$ of J^+ are both defined to quantitatively evaluate the performance of XFEM

or FEM on the coarse mesh with respect to the FEM on the fine mesh:

$$\mathrm{CORR}\left(J_{X/F}^{+}, J_{\mathrm{Fine}}^{+}\right)$$

$$= \frac{\sum_{i=1}^{N_s}\left(J_{i,X/F}^{+} - \overline{J}_{i,X/F}^{+}\right)\left(J_{i,\mathrm{Fine}}^{+} - \overline{J}_{i,\mathrm{Fine}}^{+}\right)}{\sqrt{\left(\sum_{i=1}^{N_s}\left(J_{i,X/F}^{+} - \overline{J}_{i,X/F}^{+}\right)^2\right)\left(\sum_i\left(J_{i,\mathrm{Fine}}^{+} - \overline{J}_{i,\mathrm{Fine}}^{+}\right)^2\right)}}, \quad (14)$$

$$\mathrm{MRNE}_{\mathrm{Fine}} = \frac{\sum_{i=1}^{N_s}\left(J_{i,X/F}^{+} - J_{i,\mathrm{Fine}}^{+}\right)/J_{i,\mathrm{Fine}}^{+}}{N_s}, \quad (15)$$

where $J_{i,X/F}^{+}$ is the J^{+} compute by XFEM or FEM, $J_{i,\mathrm{Fine}}^{+}$ is compute by FEM on the fine mesh. N_s is the number of the sampling point, $\overline{J}_{i,X/F}^{+}$ and $\overline{J}_{i,\mathrm{Fine}}^{+}$ is mean value of $J_{i,X/F}^{+} J_{i,\mathrm{Fine}}^{+}$ in (14). $\mathrm{CORR}\left(J_{X/F}^{+}, J_{\mathrm{Fine}}^{+}\right) = 1$ illustrates the two data is identical after normalization and can be used to assess the degree of closeness between the two data. Similarly, by substituting the $J_{i,\mathrm{MC}}^{+}$ for the $J_{i,\mathrm{Fine}}^{+}$ in (14) and (15), the mean relative numerical error $\mathrm{MRNE}_{\mathrm{MC}}$ and correlation coefficient $\mathrm{CORR}\left(J_{X/F/\mathrm{MC}}^{+}, J_{\mathrm{MC}}^{+}\right)$ of XFEM and FEM with respect to MC method can be obtained for evaluation.

3. Results and Discussion

3.1. Regular Phantom Experiment Compared with Standard FEM. The validation studies were performed using a cylindrical phantom of 30 mm height and 10 mm radius to model a mouse. It consisted of ellipsoids or cylinders to represent the tissues of mouse as shown in Figure 3(a). A solid sphere source of 1 mm radius and 0.238 nW/mm³ power density was centered at (3, 5, 0) inside the right lung. The relevant optical parameters from literature [14] are listed in Table 1. Numerical simulations are carried out to compare the standard FEM and XFEM for SP$_N$ equations. The coarse mesh containing 3459 nodes without internal mesh boundary generation was used for XFEM and another coarse mesh contained 3573 nodes for FEM as shown in Figures 3(b) and 3(c). Because the precise analytic solutions for heterogamous phantom is difficult to obtained, the FEM on the fine mesh can get relative accurate solution according to the classical finite element analysis:

$$\left\|\varphi - \varphi^h\right\| \longrightarrow 0 \quad \text{as } h \longrightarrow 0, \quad (16)$$

where the numerical solution φ^h of FEM converges to the exact solution φ as the mesh size h decreases. We choose the result of the FEM on the fine mesh containing 12312 nodes as the standard for comparison. The program of FEM and XFEM is coded in MATLAB on the desktop computer (Intel(R) Xeon(R) 2 CPU E5430 @ 2.66 GHz, and 8 G RAM).

345 interpolate points whose value is nonzero are uniformly sampled around the phantom surface. Choosing the FEM on the fine mesh as the standard, the absolute value of exiting partial current $J_{X/F}^{+}$ on the sampling points are arranged in ascending order. Then the results of XFEM,

TABLE 1: Optical parameters of phantom.

Material	$\mu_a[\mathrm{mm}^{-1}]$	$\mu_s[\mathrm{mm}^{-1}]$	g
Muscle	0.01	4.0	0.9
Lung	0.35	23.0	0.94
Heart	0.2	16.0	0.85
Liver	0.002	20.0	0.9
Bone	0.035	6.0	0.9

TABLE 2: Results of XFEM and FEM for SP$_N$ approximations.

Method		FEM	XFEM
Number of nodes		3459	3573
Number of elements		15540	15987
$\mathrm{CORR}\left(J_{X/F}^{+}, J_{\mathrm{Fine}}^{+}\right)$	DA	0.85	0.94
	SP3	0.88	0.95
	SP7	0.88	0.90
$\mathrm{MRNE}_{\mathrm{Fine}}$	DA	31%	16%
	SP3	22%	14%
	SP7	22%	15%

and FEM on the comparable coarse mesh at these sampling points can be obtained by interpolation. The comparison results with Diffusion, SP$_3$, SP$_7$ approximations are showed in Figures 4(a), 4(b), and 4(c), respectively. Although the curve has fluctuation caused by the discretization error, the results have similar tendency. It is clear that the blue curve of XFEM on coarse mesh is closer to the green curve of FEM on fine mesh than that of FEM on the coarse mesh. Compared with the FEM, the $\mathrm{CORR}\left(J_{X/F}^{+}, J_{\mathrm{Fine}}^{+}\right)$ of XFEM is closer to 1 for DA, SP$_3$, and SP$_7$ with 0.94, 0.95, and 0.90. The two methods is carried out on comparative mesh even the XFEM on the slightly less coarse mesh (3459 nodes). Thus the solution using XFEM is nearer to the true solution. The $\mathrm{MRNE}_{\mathrm{Fine}}$ of XFEM is 14%, 15%, and 16%, which is much smaller than FEM, whose $\mathrm{MRNE}_{\mathrm{Fine}}$ is 31%, 22%, and 22% for DA, SP$_3$, and SP$_7$ equations respectively. The results listed in Table 2 using the two evaluation indexes demonstrate the validity of the proposed method for SP$_N$ equations. Moreover, it indicates that the XFEM is superior to the standard FEM and the solution of XFEM is more closer to the true solution for SP$_N$ equations.

3.2. Digital Mouse Experiment Compared with Monte Carlo Method. In this experiment, a digital heterogeneous mouse from CT and cryosection data consisting of several organs is adopted to evaluate the performance of the XFEM. The mouse is shown in Figure 5(a) [15]. A cylinder source of 0.8 mm radius, 1.6 mm height, and 0.311 nW/mm³ power density was centered at (12, 7.5, 50.5) inside the liver. The optical parameters of the organs at 670 nm wavelength computed by the literature [14] are employed and shown in Table 3. The experiments were conducted using XFEM on the coarse mesh, FEM on the fine mesh, and classical MC method. The surface detector readings using MC method was achieved from MOSE [16] using 5 million photons. Since there are still errors between the SP$_N$ equation and MC to

(a)

(b)

(c)
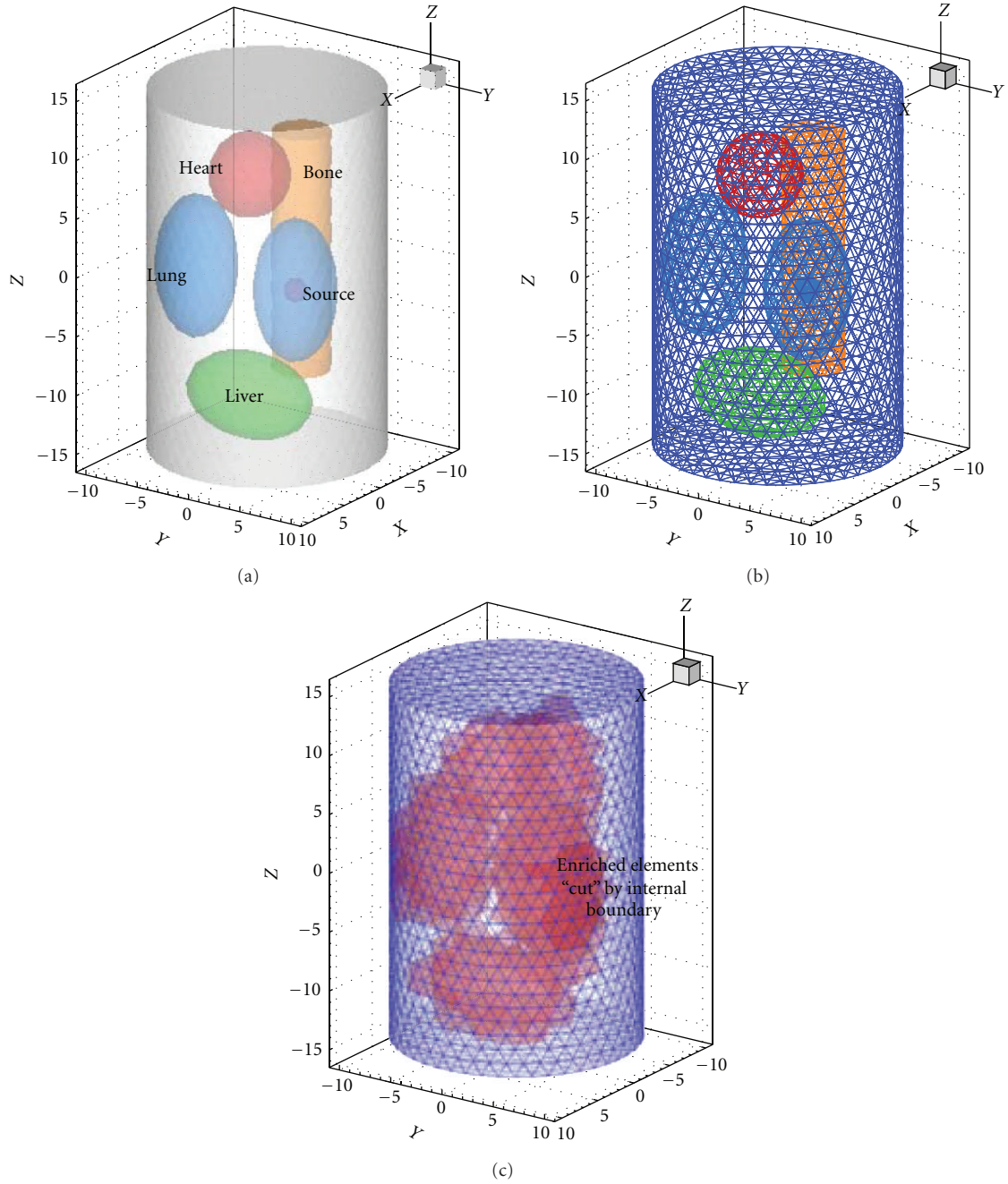
FIGURE 3: (a) Heterogeneous phantom, (b) the element mesh of FEM, and (c) element mesh of XFEM without internal boundary generation, red region is the enriched region.

depict the light propagation, the XFEM is compared with the FEM and MC simultaneously.

As shown in Figure 4(d), it is clear that the blue and green curves agree well. Considering the difference of the results between SP_3 and SP_7 equations are very small, for simplicity, only the SP_3 results are presented in Figure 6. The absolute values on the surface using the FEM, XFEM and MC methods agree well and have the similar tendency in general. For detailed comparison, 521 points with nonzero values on the mouse surface are sampled, and then the surface value of these points using the three methods are obtained by interpolation and shown in Figure 7. Choosing the MC method

as standard, the mean relative numerical error $MRNE_{MC}$ and correlation coefficient $CORR(J_{X/F/MC}^+, J_{MC}^+)$ of XFEM and FEM with respect to MC method are obtained and shown in Table 4. The $MRNE_{MC}$ and $CORR(J_{X/F/MC}^+, J_{MC}^+)$ of FEM is 0.86 and 44%, while those of XFEM is 0.93 and 45%. The computational time of several modules that perform specific computational tasks, mesh generation, matrix assembly and solver. Since the XFEM does not have the complex mesh generation as FEM, the total time cost of matrix assembly and solver is considered for comparison. It is clear that the time cost of XFEM is only 367 seconds on the coarser mesh
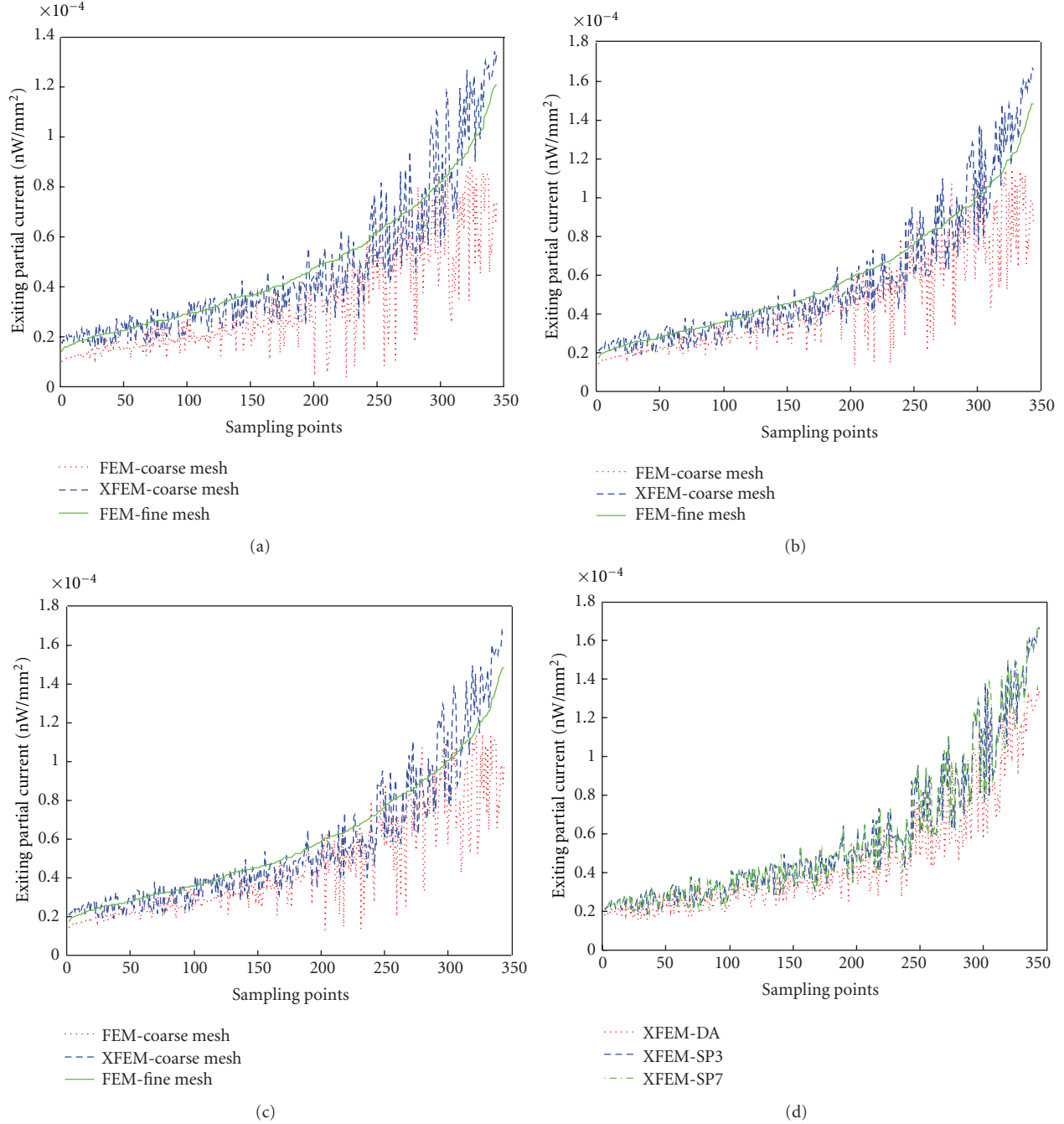
FIGURE 4: Surface detector readings J^+ for SP_N equations around the regular heterogeneous phantom using the two methods on the coarse mesh compared with the FEM on the fine mesh. (a) DA equations. (b) SP_3 equations. (c) SP_7 equations. (d) Comparison among the results of XFEM for DA, SP_3, and SP_7 equations.

while that of FEM is 2675 seconds with the similar accuracy results. All the time cost of two methods are far less than that of MC method. The XFEM has a distinct strength on time-efficiency and this makes it more practical in imaging process.

4. Conclusion

We have derived the extended finite element method with SP_N approximations for the forward model of the three dimensional optical imaging. Considering the complex geometric object, it is necessary to have the fined mesh to conform to the internal boundary. And the mesh conformation is a difficult issue in the pretreatment in the FEM, moreover the standard FEM on the fine mesh for SP_N approximations cost too much especially for the high order approximation. Fortunately, the XFEM can deal with this problem. The XFEM includes the standard FEM as a special case, which does not require a geometric representation of the interface or any boundary mesh generation. Use the signed distance
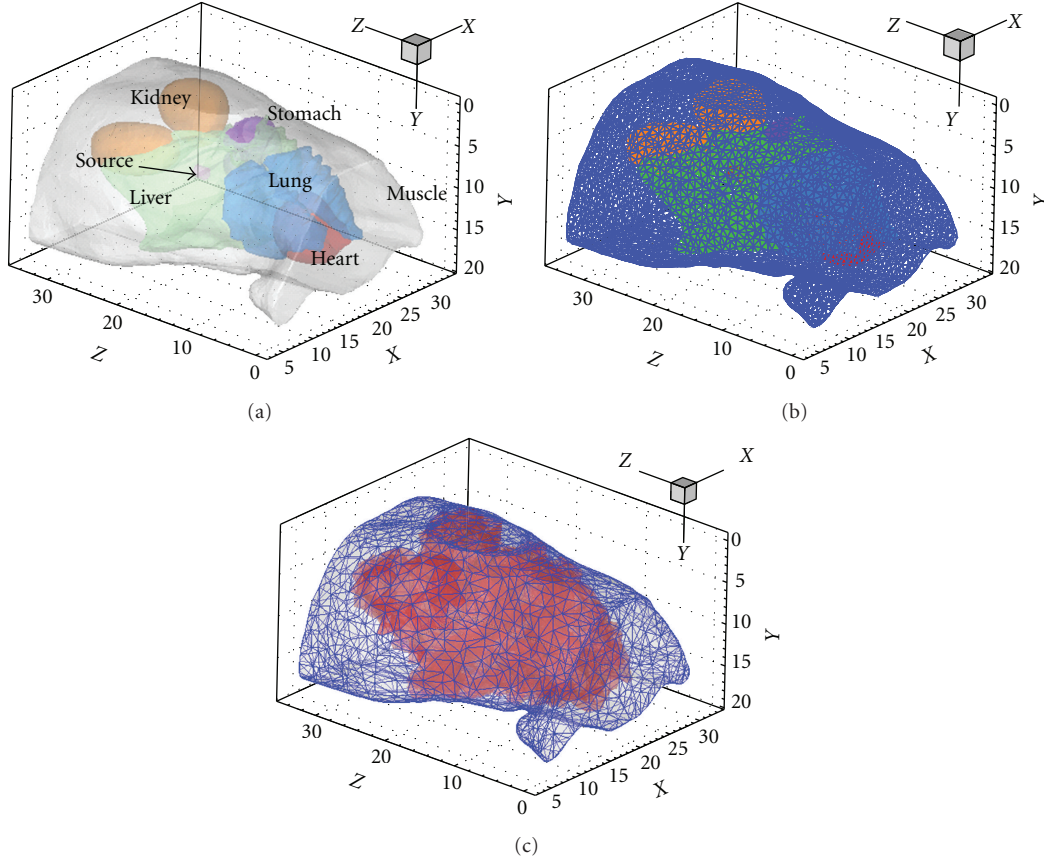
FIGURE 5: (a) The digital mouse model. (b) The external and internal boundary mesh for FEM. (c) The external boundary mesh and enriched region for XFEM.

TABLE 3: Optical parameters of the mouse organs.

Organs	$\mu_a[\text{mm}^{-1}]$	$\mu_s[\text{mm}^{-1}]$	g
Muscle	0.08697	4.29071	0.90
Heart	0.05881	6.42581	0.85
Stomach	0.01139	17.96150	0.92
Liver	0.35182	6.28066	0.90
Kidney	0.06597	16.09293	0.86
Lung	0.19639	36.23141	0.94

TABLE 4: Comparison among the FEM, XFEM and MC method.

Method	FEM	XFEM	MC method
Number of nodes	24906	6541	/
Number of elements	132202	32398	/
$\text{CORR}(J_{X/F/\text{MC}}^+, J_{\text{MC}}^+)$	0.86	0.92	1
MRNE_{MC}	44%	45%	0
Time cost [s]	2675	367	7292

functions add to the standard basis functions, the number of nodes can keep unchanged. The interface can be well depicted by the enriched functions, also the solution of SP$_N$ equation is more accurate.

The XFEM is validated through the numerical experiments. Phantom experiment was conducted and its results agreed well with the well known classical FEM. Moreover, compared with the FEM, XFEM can get more accurate result even on the slightly coarse mesh, which is closer to the FEM on the fine mesh. Digital mouse experiments further indicate that XFEM is superior to the standard FEM. The MC method was employed to evaluate the performance the proposed method. Although the relative errors and correlation coefficient using the XFEM with respect to MC method is comparable to that of standard FEM, the time

cost of XFEM is greatly decreased due to the adoption of coarser element mesh. All these indicate that the XFEM is more suitable for SP$_N$ equations especially in complex heterogeneous tissues.

Adaptive FEM method is often adopted for its high efficiency [17, 18]. But nearly all the adaptive method is based on the mesh refinement which is extremely complicate especially in three dimensions. XFEM can also be seen as an adaptive FEM method because it increases the degrees of freedom near the internal boundary while using fixed mesh, therefore the mesh refinement can be avoided. Thus the process in each adaptive level is simplified and the calculation cost is decreased. Moreover, the discretization error of the forward model is improved, which is important for the quantification reconstruction. XFEM may provide a potential tool for reconstruction algorithm.

(a)

(b)

(c)
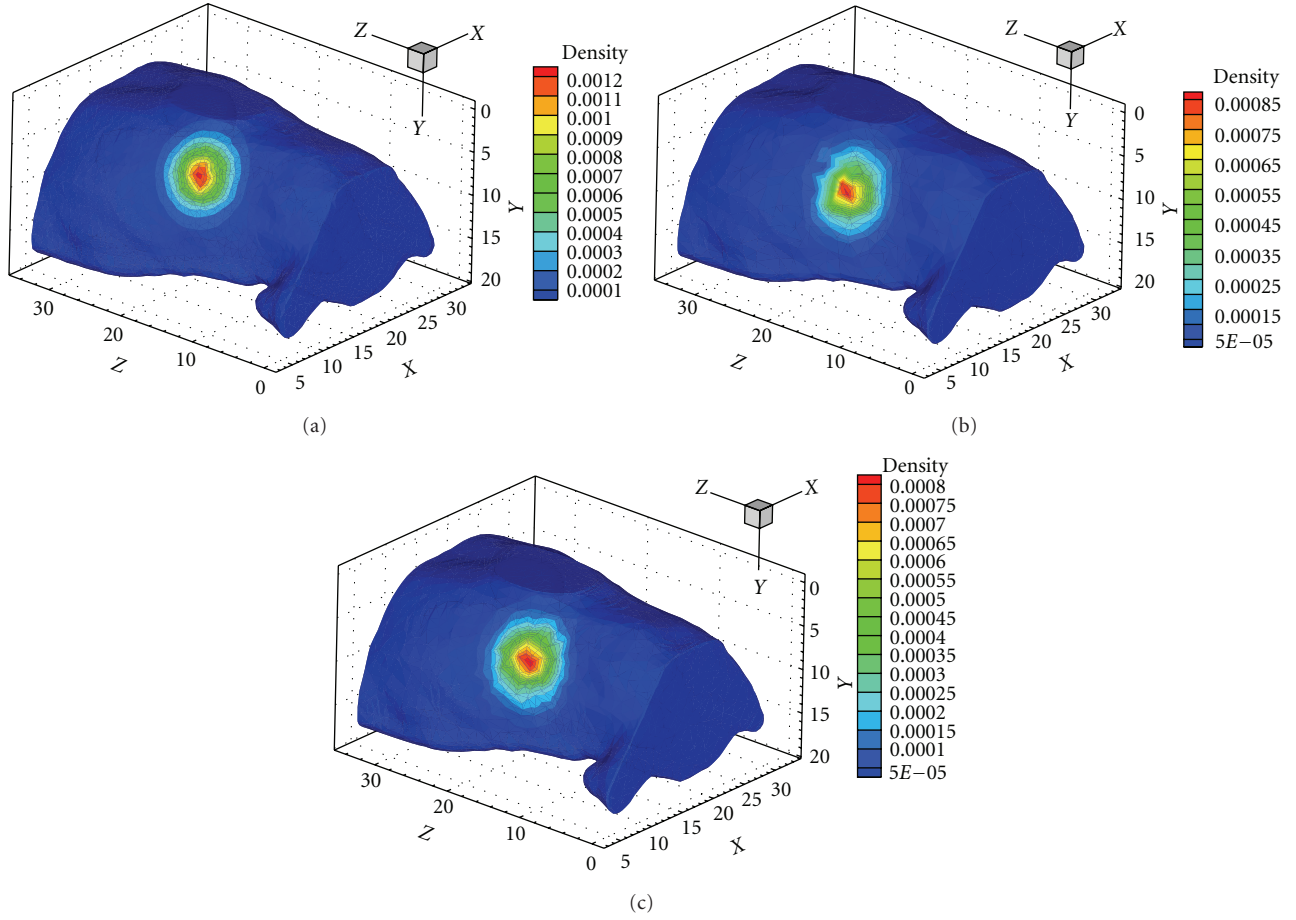
FIGURE 6: The surface detector readings with SP₃ approximation (a) using FEM on the fine mesh, (b) using XFEM on the coarse mesh, and (c) using MC method.
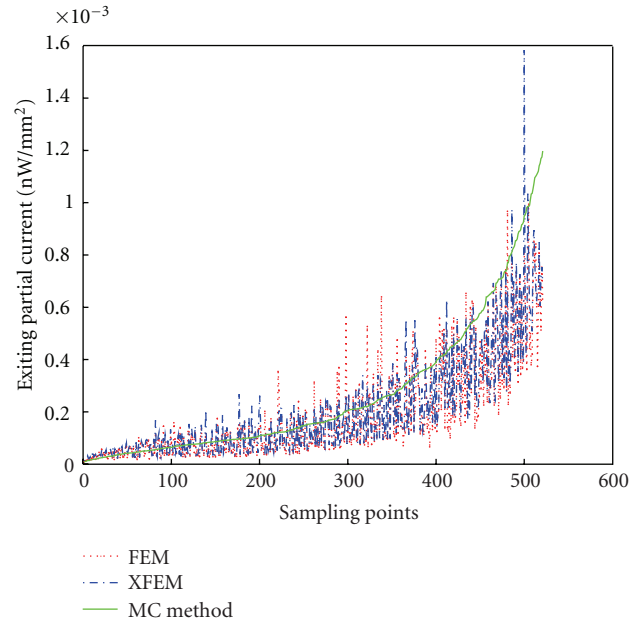
FIGURE 7: Surface detector readings J^+ using MC method, XFEM, and FEM at the sampling points. Green line denotes the results of MC method, red and blue lines are that of FEM and XFEM with SP₃ approximation.

References

[1] V. Ntziachristos, J. Ripoll, L. V. Wang, and R. Weissleder, "Looking and listening to light: the evolution of whole-body photonic imaging," *Nature Biotechnology*, vol. 23, no. 3, pp. 313–320, 2005.

[2] M. Guven, B. Yazici, X. Intes, and B. Chance, "Diffuse optical tomography with a priori anatomical information," *Physics in Medicine and Biology*, vol. 50, no. 12, pp. 2837–2858, 2005.

[3] A. P. Gibson, J. C. Hebden, and S. R. Arridge, "Recent advances in diffuse optical imaging," *Physics in Medicine and Biology*, vol. 50, no. 4, pp. R1–R43, 2005.

[4] W. Cong, G. Wang, D. Kumar et al., "Practical reconstruction method for bioluminescence tomography," *Optics Express*, vol. 13, no. 18, pp. 6756–6771, 2005.

[5] A. E. Spinelli, C. Kuo, B. W. Rice et al., "Multispectral Cerenkov luminescence tomography for small animal optical imaging," *Optics Express*, vol. 19, no. 13, pp. 12605–12618, 2011.

[6] A. D. Klose and E. W. Larsen, "Light transport in biological tissue based on the simplified spherical harmonics equations," *Journal of Computational Physics*, vol. 220, no. 1, pp. 441–470, 2006.

[7] M. Chu, K. Vishwanath, A. D. Klose, and H. Dehghani, "Light transport in biological tissue using three-dimensional frequency-domain simplified spherical harmonics equations," *Physics in Medicine and Biology*, vol. 54, no. 8, pp. 2493–2509, 2009.

[8] A. D. Klose, V. Ntziachristos, and A. H. Hielscher, "The inverse source problem based on the radiative transfer equation in optical molecular imaging," *Journal of Computational Physics*, vol. 202, no. 1, pp. 323–345, 2005.

[9] M. Schweiger, S. R. Arridge, M. Hiraoka, and D. T. Delpy, "The finite element method for the propagation of light in scattering media: boundary and source conditions," *Medical Physics*, vol. 22, no. 11 I, pp. 1779–1792, 1995.

[10] H. Jiang, "Frequency-domain fluorescent diffusion tomography: a finite-element-based algorithm and simulations," *Applied Optics*, vol. 37, no. 22, pp. 5337–5343, 1998.

[11] N. Moës, J. Dolbow, and T. Belytschko, "A finite element method for crack growth without remeshing," *International Journal for Numerical Methods in Engineering*, vol. 46, no. 1, pp. 131–150, 1999.

[12] J. Chessa, P. Smolinski, and T. Belytschko, "The extended finite element method (XFEM) for solidification problems," *International Journal for Numerical Methods in Engineering*, vol. 53, no. 8, pp. 1959–1977, 2002.

[13] Y. Lu, B. Zhu, H. Shen, J. C. Rasmussen, G. Wang, and E. M. Sevick-Muraca, "A parallel adaptive finite element simplified spherical harmonics approximation solver for frequency domain fluorescence molecular imaging," *Physics in Medicine and Biology*, vol. 55, no. 16, pp. 4625–4645, 2010.

[14] G. Alexandrakis, F. R. Rannou, and A. F. Chatziioannou, "Tomographic bioluminescence imaging by use of a combined optical-PET (OPET) system: a computer simulation feasibility study," *Physics in Medicine and Biology*, vol. 50, no. 17, pp. 4225–4241, 2005.

[15] B. Dogdas, D. Stout, A. F. Chatziioannou, and R. M. Leahy, "Digimouse: a 3D whole body mouse atlas from CT and cryosection data," *Physics in Medicine and Biology*, vol. 52, no. 3, pp. 577–587, 2007.

[16] MOSE, http://www.mosetm.net/.

[17] Y. Lv, J. Tian, W. Cong et al., "A multilevel adaptive finite element algorithm for bioluminescence tomography," *Optics Express*, vol. 14, no. 18, pp. 8211–8223, 2006.

[18] A. Joshi, W. Bangerth, and E. M. Sevick-Muraca, "Adaptive finite element based tomography for fluorescence optical imaging in tissue," *Optics Express*, vol. 12, no. 22, pp. 5402–5417, 2004.

Research Article

A Workflow for Patient-Individualized Virtual Angiogram Generation Based on CFD Simulation

Jürgen Endres,[1] Markus Kowarschik,[2] Thomas Redel,[2] Puneet Sharma,[3] Viorel Mihalef,[3] Joachim Hornegger,[1,4] and Arnd Dörfler[5]

[1] *Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander University of Erlangen-Nuremberg, Martensstrasse 3, 91058 Erlangen, Germany*

[2] *Angiography & Interventional X-Ray Systems, Healthcare Sector, Siemens AG, Siemensstrasse 1, 91301 Forchheim, Germany*

[3] *Corporate Research and Technology, Siemens Corporation, 755 College Road East, Princeton, NJ 08540, USA*

[4] *Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander University of Erlangen-Nuremberg, 91052 Erlangen, Germany*

[5] *Department of Neuroradiology, Friedrich-Alexander University of Erlangen-Nuremberg, Schwabachanlage 6, 91054 Erlangen, Germany*

Correspondence should be addressed to Jürgen Endres, juergen.endres@cs.fau.de

Received 1 June 2012; Revised 14 August 2012; Accepted 31 August 2012

Academic Editor: Huafeng Liu

Increasing interest is drawn on hemodynamic parameters for classifying the risk of rupture as well as treatment planning of cerebral aneurysms. A proposed method to obtain quantities such as wall shear stress, pressure, and blood flow velocity is to numerically simulate the blood flow using computational fluid dynamics (CFD) methods. For the validation of those calculated quantities, virtually generated angiograms, based on the CFD results, are increasingly used for a subsequent comparison with real, acquired angiograms. For the generation of virtual angiograms, several patient-specific parameters have to be incorporated to obtain virtual angiograms which match the acquired angiograms as best as possible. For this purpose, a workflow is presented and demonstrated involving multiple phantom and patient cases.

1. Introduction

Cerebrovascular diseases are beneath cardiovascular diseases the leading cause of death among industrialized countries [1]. One clinical pathology concerning the cerebrovascular system is intracranial aneurysms, abnormal bulges within the vasculature. According to the study in [2], the prevalence of unruptured intracranial aneurysms in the general population is estimated to be up to 5%. Aneurysms threaten the patients' health in case of rupture, which will lead to a subarachnoid hemorrhage (SAH) and hence may cause a hemorrhagic stroke with severe clinical consequences. For the case of intracranial aneurysms, 30% of all patients will die within the next 30 days, 30% will develop disabilities, and only the remaining part will almost completely recover

[3, 4]. However, most of the aneurysms will never rupture. For example, out of more than 10–12 million people in the US which are estimated to have an intracranial aneurysm, about 27,000 cases per year will suffer from subarachnoid hemorrhage caused by rupture events [3].

In the management of unruptured intracranial aneurysms, different preventive treatment options are established. In a neurosurgical procedure, a metal clip is placed during an open surgery along the neck of the aneurysm to prevent blood from flowing into the aneurysm dome and hence disable the possibility of rupture [5]. In an endovascular treatment, small coils are placed within the aneurysm dome. The intention of those coils is to reduce the blood flow inside the aneurysm, leading to thrombosis and finally to an occlusion of the aneurysm. A recent interventional approach

is based on the placement of flow diverting devices within the parent artery, which also aims at reducing blood flow inside the aneurysm [5].

For endovascular treatment, X-ray angiography [6] is mandatory to visualize the aneurysm as well as parent vasculature. By intraarterially injecting contrast agent, vessel structures can be visualized in addition to catheter devices. Modern systems, where source and detector are mounted at both ends of a movable, C-shaped fixture (C-arm), are capable of acquiring 2D digital subtraction angiography (DSA) images at high frame rates, which allows to observe the distribution of injected contrast agent over time. Additional plane DSA sequences unveil a lot of flow dynamic information about the hemodynamic behaviour. In addition, by rotating the C-arm around the object, static volumes can be reconstructed in a CT-like fashion (3D rotational angiography, 3D RA) [7].

However, since all treatment options imply risks for the patients, reliable parameters for aneurysm risk classification, treatment planning, and assessment are needed. Besides geometric properties of the aneurysm itself, an increasing interest is shown for hemodynamic parameters such as pressure, wall shear stress, and blood flow velocity. Due to insufficient methods of measuring those quantities *in vivo*, computational methods—that is, numerical simulations—are investigated in order to obtain those quantities, as, for example, given in [8, 9].

However, a reliable validation of the simulation results is required prior to applications in clinical environments, for which Ford et al. [10] suggested the generation of virtual/synthetic angiograms based on CFD simulation results, and a succeeding comparison of virtual and the corresponding real angiograms. One major aspect concerning this validation method for CFD simulation results is the definition of patient-specific boundary conditions. Since those patient-specific parameters are generally not available for acquired 2D DSA sequences, *in vitro* studies based on cerebral aneurysm phantoms have been performed, where parameters such as blood flow velocities at vessels proximal to the aneurysm are known [11, 12].

Furthermore, in certain DSA acquisitions, the injection of the contrast agent is done manually which leads to variations in the injection profile as well as in the timing with respect to the patient's heart phase. Using standardized injection profiles for virtual angiography will then lead to deviations of the virtual angiogram when compared to the real one.

Beneath the aspect of validation, further studies have been published using the virtual angiography technique; for example, for visualizing outcomes of virtual treatment techniques [13, 14] or for evaluating the outcome of CFD simulation results [15–17]. Clinical applications may benefit from virtual angiograms in a way that these image sequences can be generated without the use of applying further X-ray radiation dose and injecting additional contrast agent, and they can be generated for arbitrary angulations, independent of mechanical limitations such as unreachable C-arm angulations. As a visionary future aspect—not taking into

consideration the need for a validation of CFD approaches—virtual angiograms may eventually completely replace real angiograms, such that only a 3D RA is acquired for diagnostics, whereas all dynamic information is entirely based on CFD simulation and virtual angiogram generation. Finally, virtual angiography represents a familiar way for illustrating CFD simulation results, which would be hard to interpret otherwise.

In this paper, we extend the methods proposed in [10] by further incorporating patient- and treatment-specific parameters to obtain virtual angiograms aimed at matching the corresponding real angiograms as accurately as possible. Our basic workflow has already been published in [18]. The contributions of this extended work are an additional synchronization of the heart state at the beginning of virtual and acquired DSA sequences as well as a more detailed examination of the accuracy of the results with respect to quantitative error measurements.

Our paper is structured as follows. In Section 2, the basics of the underlying CFD computations are presented, the mathematical model of virtual contrast agent (contrast medium) injection and propagation is demonstrated, and methods for extracting several patient-individual parameters are covered. Furthermore, Section 2 details the generation of the virtual angiograms as well as our approaches towards the quantitative comparison of virtual and real angiograms. These methods are embedded in a workflow which is then applied to both phantom and patient cases. In Section 3, results are presented and discussed. We finally draw our conclusions in Section 4.

2. Materials and Methods

Figure 1 provides a schematic overview of the distinct steps for generating virtual angiograms. For these methods, two different types of imaging data are used. On one hand, a volumetric 3D RA image provides geometric information for both the CFD simulation and the virtual angiography. On the other hand, a 2D DSA sequence—ideally acquired at a high frame rate (e.g., 30 frames per second or higher)—will serve as input data for a patient-specific parameter extraction and, afterwards, as ground truth for comparison. As a result, virtual (synthetic) 2D DSA sequences from arbitrary viewing directions are generated.

First, patient-specific information concerning the heart rate and heart state will be extracted from the 2D DSA sequence. This information will then be used for adapting CFD simulation parameters. Second, as additional patient-specific information, the contrast bolus injection profile is extracted from the acquired angiogram. An individualized virtual angiogram is subsequently created based on the CFD output. Finally, the resulting virtual angiogram is compared both qualitatively and quantitatively with the real angiogram.

This approach represents an essential step towards the validation of the CFD results. If the virtual angiogram matches the real angiogram closely, the user may become confident of the application of CFD methods and hence generate and evaluate further virtual angiograms from
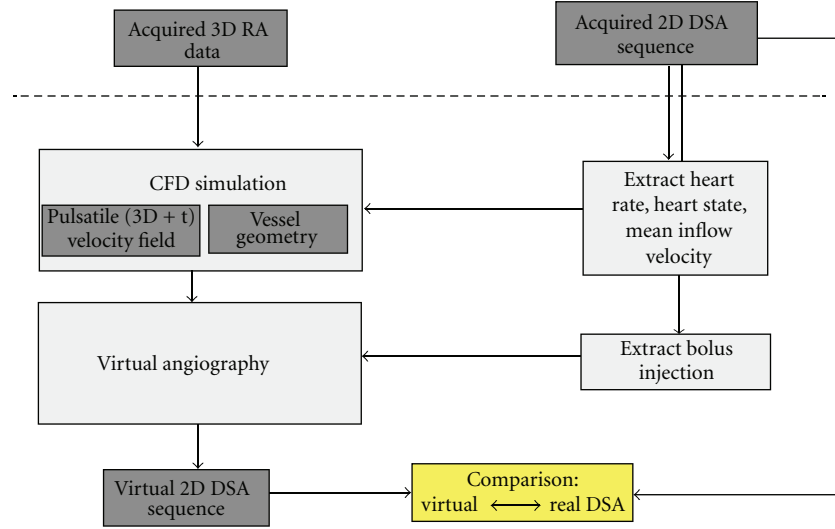
FIGURE 1: Virtual angiography workflow.
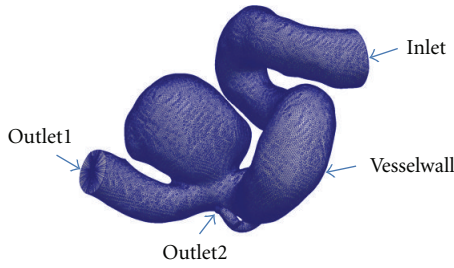
(a) Vessel geometry represented as surface mesh
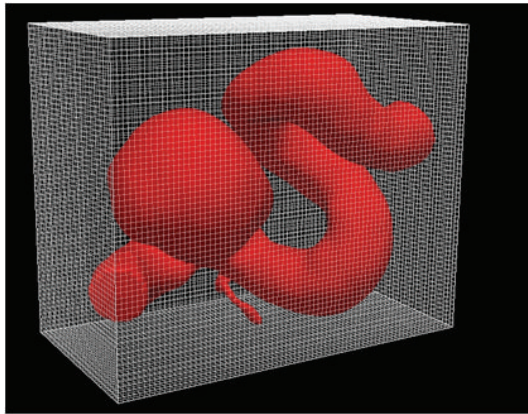
(b) Vessel geometry embedded in a Cartesian grid by using a level set

FIGURE 2: Vessel geometry for CFD simulation.

additional viewing directions without applying additional X-ray dose to the patient and without injecting further contrast medium. This means that, eventually, the computation of further virtual angiograms might replace the acquisition of further real angiograms. Note that, using virtual angiography, even viewing directions are possible that cannot be

reached by the C-arm due to mechanical limitations (e.g., due to patient/table collision).

2.1. CFD Simulation—Hemodynamic Simulation of Cerebral Blood Flow. For the computation of the flow in the cerebral vessels, the blood is modeled as a Newtonian fluid with prespecified density (ρ = 1050 kg/m^3) and viscosity (μ = 0.004 Pa · s). The basic principles of conservation of mass and momentum are applied by numerically solving the Navier-Stokes equations under appropriate boundary conditions. Under our simulation framework, the complex vessel geometry, as shown in **Figure** 2(a), which is provided as a surface mesh, is embedded in a Cartesian grid by using a level set, compare **Figure** 2(b) [19]. This provides an automatic domain setup and allows the user to bypass the time-consuming step of mesh generation [20].

After computing a level set φ with positive values inside the vessel, we solve the Navier-Stokes equations

$$\rho(\varphi)\left(\frac{\partial u}{\partial t} + u \cdot \nabla u\right) = -\nabla p + \mu(\varphi)\Delta u + F,$$

$$\nabla \cdot u = 0,$$

$$\rho(\varphi) = \rho^1(\varphi)H(\varphi) + \rho^2(\varphi)(1 - H(\varphi)),$$

$$\mu(\varphi) = \mu^1(\varphi)H(\varphi) + \mu^2(\varphi)(1 - H(\varphi)),$$

$$H(\varphi) = \begin{cases} 1, & \varphi > 0, \\ 0, & \varphi < 0. \end{cases} \tag{1}$$

The Heaviside function H distinguishes sharply between the solid and the fluid components of the domain, while we use second-order accurate spatial extrapolation across the boundary when imposing boundary conditions. The equations are discretized and solved iteratively for velocity and pressure. We use a fractional step method [21] that computes in a first step an intermediate velocity field using
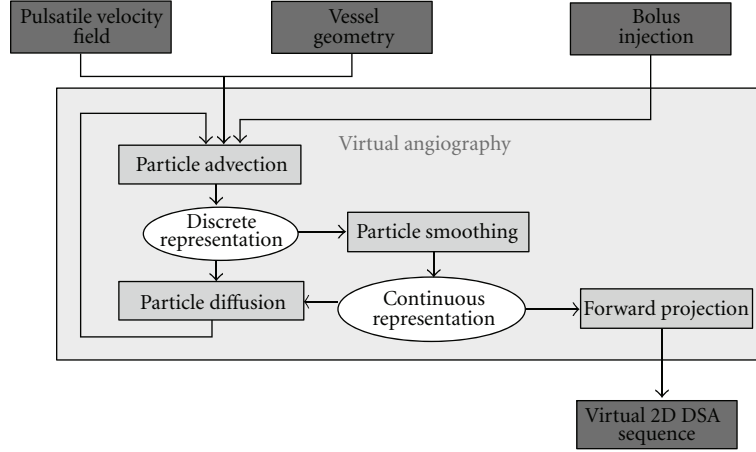
Figure 3: Virtual angiography algorithm.

(a) 1D illustration of a mixture of Gaussian distributions, red lines denote the locations of the discrete particles

(b) Particles and a rendered image of the corresponding contrast agent concentration field
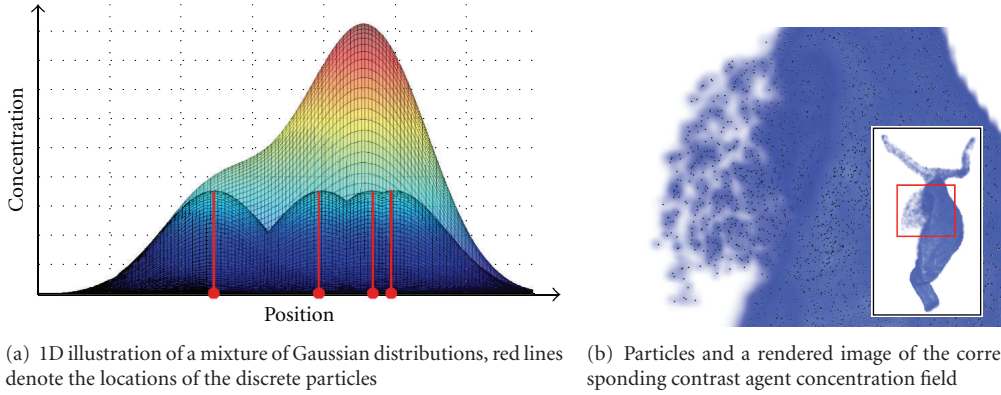
Figure 4: Particle smoothing for the sake of reprojection and the determination of contrast agent gradients.

the nonlinear advection-diffusion equation for velocity and then projects the intermediate velocity onto the field of divergence-free and tangent to the vessel boundary vector fields.

For the velocity advection, we use a second-order upwind Van-Leer slope limiting method, while for the diffusion force components, we use a semi-implicit approach as in [22] which is second-order accurate in space and unconditionally stable in 3D. The pressure Poisson equation (PPE) is solved using a multigrid preconditioned conjugate gradient solver. After the PPE is solved and the updated pressure field is determined, the fluid domain velocity is updated by subtracting the pressure gradient. The body force field F in (1) can be used to include forces due to flow diverter embedded geometries, as we proposed in [23].

For applying the boundary conditions, the inlet is completely embedded inside the Cartesian grid, and Dirichlet boundary conditions for velocity are enforced using linear extrapolation from the interior of the domain using an extrapolation routine adapted from [24]. A time-varying velocity field is applied at the inlet, which is modeled spatially as a plug profile. The outlets are modeled with constant pressure boundary conditions. The computations are performed using time steps constrained by the CFL

condition [25], while the spatial resolution was in the range of $5 \cdot 10^5$ cells, chosen such that the velocity differs less than 1% when compared to the refined grid.

2.2. Virtual Angiography—Simulated Transport of Contrast Agent and Its Visualization. In our approach, contrast agent passing through the vascular territory under consideration is modeled as a set of n discrete particles

$$\Omega = \left\{ \boldsymbol{\rho}_i \right\}, \quad \boldsymbol{\rho}_i \in \mathbb{R}^3, \ i \in 1, \dots, n, \ n \in \mathbb{N}. \tag{2}$$

The particles are assumed to be both mass- and dimensionless; hence, there is no interaction between particles (e.g., there are no (in)elastic collisions). Each particle $\boldsymbol{\rho}_i$ is defined by its location in $\mathbb{R}^3$ and is freely movable within space which means that its position is not restricted to grid points. Note that other approaches towards the generation of virtual angiograms are based on the numerical solution of an advection-diffusion equation in order to simulate the transport of contrast agent, see [10], for example. Our particle-based method can be seen as a straightforward alternative to a scheme that explicitly models the physics of contrast medium propagation using a partial differential equation [26]. By using this discrete scheme, additional

(a) Vessel geometry used for the CFD simulation

(b) User-defined line of interest (LOI) in acquired DSA series

(c) Measured intensities along LOI

(d) Zoomed view of (c)
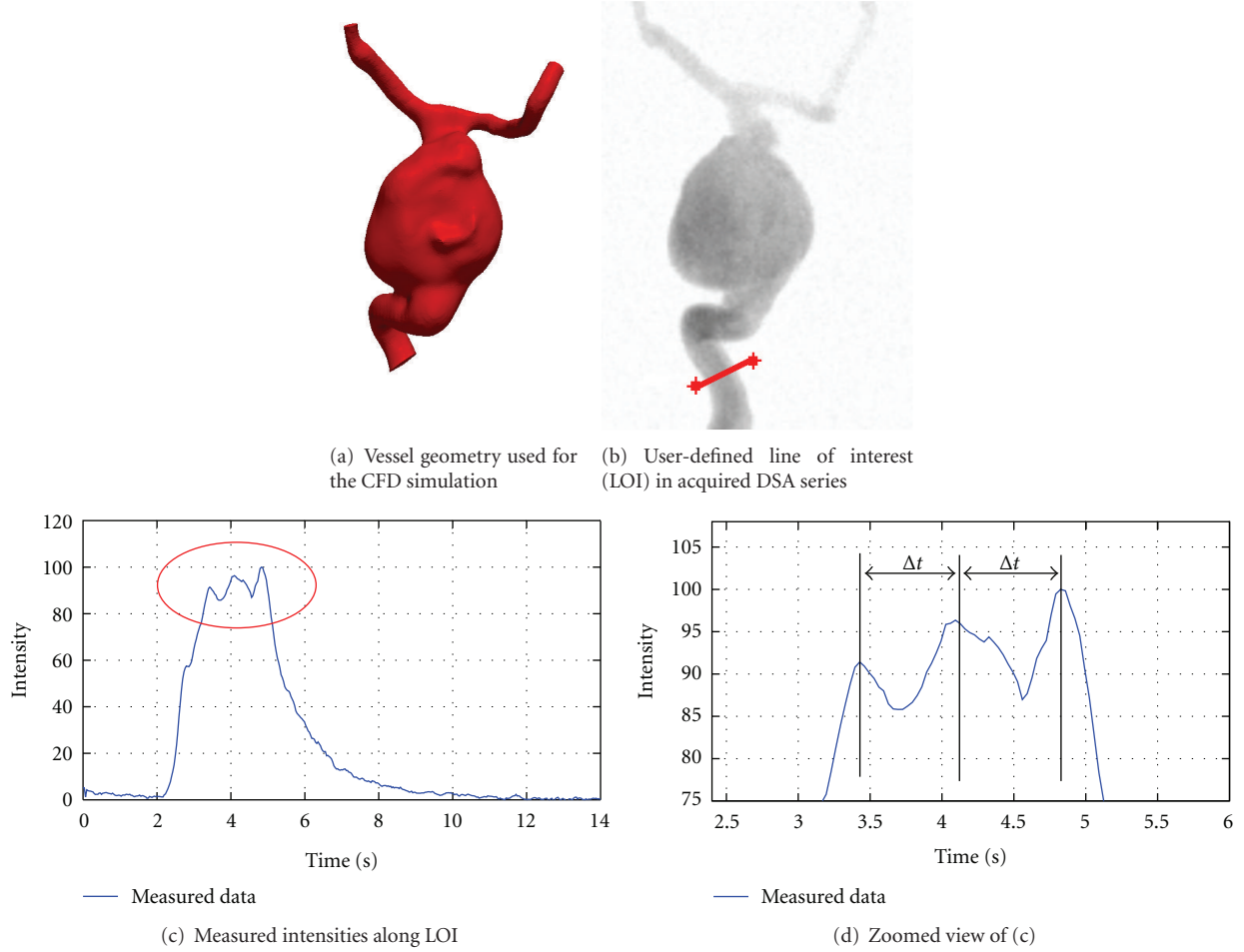
FIGURE 5: Extraction of heart rate information. The location of the line of interest (LOI), (b), corresponds to the inlet of the geometric model for the CFD simulation (a). Based on the measured intensities (c), the time between subsequent intensity peaks (d) is used for estimating the patient's heart rate.

analysis based on the particle representation can be included for flow quantification; for example, particle residence times [27] or further visualization techniques such as streamlines, streaklines, or pathlines may be employed.

Two distinct physical processes are involved in the transport of contrast medium through the vasculature. On one hand, an advective process propagates contrast agent based on an underlying velocity field, which is generated by the CFD solver. On the other hand, a diffusive process causes the contrast agent to mix autonomously with blood, which leads to a homogenization of both substances. Figure 3 illustrates the algorithm for performing the simulated transport of contrast agent consisting of advection, diffusion, and an additional smoothing procedure, which is used to transform the discrete particle set Ω into a corresponding continuous representation. In each time step, the particle set is processed sequentially. First, advection is applied to each particle. The resulting particle set is then transformed into a continuous representation (particle smoothing), from which a concentration gradient field is then obtained. Finally, the particle set is processed again according to the calculated gradient (particle diffusion).

Advection. Taking into consideration only the advective part of the transport process, the trajectory of a single particle $\boldsymbol{\rho}_i$ can be characterized independently from all other particles. This trajectory can be described as the solution of the ordinary differential equation (ODE)

$$\dot{\boldsymbol{\rho}}_i(t) = f\left(t, \boldsymbol{\rho}_i(t)\right), \tag{3}$$

where $\boldsymbol{\rho}_i$ denotes the spatial position of the particle and t represents the time. For a unique solution, an initial value

$$\boldsymbol{\rho}_i(t_i = 0) = \boldsymbol{\rho}_{i,0} \tag{4}$$

has to be specified. This value corresponds to the point in space and time where the particle gets injected into the vasculature.

In (3), $f : \mathbb{R} \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ denotes the function representing the time- and space-dependent velocity field. The function f itself is unknown; only the function values—representing the velocities at the nodes of the computational grid—are computed by the underlying CFD solver. Hence, this equation is not solvable analytically and, consequently, a
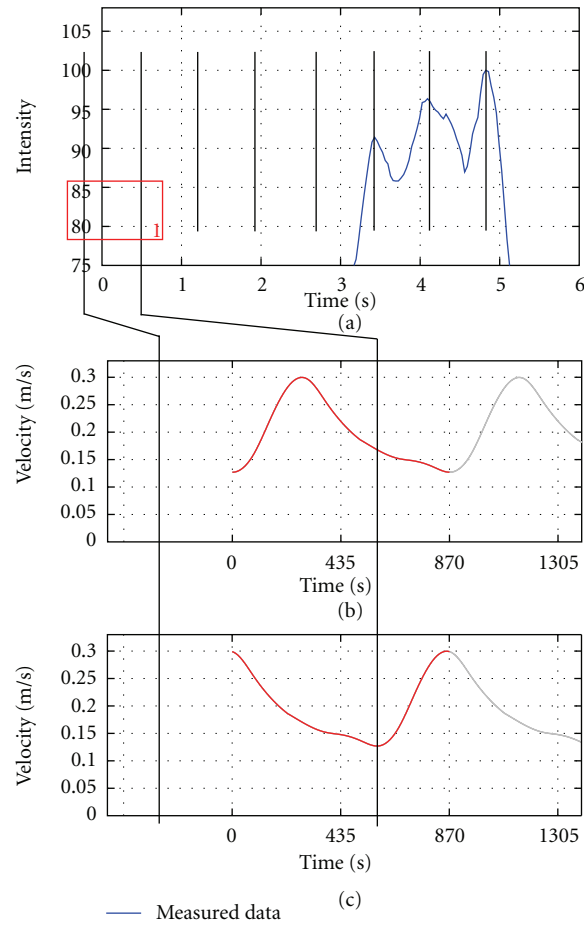
FIGURE 6: Heart-state synchronization. In (a), black vertical lines denote the extrapolated intensity peaks. In (b), the standard inflow velocity profile (red) before adaptation is shown in combination with the intensity peaks. Assuming a correspondence of low blood velocities and high opacification, the inflow velocity profile does not match to the extrapolated intensity peaks. In (c), the adapted (shifted) inflow velocity profile is shown which is then used for the CFD simulation.

(a) Example of fitted capacitor curve

(b) Description of capacitor curve parameters
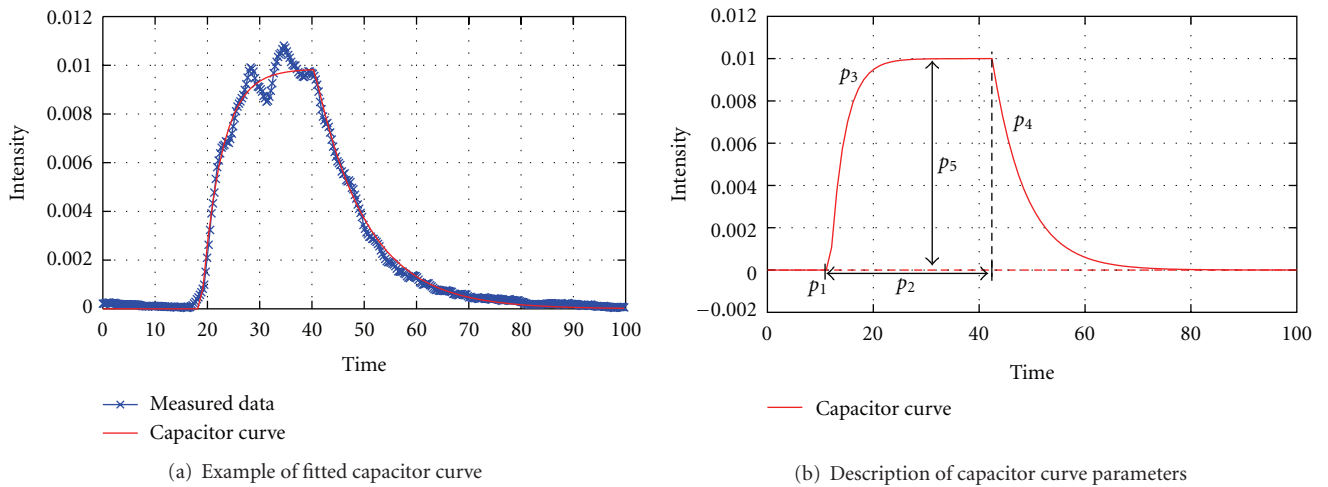
FIGURE 7: Capacitor function.
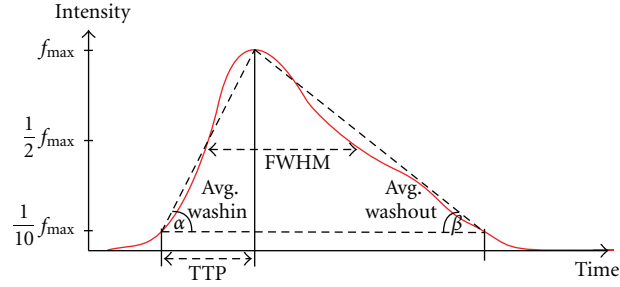
FIGURE 8: Illustration of quantitative features.

(a) $0°, 0°$

(b) $90°, 0°$

(c) $45°, 0°$

(d) $90°, 135°$

(e) $0°, 45°$

(f) $90°, 45°$
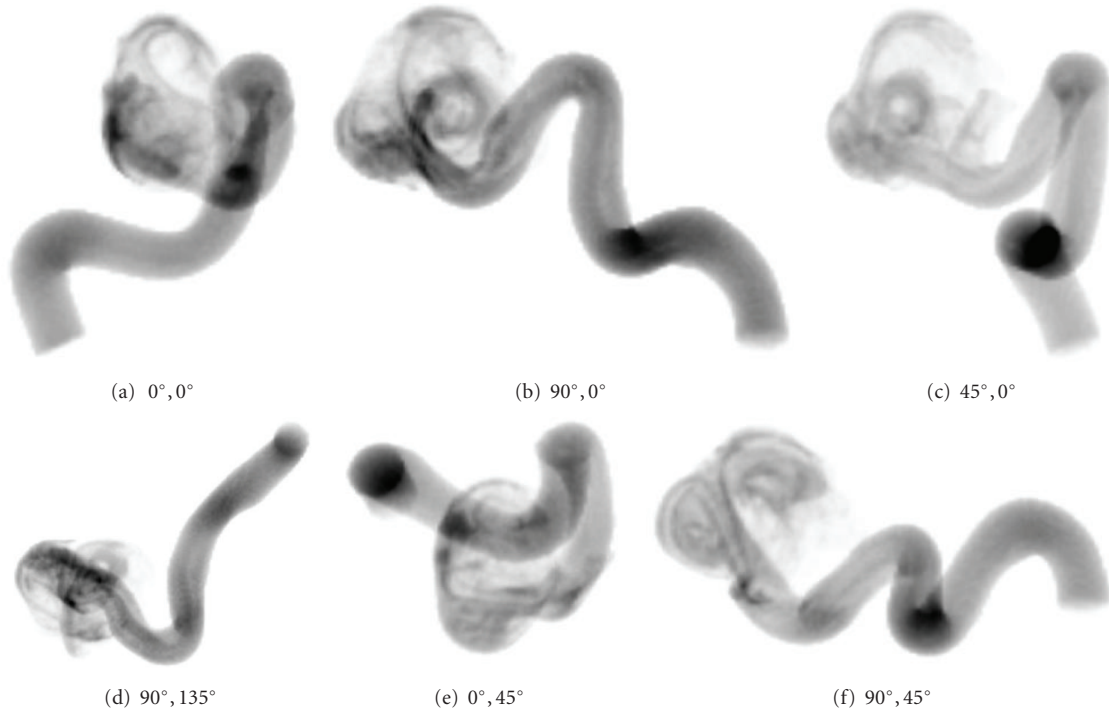
FIGURE 9: Virtual angiograms based on different projection angles for the same simulated angiography. All images show the contrast agent distribution at the same point in time. The simulated position of the C-arm is denoted below each image, given as a *primary angle*, *secondary angle* pair, where *primary angle* denotes the rotation about the left/right axis and *secondary angle* about the head/feet axis.

(a) ROIs for phantom data set, C-arm angulation $(-0.4°, -0.3°)$

(b) Extracted inflow profile for phantom data set
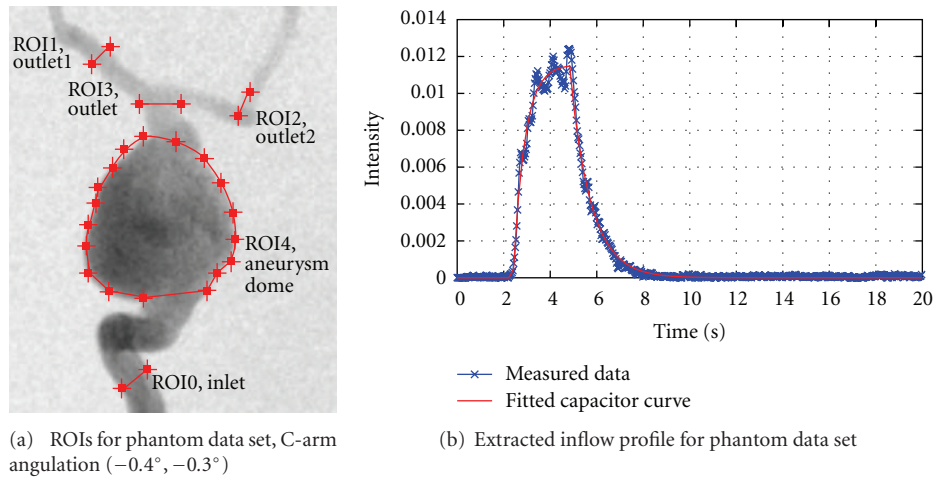
FIGURE 10: ROIs of phantom data set and used inflow profile.

(a) $t = 2.60\,\text{s}$ (b) $t = 2.66\,\text{s}$ (c) $t = 2.76\,\text{s}$ (d) $t = 2.90\,\text{s}$ (e) $t = 3.20\,\text{s}$ (f) $t = 3.36\,\text{s}$ (g) $t = 4.03\,\text{s}$

(h) $t = 5.16\,\text{s}$ (i) $t = 5.96\,\text{s}$ (j) $t = 6.33\,\text{s}$ (k) $t = 6.76\,\text{s}$ (l) $t = 7.96\,\text{s}$ (m) $t = 8.36\,\text{s}$ (n) $t = 10.49\,\text{s}$

FIGURE 11: Real (1st and 3rd row) and virtual (2nd and 4th row) angiograms of phantom data for different time steps, which are denoted below the images. $t = 0\,\text{s}$ corresponds to the respective beginning of the DSA sequence.

numerical solution has to be considered, for which an explicit fourth-order Runge-Kutta scheme, given by

$$\boldsymbol{\rho}_i(t + \delta t) = \boldsymbol{\rho}_i(t) + \frac{1}{6} \cdot (k_1 + 2k_2 + 2k_3 + k_4),$$

$$\text{where } k_1 = \delta t \cdot f\left(t, \boldsymbol{\rho}_i(t)\right),$$

$$k_2 = \delta t \cdot f\left(t + \frac{1}{2}\delta t, \boldsymbol{\rho}_i(t) + \frac{1}{2}k_1\right),$$

$$k_3 = \delta t \cdot f\left(t + \frac{1}{2}\delta t, \boldsymbol{\rho}_i(t) + \frac{1}{2}k_2\right), \quad (5)$$

$$k_4 = \delta t \cdot f\left(t + \delta t, \boldsymbol{\rho}_i(t) + k_3\right),$$

is used [26].

For the choice of δt, the CFL condition [25], which correlates the time step, the given flow velocities, and the

resolution of the underlying computational grid, is taken as a reference.

Due to the discretization in time, the particles may be advected such that they leave the vessel through the boundary, which corresponds to a flux of contrast agent through a vascular wall. To prevent this, these particles will be kept inside by bouncing them at the vascular wall back into the vessel. This represents a physically reasonable approach under the assumption of rigid vascular walls.

Diffusion. For simulating the diffusive process according to Fick's law [28], given by

$$\mathbf{v}_{\text{Diff}}(\mathbf{x}, t) = -D \cdot \frac{\partial C(\mathbf{x}, t)}{\partial \mathbf{x}}, \quad (6)$$

the discrete particle set Ω is transformed into a continuous representation $C(\mathbf{x}, t)$ describing the concentration of contrast agent (see Smoothing). According to (6), the

(a) TIC for ROI0 (inlet)

(b) TIC for ROI1 (outlet 1)

(c) TIC for ROI2 (outlet 2)

— Real angiogram
— Virtual angiogram

— Real angiogram
— Virtual angiogram

(d) TIC for ROI4 (aneurysm dome)
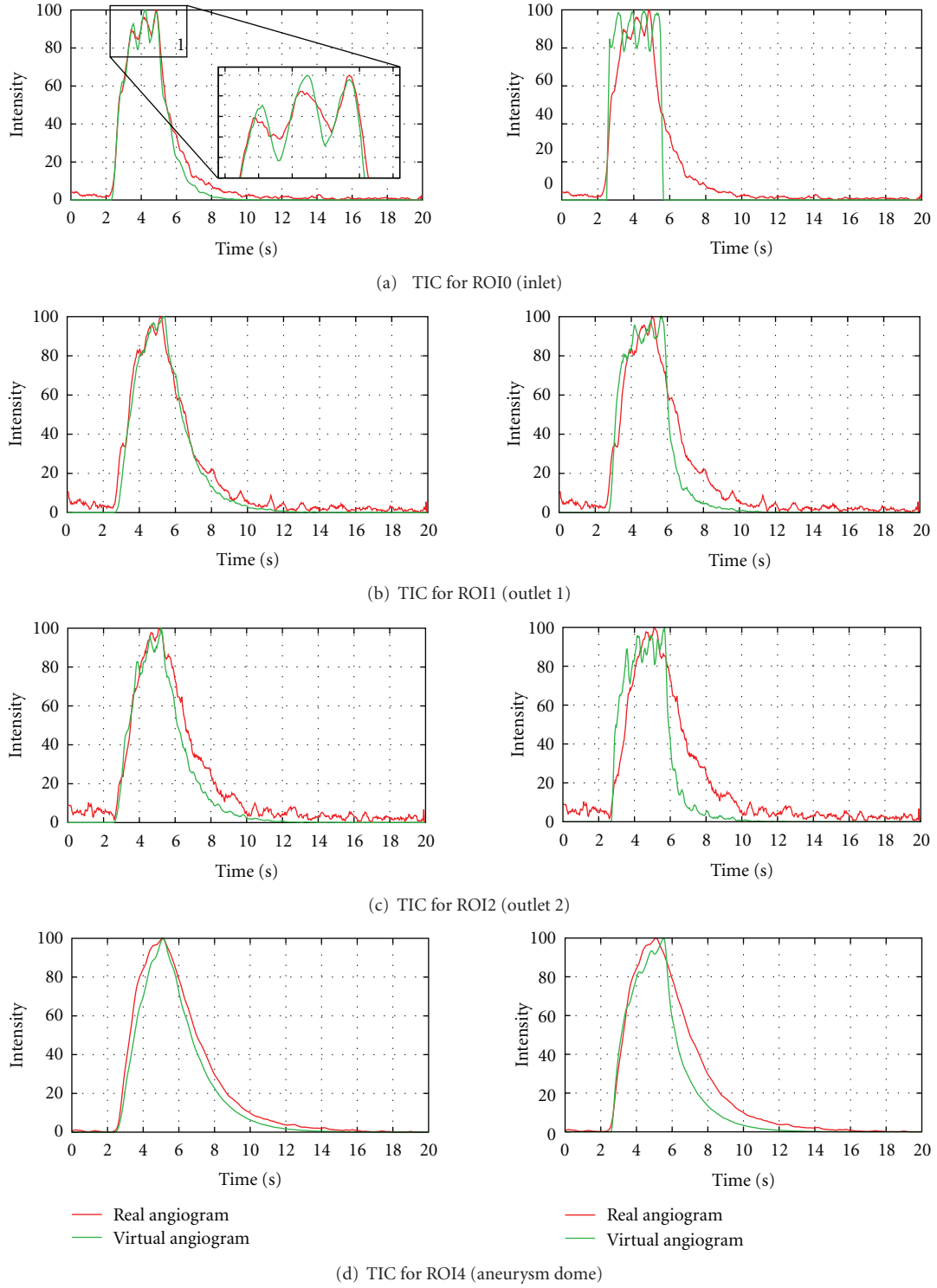
Figure 12: Time-intensity curves for phantom data set, based on real and virtual angiograms. On the left side, the heart state is synchronized and an injection bolus based on a capacitor function is used, whereas on the right side no synchronization is performed and a rectangular bolus profile is used.

(a) TIC for ROI0 (inlet)

(b) TIC for ROI1 (outlet 1)

(c) TIC for ROI2 (outlet 2)

(d) TIC for ROI4 (aneurysm dome)
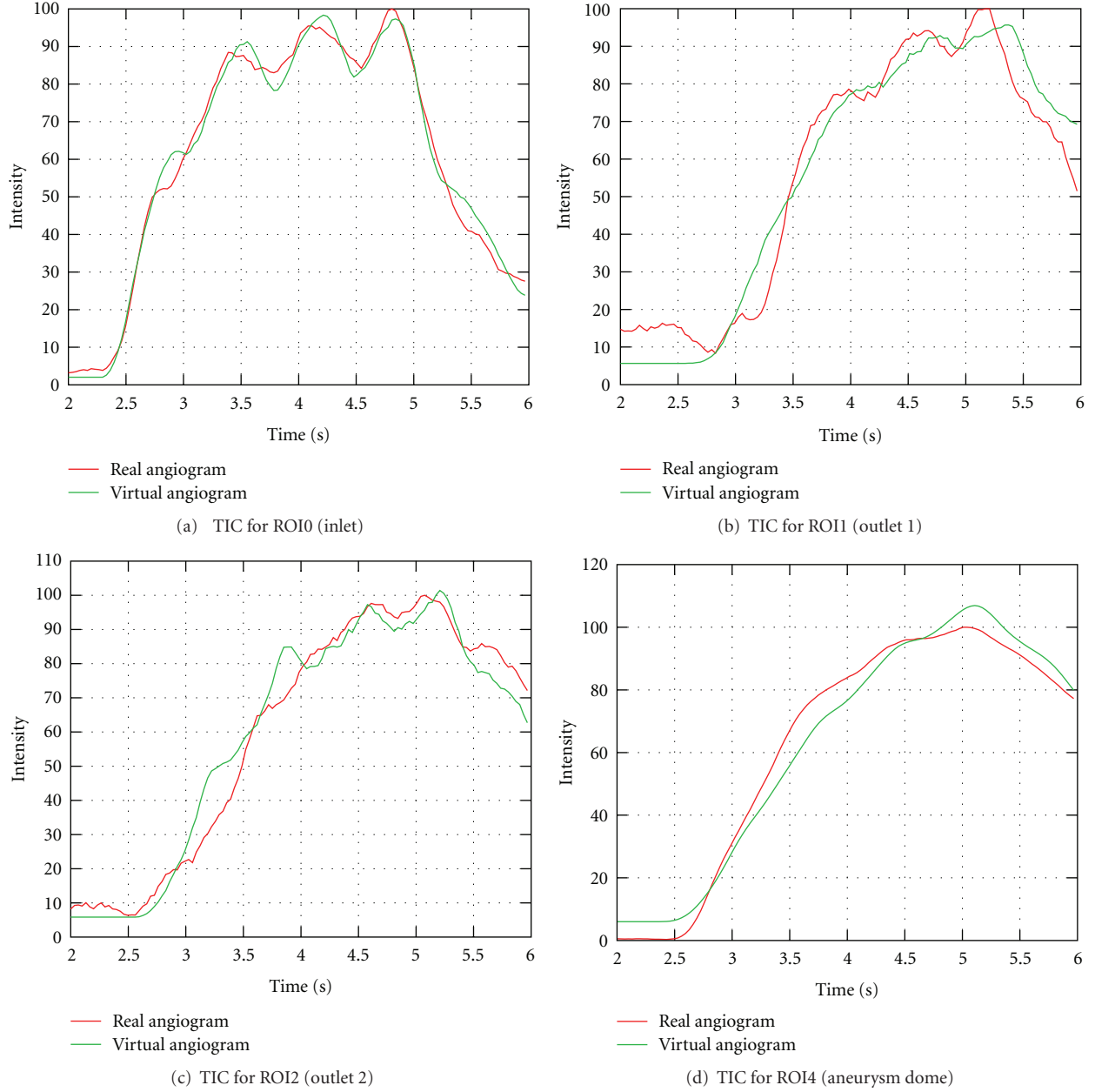
FIGURE 13: Time-intensity curves for phantom data set, based on real and virtual angiograms. The curves are cropped to arterial phase.

direction and the magnitude of the diffusive movement $\mathbf{v}_{\text{Diff}}$ is obtained by calculating the spatial gradient of the concentration image $C(\mathbf{x}, t)$, scaled by a substance-dependent diffusivity coefficient D. The resulting gradient image is subsequently used as the velocity image for advancing the particles according to diffusion.

Since contrast agent is restricted to the interior of the vessels, high concentration differences will occur at vessel boundaries, which in turn will generate large contrast medium concentration gradients. Consequently, contrast agent (i.e., particles) touching the boundary will keep on diffusing strongly towards the boundary, which results in those particles being bounced back into the vessel. Therefore,

an intermediate step is taken. After the discrete-continuous transformation, zero gradients are assured at the vessel boundary by extending the concentration from inside the vessel over the boundaries. This is achieved through the use of a distance transform [29], where each voxel outside the vessel is assigned an additional vector pointing to the closest voxel inside the vessel. This vector is then used to copy the concentration values from voxels inside the vessel to corresponding voxels outside the vessel.

Smoothing. For the reprojection, that is, the forward projection, of contrast agent concentration volumes as well as for the simulated diffusion process (in particular, for

(a) $t = 1.37$ s   (b) $t = 1.53$ s   (c) $t = 2.00$ s   (d) $t = 2.93$ s   (e) $t = 4.66$ s   (f) $t = 5.83$ s   (g) $t = 7.76$ s
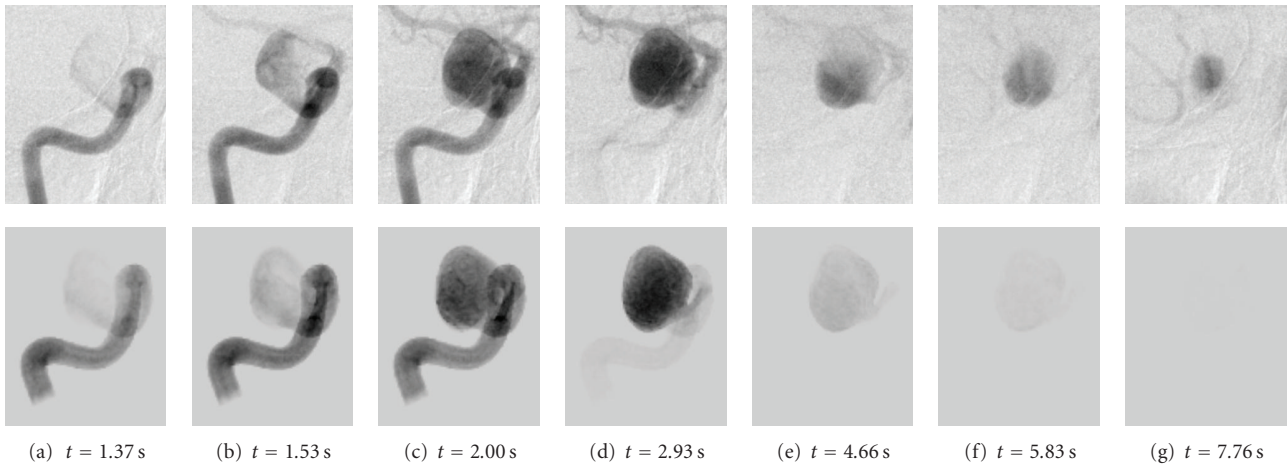
FIGURE 14: Real (1st row) and virtual (2nd row) angiograms of patient A, projection 1, for different time steps, which are denoted below the images. $t = 0$ s corresponds to the beginning of the DSA sequence.

(a) Location of ROIs for patient A, projection 1, C-arm angulation $(0°, 0°)$

(b) TIC for ROI0 (inlet)

(c) TIC for ROI1 (inflow aneurysm)

(d) TIC for ROI2 (aneurysm dome)
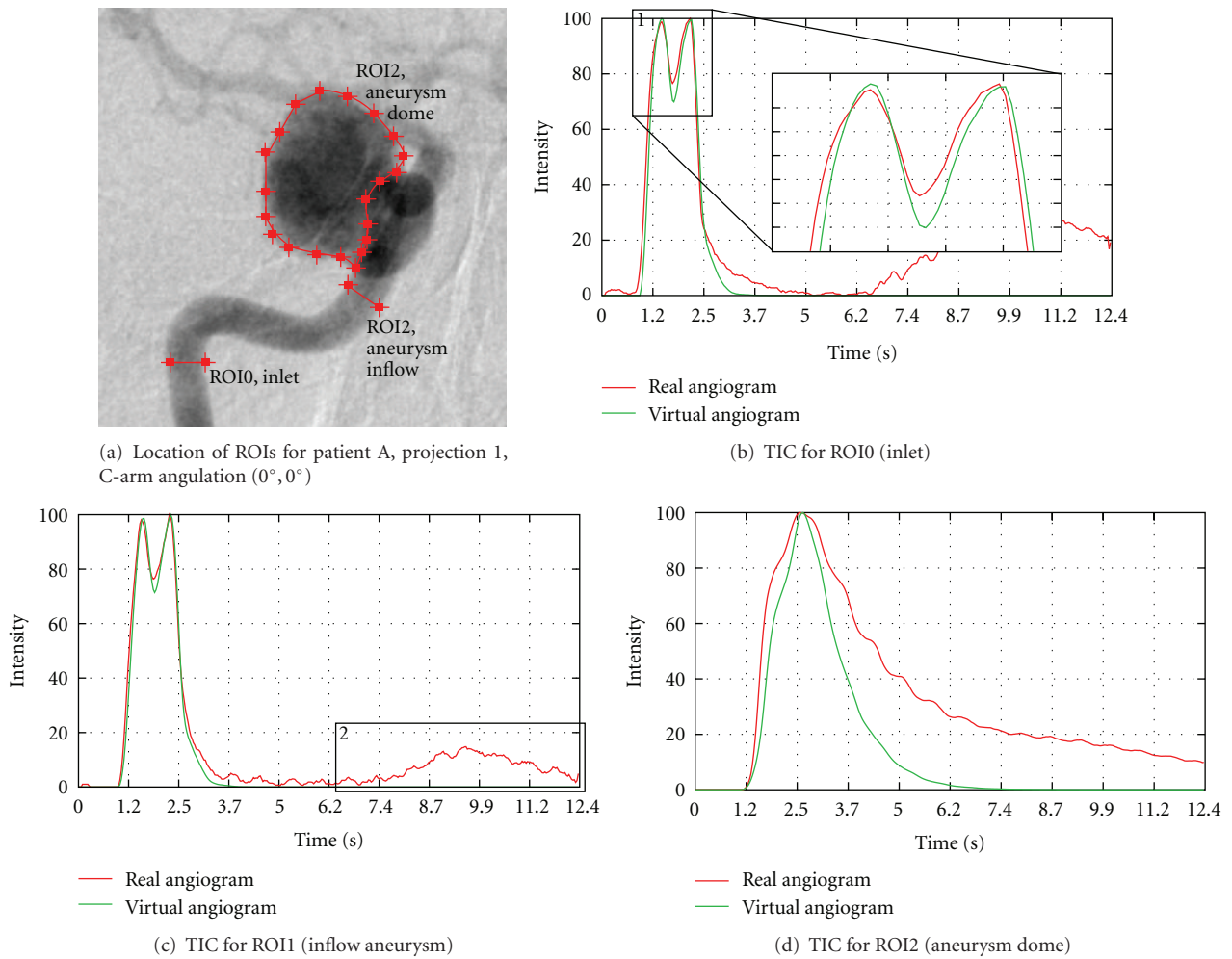
FIGURE 15: Time-intensity curves for patient A, projection 1, based on real and virtual angiograms.

(a) TIC for ROI0

(b) TIC for ROI1

(c) TIC for ROI2

FIGURE 16: Time-intensity curves for patient A, projection 1, based on real and virtual angiograms. The curves are cropped to arterial phase.

the calculation of the gradient vector field), the discrete particle representation is required to be transformed into a continuous representation of contrast agent such that its distribution is available on a regular grid. This is achieved by the following smoothing step:

$$C_{\text{discrete}} = \delta(\mathbf{x}) \longrightarrow C_{\text{continuous}} = \sum_{\boldsymbol{\rho}_i \in \Omega} f\left(\mathbf{x}, \boldsymbol{\rho}_i\right), \qquad (7)$$

where

$$\delta(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} - \boldsymbol{\rho}_i = 0, \ \boldsymbol{\rho}_i \in \Omega, \ \mathbf{x} \in \mathbb{R}^3, \\ 0, & \text{else.} \end{cases} \qquad (8)$$

This transformation describes a smearing (or smoothing) operation of a particle over its spatial neighborhood. The range and the way the particle gets smoothed is thereby specified by the function $f(\mathbf{x})$, for which a Gaussian distribution, given by

$$f\left(\mathbf{x}, \boldsymbol{\rho}_i\right) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} e^{-(1/2)(\mathbf{x}-\boldsymbol{\rho}_i)^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\rho}_i)},$$

$$\boldsymbol{\rho}_i \in \Omega, \ \mathbf{x} \in \mathbb{R}^3, \qquad (9)$$

is chosen. A smoothing parameter σ is thereby used for the covariance matrix $\Sigma = \sigma \cdot I_3$, where I_3 denotes the identity matrix, in order to parametrize the amount of smoothing.

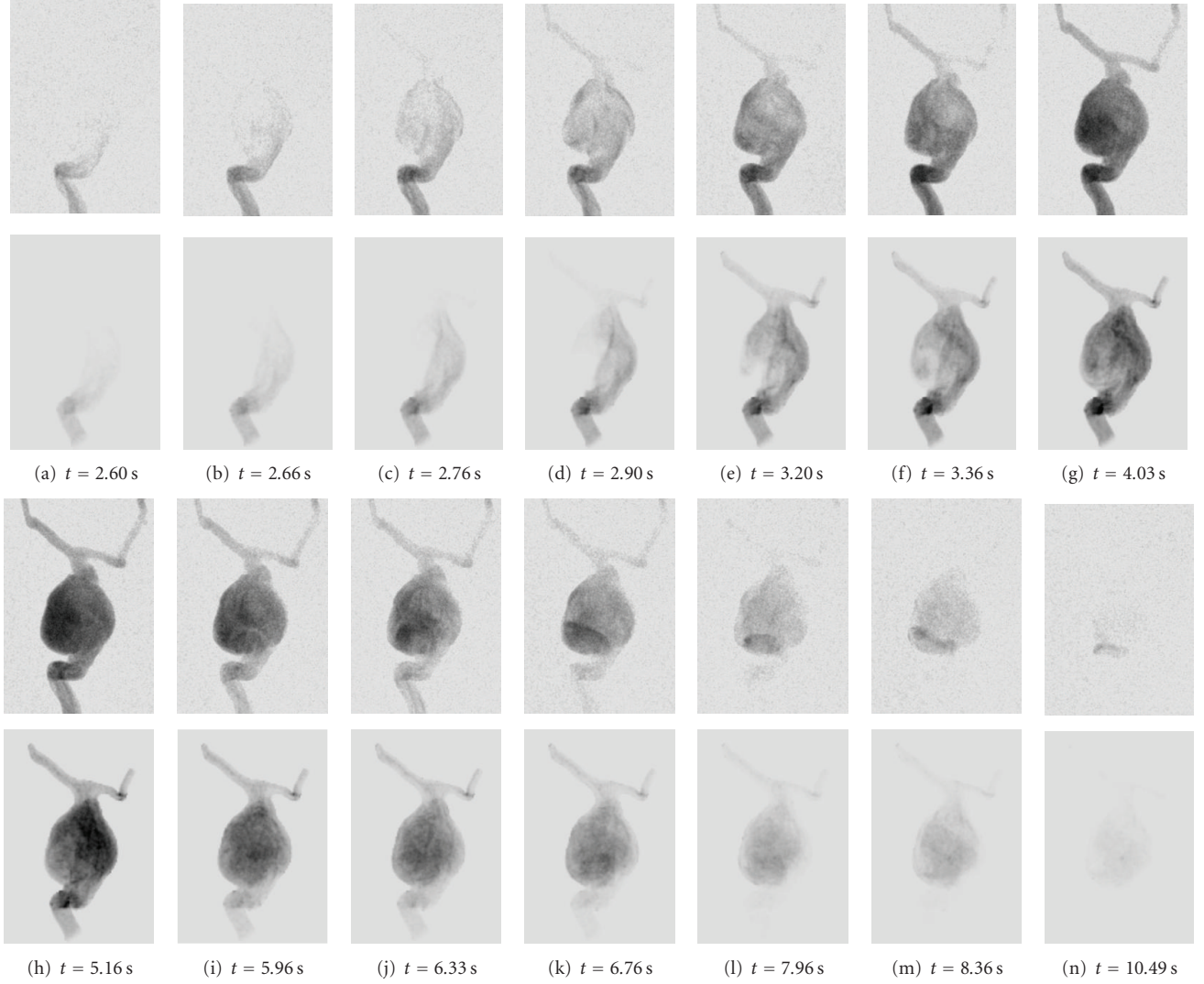(a) $t = 1.30$ s   (b) $t = 2.26$ s   (c) $t = 3.73$ s   (d) $t = 4.66$ s   (e) $t = 6.93$ s   (f) $t = 9.79$ s

FIGURE 17: Real (1st row) and virtual (2nd row) angiograms of patient A, projection 2, for different time steps, which are denoted below the images. $t = 0$ s again corresponds to the beginning of the DSA sequence.

(a)

(b)

FIGURE 18: Location of ROIs for patient A, projection 2, C-arm angulation $(-91°, -0.2°)$.

As the mean value for the Gaussian distribution, the particle position ρ_i is used.

Taking all particles into consideration, the final continuous distribution is hence a mixture of single density distributions (see Figure 4). This mixture density can then be sampled on the desired grid, which is here chosen in correspondence to the grid from the CFD solver (i.e., number of grid points per dimension and grid spacing) to keep the properties of the vessel geometry.

Using a Gaussian distribution as smoothing function has several benefits. First, the exact position of the particle between grid points is respected (by using it as the mean value of the Gaussian distribution); other solutions, such as assigning the particle to the nearest grid point, would typically shift the original particle position. Second, the Gaussian function allows to model a spatially symmetric smoothing. Third, when varying the smoothing parameter σ, the integral of the Gaussian and hence the total amount of spatially distributed contrast agent remains constant.

2.3. Patient-Specific Parameter Extraction

Heart Rate. The heart rate of each patient varies dependent on the patient's age, physical constitution, and so forth. For assuring a synchronized pulsatile pattern of real and virtual angiograms, the use of an average heart rate taken from the medical literature should thus be avoided. Measuring the patient-specific heart rate before treatment is also disadvantageous, because differences may occur between the resting heart rate and the heart rate during the treatment due to physiological factors such as stress or medication. Hence, it is desirable to have this information for the exact time when the patient is being treated.

To obtain the patient-specific heart rate, a user-defined line of interest (LOI) within the acquired 2D DSA series (Figure 5) is used. For each image of the 2D DSA series, the image intensities are integrated along this LOI. This results in a time-intensity curve which characterizes the concentration of contrast agent over time. Succeeding intensity peaks of the

(a) TIC for ROI0 (inlet)

(b) TIC for ROI1 (complete aneurysm dome)

(c) TIC for ROI2 (with contrast settling)

(d) TIC for ROI3 (without contrast settling)
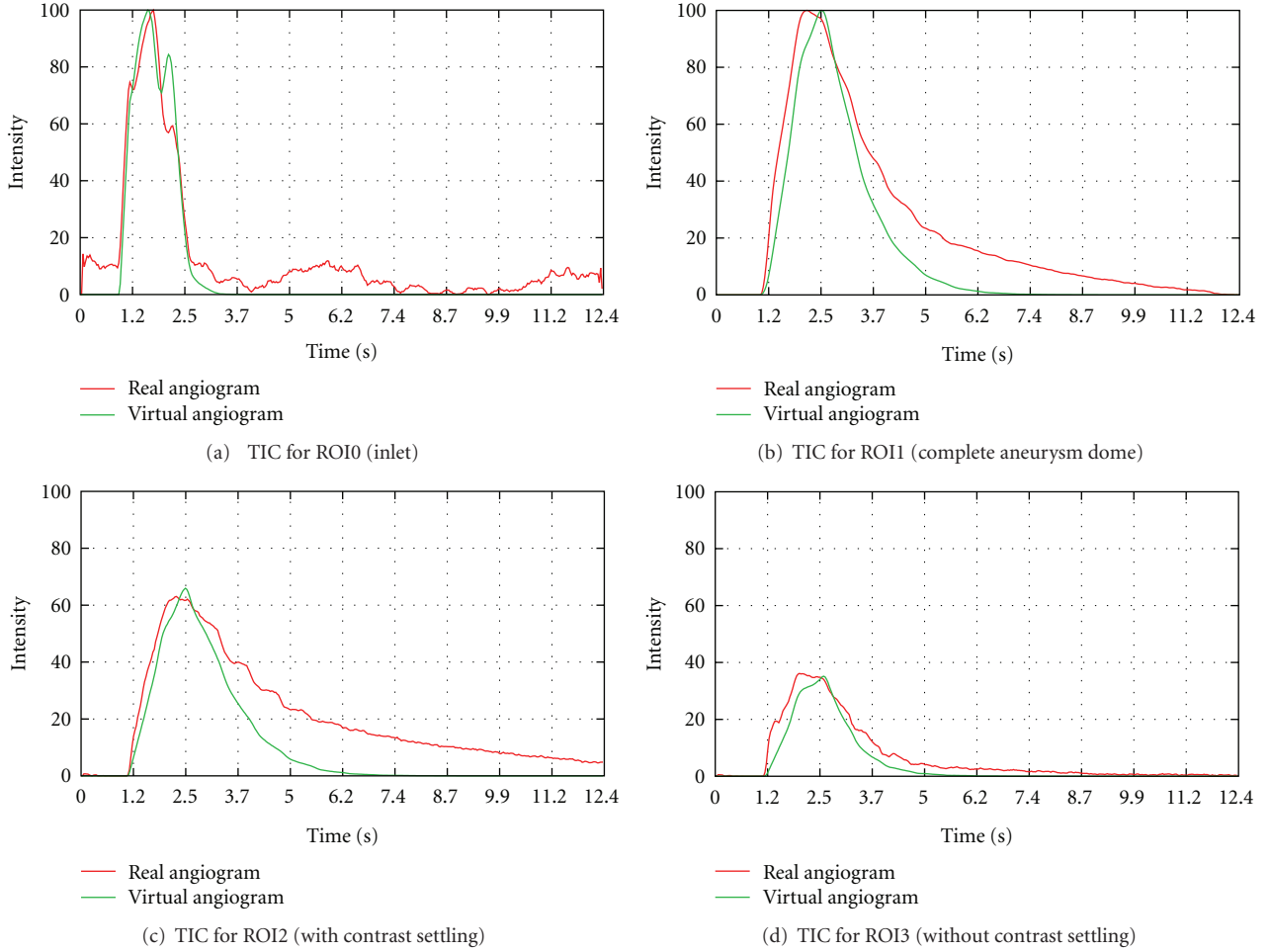
FIGURE 19: Time-intensity curves for patient A, projection 2, based on real and virtual angiograms.

measured data are then identified, out of which the heart rate h is calculated using

$$h = \frac{1}{n-1}\sum_{i=2}^{n}(p_i - p_{i-1}), \tag{10}$$

where p_i denotes the time point of the ith intensity peak, and n represents the total number of identified intensity peaks within the DSA series (Figure 5).

This heart rate is then used for the CFD simulation; the inflow velocity profile is adapted according to the calculated duration.

Heart State. In order to generate virtual angiograms that match the corresponding real angiograms as accurately as possible, not only the duration of a cardiac cycle, but also the state of the heart at the beginning of both angiograms, which affects the blood flow velocities and hence the propagation of contrast agent, must be synchronized.

For this purpose, the time-intensity curve based on the proposed LOI (Figure 5(b)) is used and the intensity peaks are identified. Under the assumption of a periodic heart

beat, these peaks are extrapolated backward in time to the beginning of the time-intensity curve (Figure 6(a)).

Now, the positions of the extrapolated intensity peaks next to the beginning of the acquired DSA signal are observed (Figure 6(a), box 1). Based on these positions, the CFD inflow velocity profile, which has already been adapted for the patient-specific heart rate, is periodically shifted such that this velocity profile would produce the same pulsatility pattern and hence the same intensity peaks as acquired for the patient. The relation between intensity peaks and inflow velocity profile is given by a correspondence of high velocities and low contrast agent intensities, since high velocities cause a large amount of blood to pass the injection point per time, but still absorbing the same amount of contrast agent (assuming a constant injection rate). This leads to a lowered contrast agent concentration and thus to reduced intensities on the acquired DSAs. By shifting the inflow profile, low velocities are assured to match the intensity peaks (Figure 6(c)).

Inflow Velocities. Inflow velocities are estimated by subsequently performing CFD simulations with varying mean

(a) TIC for ROI0 (inlet)

(b) TIC for ROI1 (complete aneurysm dome)

(c) TIC for ROI2 (with contrast settling)

(d) TIC for ROI3 (without contrast settling)
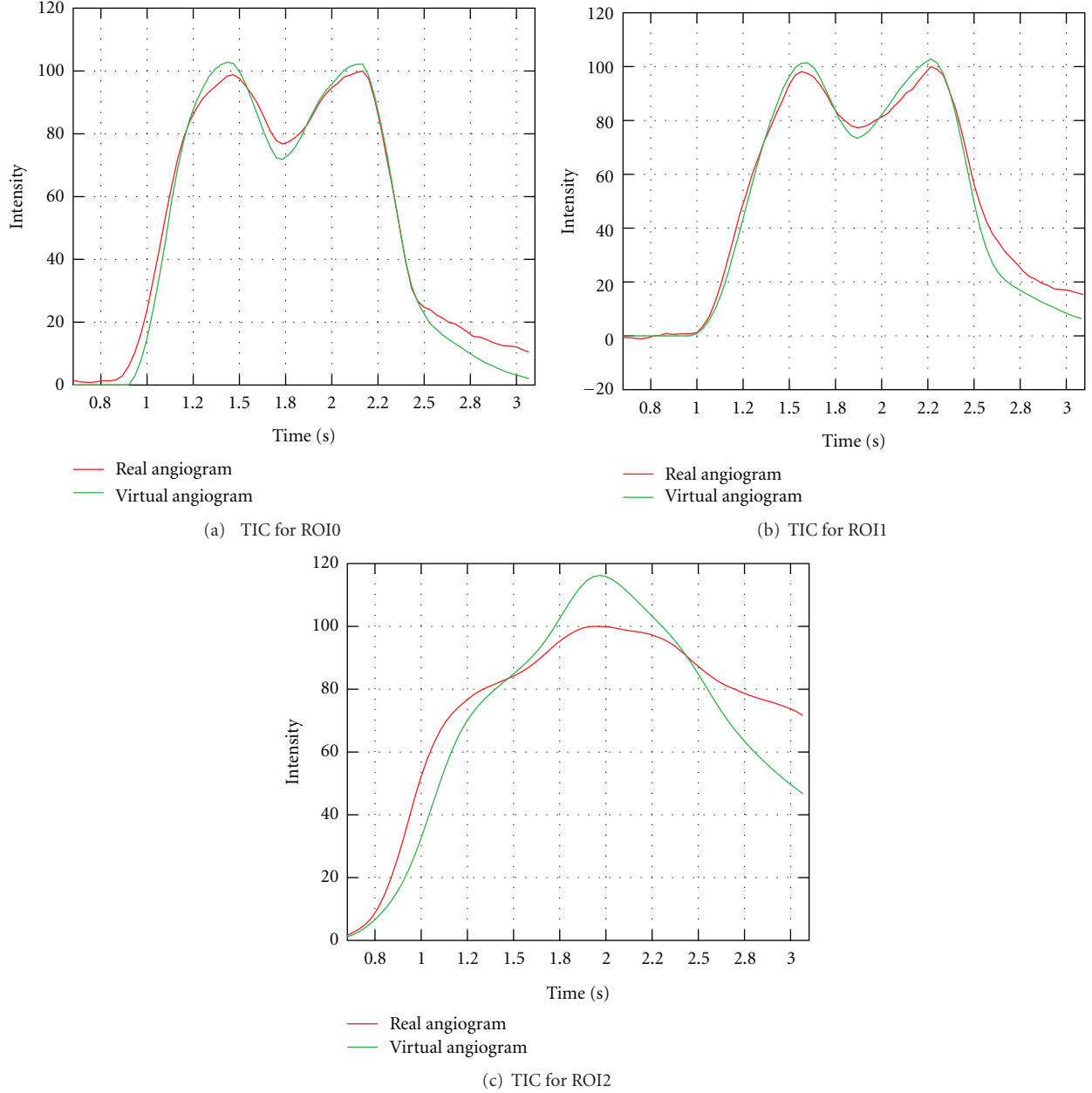
FIGURE 20: Time-intensity curves for patient A, projection 2, based on real and virtual angiograms. The curves are cropped to arterial phase.

inflow velocity. For each CFD simulation, a virtual angiogram is generated, and time-intensity curves are acquired at certain regions of interest for the real and virtual angiogram; for example, proximal to the aneurysm. Based on a comparison of these time-intensity curves, the mean inflow velocity is optimized manually to match proximal flow patterns.

Bolus Injection Profile. In angiographic procedures, injecting contrast agent is either performed using a mechanical injection regulator or an injection by hand. Especially for manual injections, the profile (e.g., duration, pressure) at

which a certain amount of contrast agent is inserted into the respective artery differs from injection to injection.

Therefore, it is not reasonable to use generic injection boli for virtual angiography, but to gain this information from the current patient case instead. Furthermore, additional physical and physiologic effects cause the injection bolus to alter, which means that the bolus does not arrive at the aneurysm as set up originally (e.g., constant injection rate of 2 mL/s for 3 s, resulting in an ideal rectangular bolus). These effects are, for instance, caused by the inner resistance of the catheter [30] and the contrast agent diffusion, since the point of injection is usually located at a certain distance proximal to the aneurysm.

(a) $t = 1.10\,\text{s}$ (b) $t = 1.17\,\text{s}$ (c) $t = 1.30\,\text{s}$ (d) $t = 1.60\,\text{s}$ (e) $t = 1.76\,\text{s}$ (f) $t = 1.93\,\text{s}$ (g) $t = 2.26\,\text{s}$
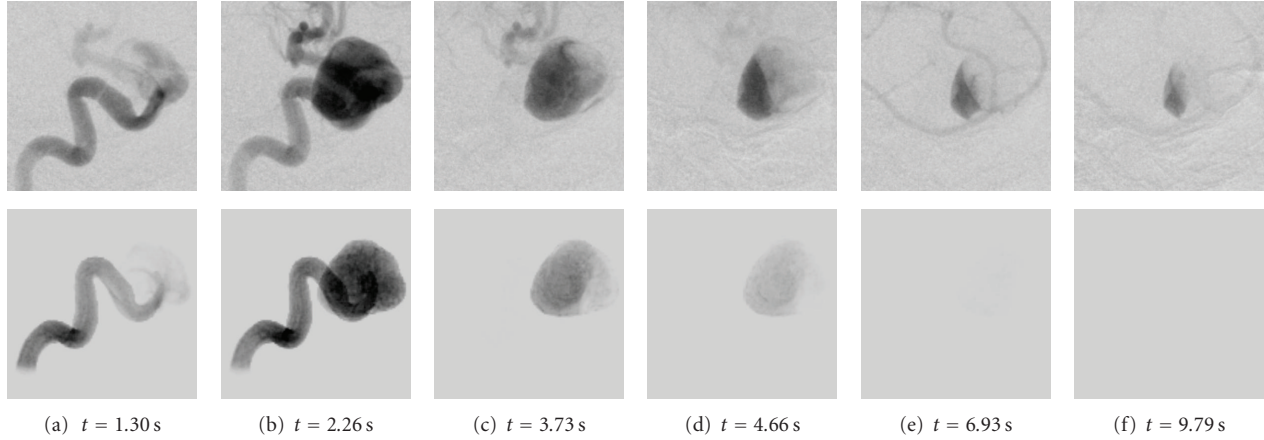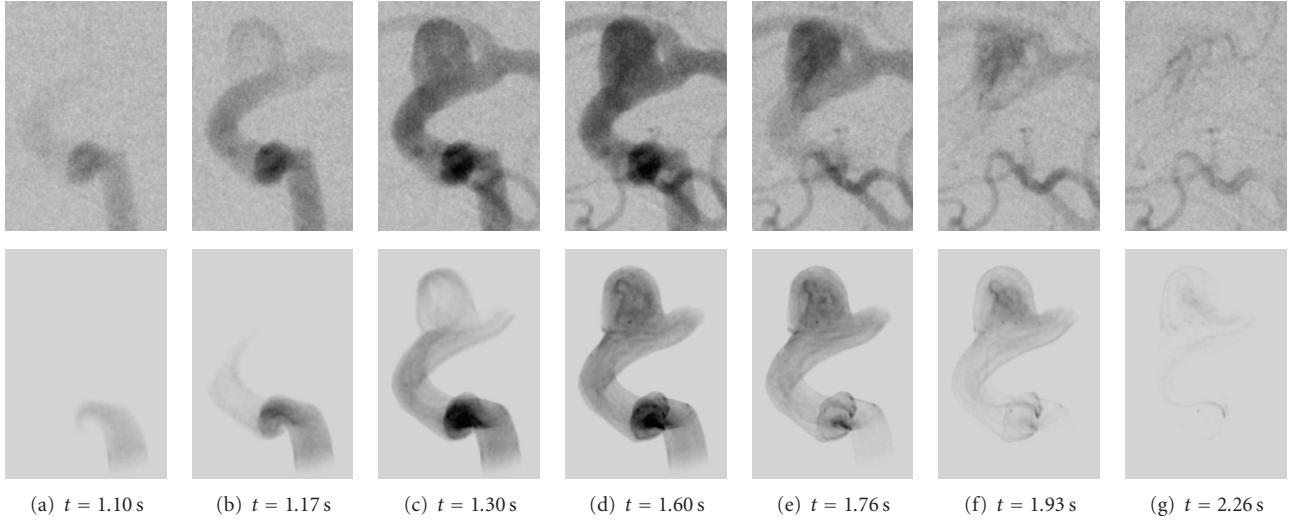
FIGURE 21: Real (1st row) and virtual (2nd row) angiogram of patient B for different time steps, which are denoted below the images. $t = 0\,\text{s}$ corresponds to the beginning of the DSA sequence.
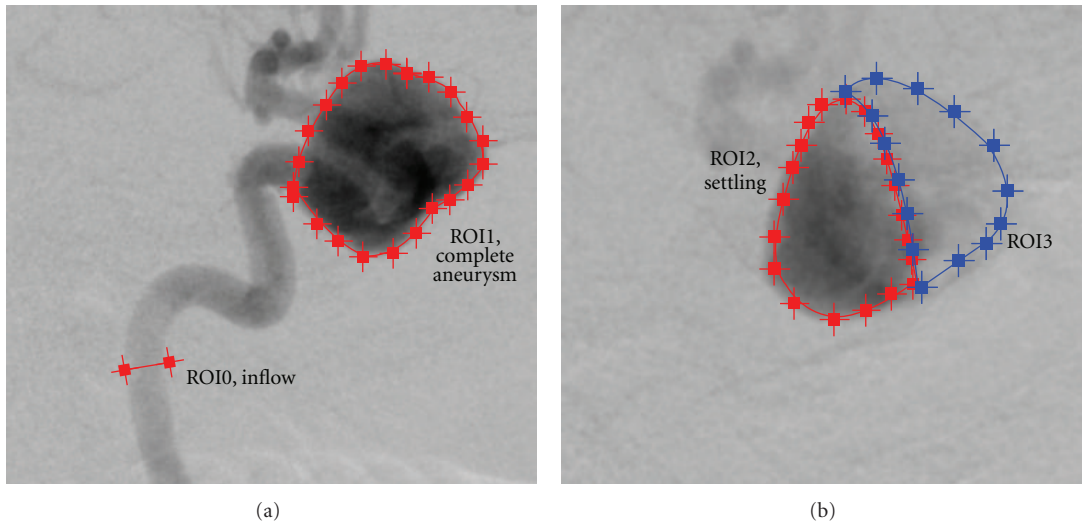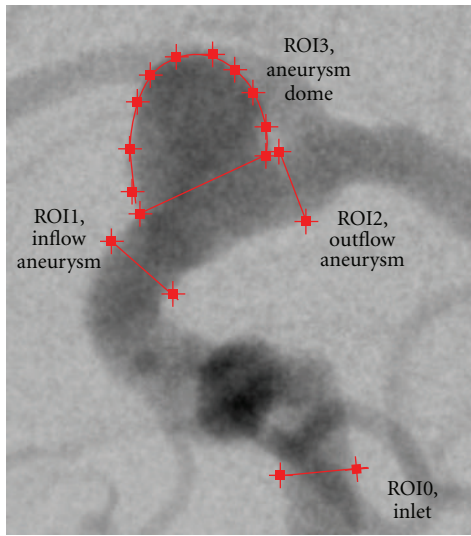
FIGURE 22: ROIs for patient B.

In order to consider these aspects for the virtual angiography, the injection bolus profile that we use in our approach is directly extracted from the acquired 2D DSA series. For this purpose, a time-intensity curve—as used for extracting heart rate and heart state information—is again employed, whose LOI is supposed to be located at the inlet plane of the CFD simulation, since that is where the virtual contrast agent is injected, as shown in Figure 5.

However, due to its mixing with blood, the contrast agent and hence the acquired time-intensity curve reflect the pulsatility caused by the patient's cardiac activity. Using this time-intensity curve directly as injection bolus profile would thus lead to the measured pulsatile pattern being included in the virtual angiography simulation. Since the velocity field generated by the CFD solver already implies a pulsatile

pattern itself, this means that pulsatility impact would in fact be considered twice.

To eliminate this pulsatile pattern and further background noise, the measured data will be fit to a predefined function. In order to determine the best-fit model parameters, the Levenberg-Marquardt optimization algorithm is used [31], which iteratively solves least-square optimization problems for nonlinear functions using a combination of the steepest descent and the Gauss-Newton method.

For the sake of representing the profile of the bolus injection over time, an adapted function based on [30] is used. Analogous to the electrical behavior of a capacitor, this function is given by

$$
I(\mathbf{p}, t)
= \begin{cases}
t < p_1 & 0, \\
t \geq p_1 \wedge t < p_1 + p_2 & p_5 \cdot \left(1 - e^{-(t-p_1)/p_3}\right), \\
\text{else} & p_5 \cdot \left(1 - e^{-p_1/p_4}\right) \\
& \quad \cdot \left(-e^{-(t-(p_1+p_2))/p_4}\right),
\end{cases}
\tag{11}
$$

where the parameter set $\mathbf{p} = \{p_i\}$, $i \in 1, \ldots, 5$ describes the curve according to Figure 7(b). To some extent, the curve resembles a rectangular function and hence rather originates from a constant contrast bolus injection which is just slightly altered by diffusive and inner resistance of the catheter. In Figure 7(a), an example of a capacitor curve fit is illustrated.

The fitted injection bolus curve $I(t)$ is finally resampled at n points, where n denotes the number of time steps to simulate. By normalizing $\sum_{i=1}^{n} I_i = 1$, a multiplication of I_i with the total number of particles to be injected results in the number of particles to be injected at time i.

2.4. 3D/2D Transformation—Forward Projection. Creating 2D projection images out of contrast agent concentration
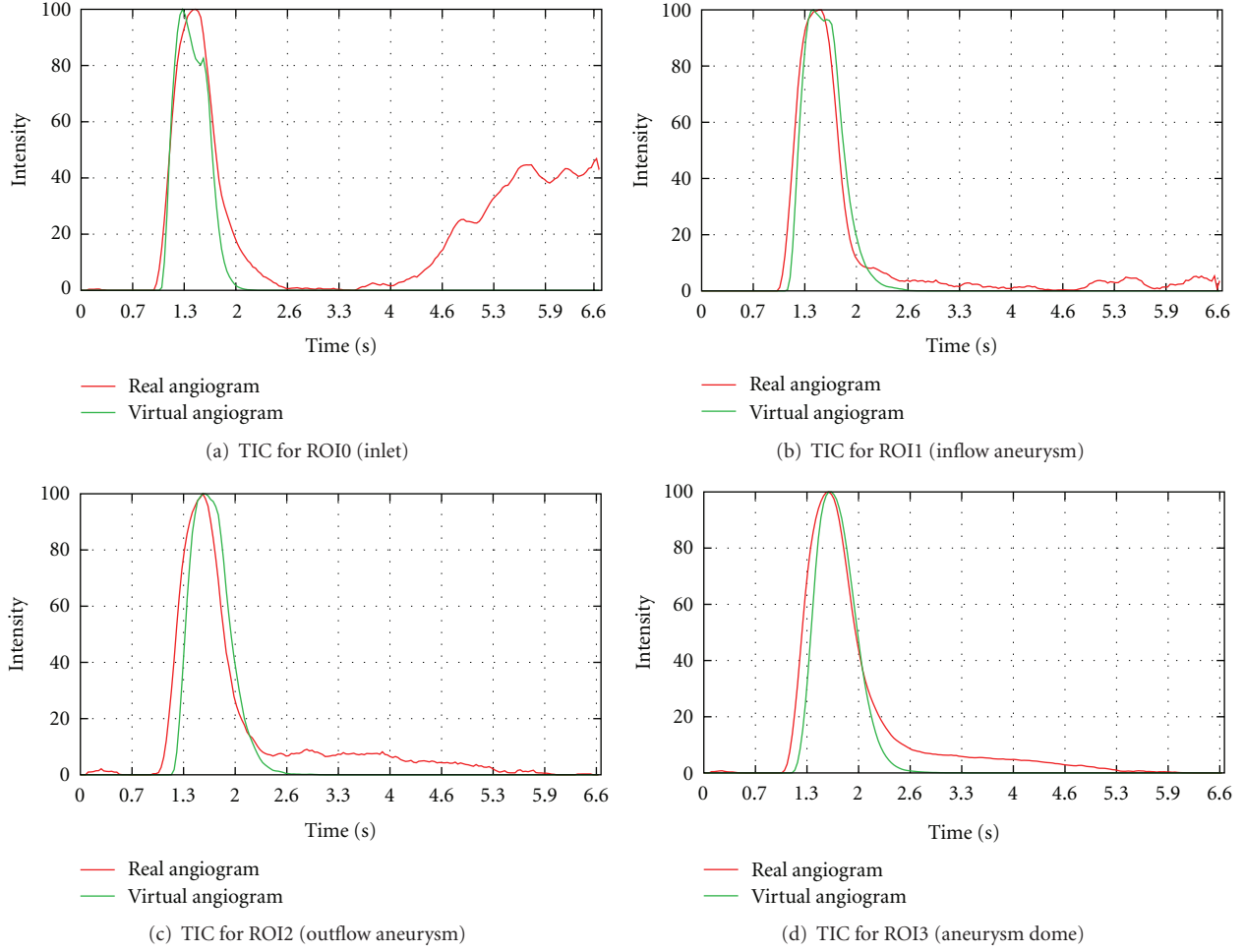
FIGURE 23: Time-intensity curves for patient B, based on real and virtual angiograms.

volumes can be described by a transformation $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$, which corresponds to the X-ray acquisition in real angiographic procedures. For the virtual angiography, the simulated projection is supposed to be computed such that the viewing directions of real and virtual angiograms match.

X-ray imaging is based on a source emitting and an image detector collecting photons. Since the (idealized) X-ray source is a point source, this system can geometrically be described by a pinhole camera model based on perspective projections [32]. Within this model, the mapping T of a point $\mathbf{p} \in \mathbb{R}^3$, located between the source and the detector, onto the image plane can be expressed as a linear transformation (in case homogeneous coordinates are used [32]) and performed using matrix calculations. This allows to represent T by

$$\mathbf{p}'_{\text{proj}} = T(\mathbf{p}) = P \cdot \mathbf{p}', \qquad (12)$$

where $\mathbf{p}' \in \mathbb{R}^4$ is the homogeneous representation of the point $\mathbf{p}$, $P \in \mathbb{R}^{3 \times 4}$ is the projection matrix, and $\mathbf{p}'_{\text{proj}}$ is the projected point, given in homogeneous coordinates.

For the case of C-arm imaging in the angiography suite, projection matrices are used for several applications; for example, for the 3D image reconstruction process based on

3D RA acquisitions [33]. For that purpose, the individual positions of the C-arm during the acquisition are specified by the used protocol and hence known in advance, which allows to generate the projection matrices within a calibration run once when the system is installed or maintained [6]. However, the acquisition of 2D DSA series is in general performed using an arbitrary C-arm angulation, for which no calibrated projection matrix is available. Nevertheless, those projection matrices are needed for generating the virtual angiograms which correspond to the real ones concerning the viewing direction.

According to the study in [34], the projection matrices can generally be calculated based on available information of the C-arm system. For this purpose, the angulation of the C-arm in left/right (LAO/RAO) as well as in head/feet (CRAN/CAUD) direction, the pixel spacing of the detector, the source-image-distance (SID), the source-to-isocenter distance (SISOD), and the coordinates of the isocenter related to the image plane are required. This information can be retrieved from the DICOM header of the particular DSA series.

For these calculations, an idealized projection geometry of the system is supposed, which neglects for instance gantry

(a) TIC for ROI0 (inlet)

(b) TIC for ROI1 (inflow aneurysm)
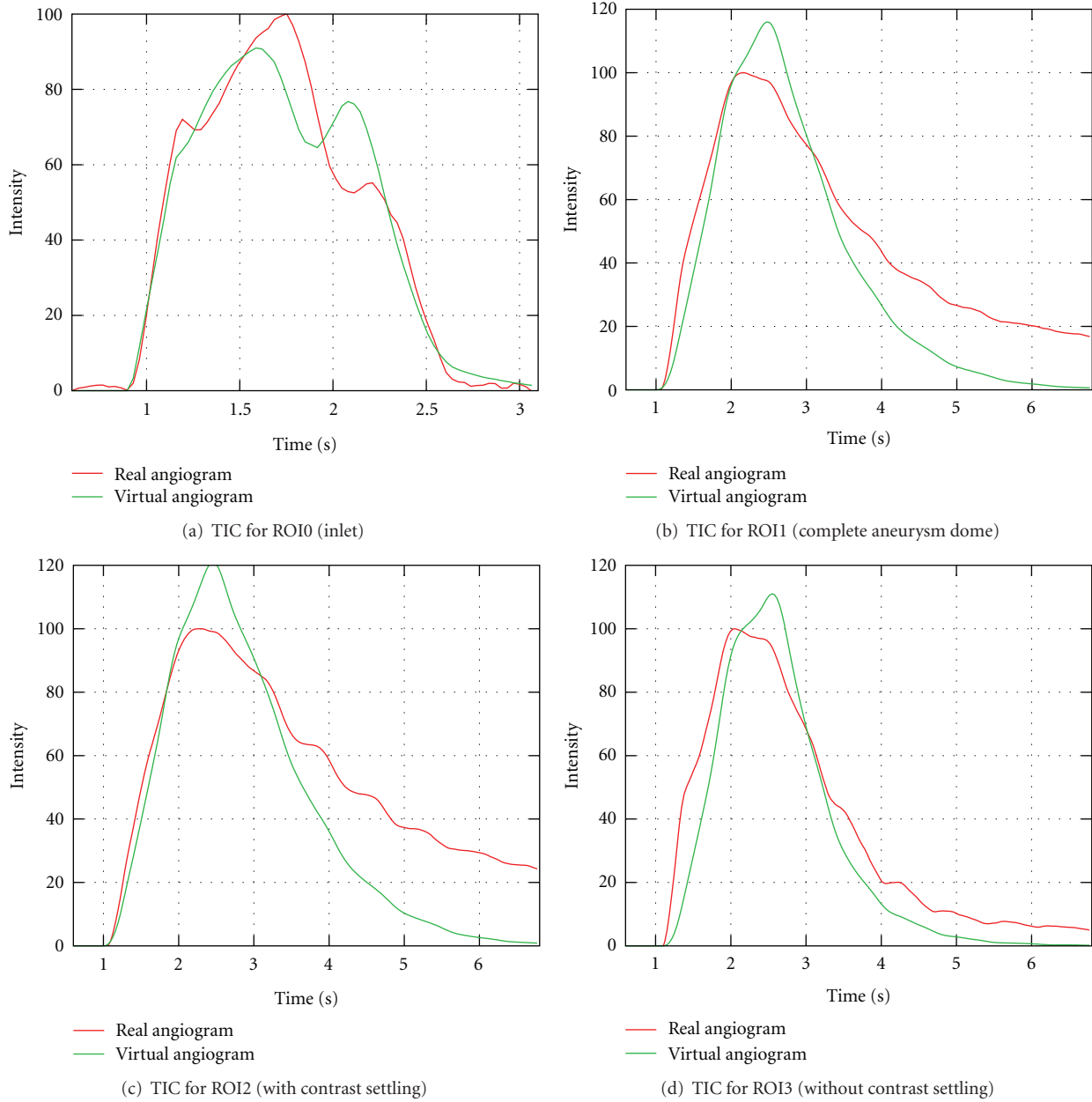

(c) TIC for ROI2 (outflow aneurysm)

(d) TIC for ROI3 (aneurysm dome)
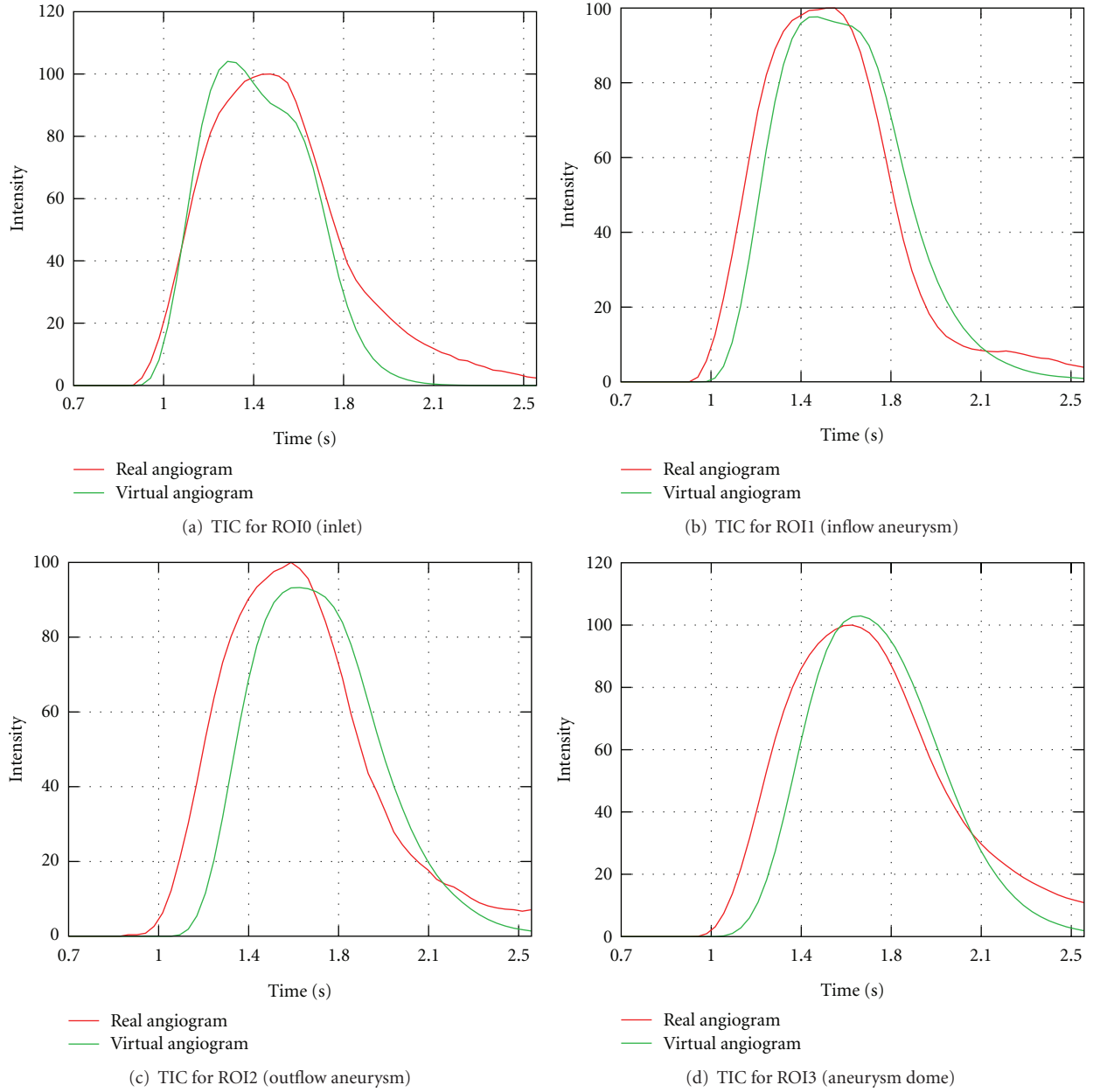
FIGURE 24: Time-intensity curves for patient B, based on real and virtual angiograms. The curves are cropped to arterial phase.

motion as well as mechanical instabilities. This in general leads to projection matrices being less accurate than the calibrated ones, see [34] for details. The resulting virtual projection images may then slightly differ from the acquired images with respect to the viewing direction.

The calculated matrices are eventually used to project the contrast medium concentration volume onto the virtual image detector. Our framework uses an implementation based on the ray casting technique [35], which generates a ray for each pixel of the virtual image plane that intersects the corresponding pixel, the virtual X-ray source, and the concentration volume that is located between the virtual source and the virtual detector. The rays are then sampled

equidistantly and the concentration values for these positions, which correspond to X-ray attenuation coefficients, are added up. This summation corresponds to the numerical approximation of the line integral of X-ray attenuation values along the respective ray.

2.5. Comparison/Evaluation—Methods. In total, three different cases were used for testing, each of them consisting of a 3D RA data set and at least one high-speed DSA series (30 fps) showing a complete bolus passage, see Table 1. For one case (patient A), two DSA series are available, which show the bolus injections from different C-arm angulations. The first case (phantom data) is a medical phantom of a

giant artificial aneurysm, whose shape is based on a real patient case. The pulsatile blood flow is modeled using a combination of a steady and a pulsatile pump. The other cases (patient A, patient B) represent data from two patients with aneurysms at the internal carotid artery. The data was provided by Stony Brook University, New York (phantom data), and the Department of Neuroradiology, University of Erlangen-Nuremberg (patient A, patient B).

The evaluation is performed on a qualitative and a quantitative basis. For qualitative comparisons, features such as synchronization of time, a match of global flow patterns, zones of recirculation, and the location of the inflow jet are inspected.

For the comparison between virtual and real angiograms, a pixelwise correspondence is not given (see Section 2.4): Therefore, quantitative measurements are based on time-intensity curves (TICs) of certain regions of interest (ROI). These regions are selected by hand at corresponding positions of real and virtual angiograms.

In detail, the quantitative features, as depicted in Figure 8, are as follows [17, 36]:

(i) Full width at half maximum (FWHM): this measurement describes the duration $d_{FWHM} = t_2 - t_1$ between two points in time where the measured intensities reach half of the maximum intensity during wash-in and wash-out phases; that is, $f(t_1) = f(t_2) = (1/2)f_{max}$. This parameter indicates the rate at which blood (or contrast agent) is exchanged in the selected region of interest and is practically used for evaluating the outcome of a treatment by comparing pre- and postmeasurements.

(ii) Time to peak (TTP): time to peak describes the duration d_{TTP} until the maximum opacification f_{max}, that is, intensity, is reached. The duration is measured from the time when the opacification reaches 10% of total opacification for the first time. This parameter quantifies the wash-in phase.

(iii) Average washin/washout: the average slope at which the time-intensity curve increases and decreases, respectively. For that purpose, the durations from 10% of maximum opacification until f_{max} is reached and the decrease from f_{max} to 10% of maximum opacification is used. The parameter is taken for describing the inflow and the outflow behavior, respectively.

(iv) Relative root-mean-square error (rRMSE, [17]) between time-intensity curves of real and virtual angiograms. The TIC corresponding to the real angiogram is scaled such that its values lie in the range between 0 and 100; the TIC of the virtual angiogram is then shifted and scaled such that the rRMSE measurement gets minimal. The rRMSE is defined as

$$\text{rRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{P_i - T_i}{T_m}\right)^2}, \qquad (13)$$

where P_i and T_i denote the (normalized) intensities of the TICs of real and virtual angiograms. T_m represents the mean value of the TIC of the acquired angiogram. For certain patient cases, venous structures may overlay the observed vascular segment in the background, leading to a rerise of intensities after a certain time. Since, for simulated cases, this situation does not occur, this measurement may be strongly corrupted by diverging time-intensity curves between real and virtual angiograms when the contrast agent reaches those venous structures. To avoid that problem, this curve-based measurement is restricted to a selected part of the time-intensity curve only instead of the total simulated duration.

3. Results and Discussion

3.1. Arbitrary Projection Angles. As was mentioned in Section 2.4, the used projection matrices are generated from C-arm angulation information in order to reproject the calculated contrast agent concentration volume according to the projection direction of the acquired 2D DSA sequence. This step is needed to compare the real and virtual angiograms side by side for the purpose of validating CFD simulation results.

In principle, by choosing arbitrary values for the C-arm angulation (rotation in left/right and head/feet direction), arbitrary projection geometries are possible, compare Figure 9; this particularly enables angulations which cannot be reached in reality due to mechanical limitations.

It is important to note that the generation of additional virtual angiograms requires no further X-ray exposure or additional contrast agent delivery to the patient. Any number of virtual angiograms can be generated from any desired viewing angle without the need for additional imaging of the patient, thus potentially becoming an important tool in the treatment planning process for a number of cerebral vascular disorders.

3.2. Comparison of Real and Virtual Angiograms—Phantom Data. For the phantom data set, the simulated angiography was performed using a total of 10^6 particles. δt was chosen to be 0.002 s, the diffusivity coefficient $D = 0.1$, and the smoothing parameter $\sigma = 1.0$. In total, 20.04 s were simulated, corresponding to the acquired DSA series (601 frames at 30 fps). The injection bolus was modeled using the presented capacitor function (Figure 7, (11)). From the acquired 2D DSA series, a heart rate of 84 bpm was extracted. The underlying CFD simulation was performed using a blood viscosity of 0.004 Pa·s with a density of 1050 kg/m³. The inflow profile was synchronized with the heart state as described, with velocities ranging from 0.28 to 0.37 m/s and an average of 0.31 m/s.

The real and the corresponding virtual angiogram is shown in Figure 11. Overall, the angiograms show satisfying accordance concerning the temporal synchronization and global flow patterns. In both angiograms, the inflow jet enters the aneurysm at $t = 2.6$ s and proceeds along the right

TABLE 1: Cases used for testing and evaluation.

Data set	3D RA data		2D DSA	
	Size (voxel)	Resolution (mm)	Resolution (mm)	Frame rate
Phantom data	$512 \times 512 \times 512$	$0.46 \times 0.46 \times 0.46$	0.308×0.308	30 fps
Patient A	$512 \times 512 \times 396$	$0.28 \times 0.28 \times 0.28$	0.308×0.308	30 fps
Patient B	$512 \times 512 \times 396$	$0.1 \times 0.1 \times 0.1$	0.154×0.154	30 fps

TABLE 2: Quantitative measurements for phantom data.

DSA	ROI	FWHM	$\varnothing$ Washin (rad)	$\varnothing$ Washout (rad)	TTP	rRMSE
Real	ROI0,	2.74 s	0.89	−0.84	2.54 s	5.3%
Virtual	Inlet	2.67 s	1.05	−0.89	1.82 s	
Real	ROI1,	3.17 s	0.89	−0.66	2.51 s	13.3%
Virtual	Outlet1	2.84 s	0.89	−0.79	2.51 s	
Real	ROI2,	3.10 s	0.90	−0.59	2.44 s	10.8%
Virtual	Outlet2	2.71 s	0.91	−0.80	2.44 s	
Real	ROI3,	3.17 s	0.87	−0.61	2.67 s	7.6%
Virtual	Outlet	2.81 s	0.90	−0.80	2.44 s	
Real	ROI4,	3.73 s	0.90	−0.55	2.48 s	9.3%
Virtual	Aneurysm dome	3.20 s	0.91	−0.62	2.41 s	

aneurysm wall, as can be seen in Figures 11(a) and 11(b). Within the aneurysm dome, the contrast agent further circulates counter-clockwise Figures 11(c)–11(f). In Figures 11(c)–11(f), the real angiogram appears to have a slightly faster filling of the aneurym with contrast agent; having a closer look at the inflow jet, one can observe that the contrast agent for the real angiogram is distributed homogeneously very fast within the aneurysm dome, whereas the virtual contrast agent stays dense on the right side of the aneurysm Figure 11(d). Presumably, this different behavior might be a mismatch of assumed and real properties of blood, for instance viscosity, whose effects have not been observed in this study. For both cases, the opacification in the aneurysm reaches its maximum at approximately 4-5 s, compare Figures 11(g)-11(h). At $t = 5.2$ s (h), the inflow of contrast agent reduces and the remaining contrast agent is flushed out of the aneurysm Figures 11(i)–11(n). The outflow phase is well synchronized here. As can be seen in Figures 11(k)–11(n), a small amount of contrast agent remains at the bottom left side of the aneurysm in the real angiogram.

For this case, multiple regions of interest for measuring time-intensity curves are chosen, as depicted in Figure 10(a), to measure certain effects. ROI0, which is placed at a short distance behind the particle injection area, is supposed to show that certain conditions (heart rate, heart state, contrast bolus, synchronization of time) are in agreement for the beginning of the simulated and the real domain. ROI1, ROI2, and ROI3 measure the contrast bolus at the end of the simulated domain. Assuming identical measurements between real and virtual angiograms at ROI0, differences which occur in one of these ROIs might indicate a different behavior of contrast agent for the real and simulated

environment for preceding areas. Finally, ROI4 is chosen to cover the whole aneurysm dome to measure the global behavior of contrast agent within that domain.

In Figure 12, the measured time-intensity curves for the regions of interest depicted in Figure 10(a) are shown. The curves are normalized such that the measured intensities range from 0 to 100 for both angiograms. For the real angiogram, noise and motion cause the time-intensity curves to appear more ragged.

To demonstrate the importance of extracting the injection-specific bolus profile and the synchronization of the virtual contrast agent injection with the patient's cardiac activity according to the presented methods, time-intensity curves are additionally presented for a virtual angiography based on a generic, rectangular injection bolus. For that case, the duration of the injection is adapted to the FWHM measurement of a time-intensity curve acquired at the LOI shown in Figure 5(b).

In Figure 12(a), the time-intensity curves are given for the region of interest located at a short distance behind the particle injection area (ROI0). The curves are well synchronized in shape and time, which indicates a proper extraction of the contrast bolus injection profile, see Figure 10(b). The pulsatility in both angiograms further coincides well concerning the duration of the cardiac cycle and the synchronization of the heart phase, as can be seen in box 1, Figure 12(a). Differences in the height of the intensity peaks are possibly caused by the inflow velocities which are used for the CFD simulation. For this case, the range of presumed inflow velocities (0.28–0.37 m/s) is propably chosen too wide which may cause the intensities of the virtual angiogram to oscillate stronger than those of the real angiogram.

TABLE 3: Quantitative measurements for patient A, projection 1.

DSA	ROI	FWHM	∅ Washin (rad)	∅ Washout (rad)	TTP	rRMSE
Real	ROI0,	1.29 s	1.18	−1.28	1.29 s	9.2%
Virtual	Inlet	1.25 s	1.42	−1.16	0.5 s	
Real	ROI1,	1.29 s	1.20	−1.29	1.22 s	9.9%
Virtual	Aneurysm inflow	1.22 s	1.20	−1.34	1.22 s	
Real	ROI2,	2.87 s	1.20	−0.29	1.22 s	16.2%
Virtual	Aneurysm dome	1.72 s	1.20	−0.93	1.22 s	

TABLE 4: Quantitative measurements for patient A, projection 2.

DSA	ROI	FWHM	∅ Washin (rad)	∅ Washout (rad)	TTP	rRMSE
Real	ROI0,	1.29 s	1.29	−1.20	1.29 s	23.4%
Virtual	Inlet	1.25 s	1.37	−1.24	0.5 s	
Real	ROI1,	2.21 s	1.25	−0.50	1.22 s	33.6%
Virtual	Complete aneurysm	1.65 s	1.18	−0.92	1.22 s	
Real	ROI2,	2.61 s	1.23	−0.44	1.22 s	38.0%
Virtual	aneurysm part w/settling	1.78 s	1.18	−0.88	1.22 s	
Real	ROI3,	1.95 s	1.28	−0.78	1.22 s	30.3%
Virtual	aneurysm part w/o settling	1.45 s	1.18	−1.08	1.22 s	

The time-intensity curves for the outlets, that is, ROI1, Figure 12(b) and ROI2, Figure 12(c), show a strong correspondence for the rising edge between real and virtual angiograms, meaning that contrast agent arrives synchronously in time within the real and virtual angiogram.

In Figure 12(d), the time-intensity curves are compared for the region of interest covering the complete aneurysm dome. According to this figure, the maximum opacification is reached for both angiograms identically after 5 s, which has also been observed in Figure 11.

Table 2 denotes the quantitative values for the measured time-intensity curves. The measured FWHM is slightly increased (0.53 s) for the real aneurysm dome in contrast to that from the virtual angiogram. The FWHM at ROI0 (inlet) is in good agreement (difference: 0.07 s), but for the successive regions (outlet ROIs), an increased gap is detected (differences: 0.33 s–0.39 s).

The time to peak measurements coincide for ROI1–ROI4; the large difference for the inlet ROI (ROI0) is caused by the fact that the maximum opacification of the real curve is reached at the third main intensity peak, while the one of the virtual angiogram is already reached at the second peak, see box 1 in Figure 12(a).

For the washin phase, the values correspond to a great extent; the average washin for ROI0, which is increased for the real angiogram, arises from the peak-to-peak shift described above. Concerning the washout phase, the real angiogram has a decreased average washout rate (ROI1–ROI4), which might be caused by the remaining contrast agent within the aneurysm dome.

The calculated relative root mean square errors for the arterial phase of the time-intensity curves, see Figure 13, are in the range of 5.3%–13.3% and thus comparable to [17].

3.3. Comparison of Real and Virtual Angiograms—Patient A.
For patient A, who has a large/giant aneurysm at the internal carotid artery, two angiograms with C-arm angulations, (0°, 0°; anteroposterior view) and (−91°,−0.2°; lateral view), were evaluated. The parameter set is chosen for both cases as follows: 10^6 particles were used, δt was chosen to be 0.002 s, diffusivity $D = 1$, and smoothing factor $\sigma = 1$. The acquired DSA sequence has a duration of 12.4 s. The extracted heart rate is 81 bpm. The CFD inflow velocities have a mean of 0.31 m/s, ranging from 0.25 to 0.42 m/s. Blood specific parameters were chosen identical as for the phantom data.

3.3.1. Projection 1 (0° RAO/LAO, 0° CRAN/CAUD).
Figure 14 shows both angiograms. The inflow of contrast agent is well synchronized between the real and virtual angiogram, Figures 14(a) and 14(b). The inflow jet, which is clearly visible in both cases, pours into the aneurym at the bottom side and circulates clockwise. The depicted part of the vessel is homogeneously filled with contrast agent in (c). At $t = 2.93$ s, the inflow phase is finished for both angiograms.

Major differences between the real and virtual angiogram occur during the washout phase. In the virtual angiogram, the contrast agent is flushed out quite fast, whereas in the

TABLE 5: Quantitative measurements for patient B.

DSA	ROI	FWHM	∅ Washin (rad)	∅ Washout (rad)	TTP	rRMSE
Real	ROI0,	0.59 s	1.42	−1.31	0.5 s	35.2%
Virtual	Inlet	0.56 s	1.48	−1.39	0.3 s	
Real	ROI1,	0.63 s	1.41	−1.29	0.53 s	42.8%
Virtual	Aneurysm inflow	0.59 s	1.44	−1.37	0.43 s	
Real	ROI2,	0.66 s	1.41	−1.32	0.53 s	54.6%
Virtual	Aneurysm outflow	0.63 s	1.42	−1.37	0.50 s	
Real	ROI2,	0.73 s	1.41	−1.25	0.53 s	27.9%
Virtual	Aneurysm dome	0.63 s	1.43	−1.36	0.46 s	

real angiogram, a remaining part of contrast agent still stays inside the aneurysm Figures 14(e)–14(g).

Identical characteristics can be observed for the time-intensity curves in Figure 15 and the measured quantities based on these curves, as denoted in Table 3. The synchronization (time, heart rate, and heart state) for the inlet (ROI0) is given to a great extent, compare Figure 15(b). The full width at half maximum for that measurement differs only by 0.04 s or approximately by one image frame. The deviation concerning the averaged washin and washout can be explained by looking at the zoomed section, see box 1 in Figure 15(b). Both curves have two intensity peaks; for the real angiograms, the measured intensity is larger at the second peak, while the virtual angiogram reaches its maximum at the first peak. This deviation is 0.69 s, which causes the mentioned difference.

An additional effect, which can usually be avoided for medical phantoms, but which is a common observation for actual patient cases is the depicted rerise of intensity values towards the end of the sequence, see TICs in Figures 15(b) and 15(c), box 2. This effect is caused by contrast agent in arteries and veins which are in the background of the observed aneurysm, but do overlay with the aneurysm on the 2D DSA images and hence contribute to the opacification.

The rRMSE, which is calculated from the time-intensity curves cropped to arterial phase (Figure 16), is comparable for ROI0 and ROI1. For ROI2, the rRMSE is increased due to the observed remaining contrast agent inside the aneurysm.

3.3.2. Projection 2 (−91° RAO/LAO, −0.2° CRAN/CAUD). The second projection provides a lateral view (rotated by 90° in left/right direction compared to projection 1) of the aneurysm, see Figure 17. After the washin phase, the contrast agent circulates in the left part of the aneurysm, both in the real and virtual angiograms (c). As was observed in the first projection, the contrast agent flushes out quite fast in the virtual angiogram, whereas in the real angiogram, a small portion remains within the aneurym, Figures 17(d)–17(f). From this angulation, it is clearly visible that the contrast agent is settling at the left side of the aneurysm (direction of gravity). This effect, although not as intense, has also been observed for the acquired angiogram of the phantom data,

compare Figures 11(k)–11(n). This settling of contrast agent is a known physiological effect [37–39], but the reason for that behavior has not finally been discovered yet.

For taking a closer look at this effect, we measured time-intensity curves for the regions of interest depicted in Figure 18 and derived quantitative values for these regions, see Table 4. For the complete aneurysm (ROI1), the time-intensity curve of the real angiogram, Figure 19(b), has a sustained wash-out phase, similar to the corresponding curve of the previous projection, compare Figure 15(d).

In Figures 19(c) and 19(d), this settling effect is further investigated. The region of interest covering the complete aneurysm dome (ROI1) is partitioned into two distinct regions capturing the areas with (ROI2) and without (ROI3) settled contrast agent. This partitioning is illustrated in Figure 18. The corresponding time-intensity curves, as depicted in Figures 19(c) and 19(d), are normalized with respect to the normalization factors of the time-intensity curves in Figure 19(b), meaning that the addition of the corresponding curves in Figures 19(c) and 19(d) results in the curves of Figure 19(b). As can been seen in the two figures, we were able to separate the settling effect. The trailing edge of the real curve in Figure 19(b) is almost identical to the trailing edge for the curve in Figure 19(c) in its shape and height. In comparison, the real curve in Figure 19(d) has no extended outflow phase and fits quite well to the virtual curve. The TICs cropped to arterial phase, which are used for calculating the rRMSE, are provided in Figure 20.

Consequently, the principle behavior of contrast agent within the aneurysm is captured sufficiently. However, additional effects such as this mixing/settling behaviour occur in real environments and are not covered by our model yet.

3.4. Comparison of Real and Virtual Angiograms—Patient B. Patient B has a medium-sized aneurysm at the supraophthalmic internal carotid artery. One acquired angiogram is available (C-arm angulation 16.4°, 16.1°), which has a duration of 6.7 s. Contrast agent was injected for only a small time period, resulting in a short measured injection bolus. For the region of interest at a short distance behind the inlet (ROI0, Figure 22), the FWHM of the measured

time-intensity curve is 0.59 s, as given in Table 5—there was no possibility to extract information about heart rate and heart state, see Figure 23. Therefore, a heart rate of 69 bpm was assumed, and the CFD velocity inflow profile could not be synchronized with the acquired angiogram. The selected inflow velocities range from 0.44 to 0.6 m/s, the average velocity is 0.5 m/s. For the simulation, the parameters were chosen as for the other patient case (10^6 particles, $\sigma = 1$, $\delta t = 0.002$ s), with a diffusivity coefficient $D = 1.0$.

Figure 21 shows the real and virtual angiograms. The global flow patterns of both angiograms match, as can be seen from the inflow jet entering the aneurysm, compare Figure 21(b) for the real and Figure 21(c) for the virtual angiogram, and the small amount of contrast agent in the center of the aneurysm before it is washed out, see Figure 21(f). Figures 21(a)–21(c) reveal that the filling of the parent vessel and of the aneurysm itself is delayed for the virtual angiogram compared to the real angiogram, which presumably indicates an underestimation of inflow velocities for the CFD simulation.

This observation can also be retrieved from the time-intensity curves. From ROI0 to ROI2, Figures 23(a)–23(c), a shift between both curves is observable, which is progressively growing with increasing distance from the injection spot. The intensities of the real angiogram are thereby registered sooner than those of the virtual angiogram.

In general, further calculated quantities based on time-intensity curves cropped to arterial phase, Figure 24, show a concordance between real and virtual angiograms, meaning that the overall simulated behavior matches satisfactorily the *in vivo* behavior, see Table 5. However, the delay strongly affects the relative root mean square error, which is significantly larger than for the previous cases under consideration.

4. Conclusion

In this paper, we have proposed a workflow to generate patient-specific virtual angiograms based on CFD simulation results.

We have used discrete particles in order to simulate the transport of contrast agent and successfully demonstrated that this approach can properly model the behavior of contrast agent, although this is not the predominant approach used in related works [10, 30].

As a particular aspect, we have put particular emphasis on a patient-individualized generation of virtual angiograms, namely, the incorporation of the patient-individual heart rate as well as the treatment-specific bolus injection profile caused by manual contrast medium injection, for example. As a consequence of this approach, it is now possible to apply our virtual angiography method without the need of using contrast injection systems or taking care of specified contrast bolus profiles, which may improve the future usability of this application in clinical settings.

To obtain these parameters, we presented methods which solely used the acquired 2D DSA data. Satisfying results have been achieved with these methods to synchronize the heart rate, heart state, and contrast bolus injection of virtual and acquired angiograms. In general, the measured quantitative differences are approximately in the same range between real and virtual angiograms as given in [17]. Further improvements of the results may be possible due to more sophisticated methods for adapting the inflow velocities.

Finally, the overall correspondence offers the possibility to use the presented virtual angiography workflow as a tool towards the indirect validation of patient-individual CFD simulation results.

References

[1] C. D. Mathers, T. Boerma, and D. Ma Fat, "Global and regional causes of death," *British Medical Bulletin*, vol. 92, no. 1, pp. 7–32, 2009.

[2] H. R. Winn, J. A. Jane, J. Taylor, D. Kaiser, and A. G. W. Britz, "Prevalence of asymptomatic incidental aneurysms: review of 4568 arteriograms," *Journal of Neurosurgery*, vol. 96, no. 1, pp. 43–49, 2002.

[3] J. L. Brisman, J. K. Song, and D. W. Newell, "Cerebral aneurysms," *The New England Journal of Medicine*, vol. 355, no. 9, pp. 928–939, 2006.

[4] W. I. Schievink, "Intracranial aneurysms," *The New England Journal of Medicine*, vol. 336, no. 1, pp. 28–40, 1997.

[5] B. Seibert, R. P. Tummala, R. Chow, A. Faridar, S. A. Mousavi, and A. A. Divani, "Intracranial aneurysms: review of current treatment options and outcomes," *Frontiers in Neurology*, vol. 2, article 45, 2011.

[6] A. Oppelt, *Imaging Systems for Medical Diagnostics: Fundamentals, Technical Solutions, and Applications for Systems Applying Ionizing Radiation, Nuclear Magnetic Resonance, and Ultrasound*, Publicis Corporate Publishing, 2005.

[7] N. Strobel, O. Meissner, J. Boese et al., "3D imaging with flat-detector c-arm systems," in *Medical Radiology*, C. T. Multislice, M. F. Reiser, C. R. Becker, K. Nikolaou, and G. Glazer, Eds., pp. 33–51, Springer, Berlin, Germany, 2009.

[8] J. R. Cebral, M. Hernandez, and A. F. Frangi, "Computational analysis of blood flow dynamics in cerebral aneurysms from CTA and 3D rotational angiography image data," in *Proceedings of the International Congress on Computational Bioengineering*, vol. 1, pp. 191–198, 2003.

[9] D. A. Steinman, J. S. Milner, C. J. Norley, S. P. Lownie, and D. W. Holdsworth, "Image-based computational simulation of flow dynamics in a giant intracranial aneurysm," *American Journal of Neuroradiology*, vol. 24, no. 4, pp. 559–566, 2003.

[10] M. D. Ford, G. R. Stuhne, H. N. Nikolov et al., "Virtual angiography for visualization and validation of computational models of aneurysm hemodynamics," *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1586–1592, 2005.

[11] J. Durant, I. Waechter, R. Hermans, J. Weese, and T. Aach, "Toward quantitative virtual angiography: Evaluation with in vitro studies," in *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08)*, pp. 632–635, Paris, France, May 2008.

[12] Q. Sun, A. Groth, M. Bertram et al., "Phantom-based experimental validation of computational fluid dynamics simulations on cerebral aneurysms," *Medical Physics*, vol. 37, no. 9, pp. 5054–5065, 2010.

[13] J. R. Cebral, M. A. Castro, S. Appanaboyina, C. M. Putman, D. Millan, and A. F. Frangi, "Efficient pipeline for image-based patient-specific analysis of cerebral aneurysm hemodynamics:

technique and sensitivity," *IEEE Transactions on Medical Imaging*, vol. 24, no. 4, pp. 457–467, 2005.

[14] H. G. Morales, I. Larrabide, M. C. Villa-Uriol, A. J. Geers, and A. F. Frangi, "Towards the validation of a virtual coiling technique using a real versus a simulated bolus injection," in *Proceedings of the 2nd International Conference on Mathematical and Computational Biomedical Engineering (CMBE'11)*, pp. 512–515, 2011.

[15] J. R. Cebral, R. S. Pergolizzi Jr., and C. M. Putman, "Computational fluid dynamics modeling of intracranial aneurysms: qualitative comparison with cerebral angiography," *Academic Radiology*, vol. 14, no. 7, pp. 804–813, 2007.

[16] J. R. Cebral, A. Radaelli, A. F. Frangi, and C. M. Putman, "Qualitative comparison of intra-aneurysmal flow structures determined from conventional and virtual angiograms," in *Medical Imaging 2007: Physiology, Function, and Structure from Medical Images*, vol. 6511 of *Proceedings of SPIE*, pp. 65111E–1–65111E–9, San Diego, Calif, USA, February 2007.

[17] Q. Sun, A. Groth, and T. Aach, "Comprehensive validation of computational fluid dynamics simulations of in-vivo blood flow in patient-specific cerebral aneurysms," *Medical Physics*, vol. 39, no. 2, pp. 742–754, 2012.

[18] J. Endres, T. Redel, M. Kowarschik, J. Hutter, J. Hornegger, and A. Dörfler, "Virtual angiography using CFD simulations based on patient-specific parameter optimization," in *Proceedings of the 9th IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*, pp. 1200–1203, 2012.

[19] M. Sussman, P. Smereka, and S. Osher, "A level set approach for computing solutions to incompressible two-phase flow," *Journal of Computational Physics*, vol. 114, no. 1, pp. 146–159, 1994.

[20] V. Mihalef, R. Ionasec, P. Sharma et al., "Patient-specific modelling of whole heart anatomy, dynamics and haemodynamics from four-dimensional cardiac CT images," *Royal Society Interface Focus Journal*, vol. 1, no. 3, pp. 286–296, 2011.

[21] J. B. Bell, P. Colella, and H. M. Glaz, "A second-order projection method for the incompressible navier-stokes equations," *Journal of Computational Physics*, vol. 85, no. 2, pp. 257–283, 1989.

[22] J. Li, Y. Y. Renardy, and M. Renardy, "Numerical simulation of breakup of a viscous drop in simple shear flow through a volume-of-fluid method," *Physics of Fluids*, vol. 12, no. 2, pp. 269–282, 2000.

[23] V. Mihalef, P. Sharma, A. Kamen, and T. Redel, "An immersed porous boundary method for computational fluid dynamics of blood flow in aneurysms with flow diverters," in *Proceedings of the ASME Summer Bioengineering Conference*, 2012.

[24] T. D. Aslam, "A level-set algorithm for tracking discontinuities in hyperbolic conservation laws: I. Scalar equations," *Journal of Computational Physics*, vol. 167, no. 2, pp. 413–438, 2001.

[25] R. Courant, K. Friedrichs, and H. Lewy, "On the partial difference equations of mathematical physics," *IBM Journal of Research and Development*, vol. 11, no. 2, pp. 215–234, 1967.

[26] S. Larsson and V. Thomée, *Partial Differential Equations with Numerical Methods*, Texts in Applied Mathematics, Springer, Berlin, Germany, 2003.

[27] G. Y. Suh, A. S. Les, A. S. Tenforde et al., "Quantification of particle residence time in abdominal aortic aneurysms using magnetic resonance imaging and computational fluid dynamics," *Annals of Biomedical Engineering*, vol. 39, no. 2, pp. 864–883, 2011.

[28] J. A. Schetz and A. E. Fuhs, *Fundamentals of Fluid Mechanics*, John Wiley & Sons, New York, NY, USA, 1999.

[29] P. E. Danielsson, "Euclidean distance mapping," *Computer Graphics and Image Processing*, vol. 14, no. 3, pp. 227–248, 1980.

[30] Q. Sun, A. Groth, I. Waechter, O. Brina, J. Weese, and T. Aach, "Quantitative evaluation of virtual angiography for interventional X-ray acquisitions," in *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'09)*, pp. 895–898, Boston, Mass, USA, July 2009.

[31] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, Berlin, Germany, 2000.

[32] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, NY, USA, 2nd edition, 2004.

[33] N. Navab, A. Bani-Hashemi, M. Nadar et al., "3D reconstruction from projection matrices in a c-arm based 3D-angiography system," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, W. Wells, A. Colchester, and S. Delp, Eds., vol. 1496 of *Lecture Notes in Computer Science*, pp. 119–129, Springer, Berlin, Germany, 1998.

[34] A. Brost, N. Strobel, L. Yatziv et al., "Accuracy of X-ray image-based 3D localization from two C-arm views: a comparison between an ideal system and a real device," in *Medical Imaging 2009: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 7261 of *Proceedings of SPIE*, pp. 72611Z–1–72611Z–10, Lake Buena Vista, Fla, USA, February 2009.

[35] K. Engel, M. Hadwiger, J. M. Kniss, C. Rezk-Salama, and D. Weiskopf, *Real-Time Volume Graphics*, A. K. Peters, 2006.

[36] T. Struffert, S. Ott, M. Kowarschik, F. Bender, E. Adamek, and A. Dörfler, "Measurement of timecontrast curves to assess the hemodynamic effect of flow diverting devices: a feasibility study in an animal model," *Interventional Neuroradiology*, vol. 17, supplement 1, pp. 79–82, 2011.

[37] Z. J. Wang, K. R. Hoffmann, Z. Wang, S. Rudin, L. R. Guterman, and H. Meng, "Contrast settling in cerebral aneurysm angiography," *Physics in Medicine and Biology*, vol. 50, no. 13, pp. 3171–3181, 2005.

[38] L. D. Jou, A. Mohamed, D. H. Lee, and M. E. Mawad, "3D rotational digital subtraction angiography may underestimate intracranial aneurysms: findings from two basilar aneurysms," *American Journal of Neuroradiology*, vol. 28, no. 9, pp. 1690–1692, 2007.

[39] B. B. Lieber, C. Sadasivan, Q. Hao, J. Seong, and L. Cesar, "The mixability of angiographic contrast with arterial blood," *Medical Physics*, vol. 36, no. 11, pp. 5064–5078, 2009.

Research Article

Linear Program Relaxation of Sparse Nonnegative Recovery in Compressive Sensing Microarrays

Linxia Qin, Naihua Xiu, Lingchen Kong, and Yu Li

Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Linxia Qin, lxqin.echo@163.com

Compressive sensing microarrays (CSM) are DNA-based sensors that operate using group testing and compressive sensing principles. Mathematically, one can cast the CSM as sparse nonnegative recovery (SNR) which is to find the sparsest solutions subjected to an underdetermined system of linear equations and nonnegative restriction. In this paper, we discuss the l_1 relaxation of the SNR. By defining nonnegative restricted isometry/orthogonality constants, we give a nonnegative restricted property condition which guarantees that the SNR and the l_1 relaxation share the common unique solution. Besides, we show that any solution to the SNR must be one of the extreme points of the underlying feasible set.

1. Introduction

Nowadays, with the rapid development of molecular biology techniques, scientists use compressive sensing microarrays to collect the gene expression changes of patients suffer from specific diseases and test a lot of different drugs on cells genetically to look for medicine being able to change the abnormal gene expression [1, 2]. A DNA microarray is a collection of microscopic DNA spots attached to a solid surface. Each DNA spot contains a string of specific DNA sequences, known as probes. These can be a short section of a gene or other DNA element that are used to hybridize an organism's genetic sample under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of chemiluminescence-labeled targets to infer the genetic makeup in the test sample.

Although the number of DNA sequences is extremely large, not all agents are expected to be present in a significant concentration at a given time and location. In traditional microarrays, this results in many inactive probes during sensing. On the other hand, we are often interested in only a small quantity of certain harmful biological agents. Therefore, it is important to not just detect the presence of agents in a sample but also estimate the concentrations with which they are present.

Assume that there are m spots and n labeled targets, and we have far fewer spots than target agents such that $m \ll n$. Mathematically, one can represent the DNA concentration of each organism as an element in a vector $x \in \mathbb{R}^n$ and the measurements as $b \in \mathbb{R}^m$. For $1 \le i \le m$ and $1 \le j \le n$, the probe at spot i hybridizes to target j with probability a_{ij}. The target j occurs in the tested DNA sample with concentration x_j, which is clearly nonnegative. Denoting by $A := (a_{ij})_{m \times n}$, the process of DNA microarrays leads to the *sparse nonnegative recovery* (SNR) which is to find the sparsest solutions subjected to an underdetermined system of linear equations and nonnegative constraints, with the mathematical model as follows:

$$
\begin{aligned}
\min \ &\|x\|_0, \\
\text{s.t.} \ &Ax = b, \quad x \ge 0,
\end{aligned} \tag{P_0}
$$

where the variable vector $x \in \mathbb{R}^n$, $\|x\|_0$ denotes the number of the nonzero entries of x, $A \in \mathbb{R}^{m \times n}$ is the measurement matrix with full row rank, $m \ll n$, and $b \in \mathbb{R}^m$.

SNR can be regarded as a special case of the sparse recovery, which is related to program $\min\{\|x\|_0 \mid Ax = b\}$. This program has sparked the significant concern and rapid development in recent years [3–5] owing to its wide applications. However, with the nonnegativity prior information about the object to be recovered in various applications

such as CSM, solutions on (P_0) tend to be closer to the actual situations and lead to substantial improvements in the image reconstruction. Moreover, with the nonnegative constraints, the feasible set becomes a polyhedral set instead of an affine subspace. This will bring us essential hardness in projecting on the feasible set. Thus, (P_0) is more likely difficult to solve. Therefore, SNR deserves specific study.

Problem (P_0) has been shown to be NP-hard [6, 7] in general from the perspective of computational complexity. One popular approach is to reconstruct the vector via the l_1 relaxation, which refers to

$$\min \|x\|_1,$$
$$\text{s.t.} \quad Ax = b, \quad x \geq 0. \tag{P_1}$$

Since (P_1) is a standard linear program, it is easy to solve. An important issue is how to guarantee the equivalence of (P_0) and (P_1) in the sense that they have the same unique k-sparse solution under some conditions. Here, we call a vector x k-sparse if the number of its nonzero entries is no more than k. There has been some increasing interest and activity in this area; see, for example, [8–14]. Donoho and Tanner [9] firstly proposed that (P_0) and (P_1) share the common k-sparse unique solution if the polytope AT is outwardly k-neighborly, where T is the standard simplex in $\mathbb{R}^n$. Zhang [13] proved that (P_0) and (P_1) share the common k-sparse unique solution if the null space of A is strictly half k-balance. Juditsky et al. [11] developed several different necessary and sufficient conditions for the (P_0)-(P_1) equivalence in the case of general type sign restrictions, including the nonnegative constraints as its special case. When the feasible set of (P_0) is a singleton, the unknown can be recovered by optimizing any objective function over this constraint set, and (P_0) and (P_1) definitely get the same unique solution. In this case, Bruckstein et al. [8] got the uniqueness of the feasible solution under a sufficient condition that A has a rowspan intersecting the positive orthant. Furthermore, Wang et al. [14] proved that the above sufficient condition is also necessary to the uniqueness of the feasible solution. Donoho and Tanner [10] proved that the underlying feasible set is a singleton if and only if the polytope $A\mathbb{R}^n_+$ and $\mathbb{R}^n_+$ have the same number of k-faces. Khajehnejad et al. [12] gave another equivalent condition of the uniqueness property by characterizing the support size of vectors in the null space of A.

For the l_1 relaxation of sparse recovery, one of the most significant conditions is the restricted isometry property (RIP), named by Candès and Tao [15] with the ground-breaking work of Donoho et al. [16, 17]. However, to the best of our knowledge, the nonnegative case of RIP has not been investigated. This paper will deal with this issue. We begin with investigating the solution property of SNR and show that any solution to the SNR must be one extreme point of its feasible set in Section 2. We prove in Section 3 the nonemptiness and boundedness of the solution set of (P_1) and show that any solution of (P_1) could be stated as the convex combination of its optimal extreme points. In Section 4, by defining the nonnegative restricted isometry/orthogonality constants, we derive a sufficient

condition for exact recovery of the sparsest nonnegative image/signal via the linear program relaxation.

Now we give some notations used in the text. We use $\text{sol}(\cdot)$ and $v(\cdot)$ to denote the solution set and optimal value of problem $(\cdot)$. The $e_i \in \mathbb{R}^n$ would be the vector with only the ith entry 1 and the rest all 0. $e^l \in \mathbb{R}^l$ is the vector with each entry equal to 1; we also use e to demonstrate that $e^n \in \mathbb{R}^n$ for short. The $a_i \in \mathbb{R}^m$ for $i = 1, \ldots, n$ denote the column vectors of the matrix A and $A_I = (a_i)_{i \in I \subset \{1, \ldots, n\}}$. For any $x \in \mathbb{R}^n$, x_i is the ith component and $I(x)$ is the support set of x; that is, $I(x) = \{i \mid x_i \neq 0, i = 1, \ldots, n\}$. For any subset $T \subset \{1, \ldots, n\}$, T^c denotes the complement set of T out of $\{1, \ldots, n\}$.

2. Solution Property

Throughout the paper we assume that

$$\mathcal{S} := \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\} \neq \varnothing. \tag{1}$$

Apparently, $\mathcal{S}$ is a polyhedral set in $\mathbb{R}^n$. According to the representation theorem, any $x \in \mathcal{S}$ could be represented as follows:

$$x = \sum_{i=1}^{t} \lambda_i \bar{x}^{(i)} + \sum_{j=1}^{q} \mu_j d^{(j)}, \tag{2}$$

where $\sum_{i=1}^{t} \lambda_i = 1, \lambda_i \geq 0, i = 1, \ldots, t$, and $\{\bar{x}^{(i)} \in \mathbb{R}^n \mid i = 1, 2, \ldots, t\}$ are the extreme point set of $\mathcal{S}$; $\mu_j \geq 0, j = 1, \ldots, q$, and $\{d^{(j)} \in \mathbb{R}^n \mid j = 1, 2, \ldots, q\}$ are the extreme direction set of $\mathcal{S}$. Apparently, $A\bar{x}^{(i)} = b, i = 1, \ldots, t; Ad^{(j)} = 0, d^{(j)} \geq 0, j = 1, \ldots, q$.

Define subsets of $\mathbb{R}^n$ as follows:

$$S_0 = \text{sub}\{0\},$$

$$S_1 = \left\{ \bigcup_{i=1}^{n} \text{sub}\{e_i\} \right\} \setminus S_0,$$

$$S_2 = \left\{ \bigcup_{i_1, i_2 = 1}^{n} \text{sub}\{e_{i_1}, e_{i_2}\} \right\} \setminus \bigcup_{j=0}^{1} S_j,$$

$$\vdots \tag{3}$$

$$S_r = \left\{ \bigcup_{i_1, \ldots, i_r = 1}^{n} \text{sub}\{e_{i_1}, \ldots, e_{i_r}\} \right\} \setminus \bigcup_{j=0}^{r-1} S_j,$$

$$\vdots$$

$$S_n = \{\text{sub}\{e_1, \ldots, e_n\}\} \setminus \bigcup_{j=0}^{n-1} S_j,$$

where $\text{sub}\{e_{i_1}, \ldots, e_{i_k}\}$ denotes the subspace spanned by the vectors $e_{i_1}, \ldots, e_{i_k}, k = 1, \ldots, n$. Clearly, $\{S_1, S_2, \ldots, S_n\}$ forms a partition of $\mathbb{R}^n$; that is, $\bigcup_{j=1}^{n} S_j = \mathbb{R}^n$ and $S_i \bigcap S_j = \varnothing$, $i \neq j$. Moreover, $\|x\|_0 = r$ for any $x \in S_r$. Along with the nonemptiness of $\mathcal{S}$, it is easy to see that $\mathcal{S}$ must intersect one

of these sets, hence $\nu(P_0) = \min\{r \in \mathbb{R} \mid \mathcal{S} \cap S_r \neq \varnothing\}$ and $\mathrm{sol}(P_0) \neq \varnothing$. Furthermore, we have the following result for the optimal value of (P_0).

Lemma 1. *Assume that $\nu(P_0) = k$, one must have $k \leq m$.*

Proof. Suppose that the conclusion is not true; that is, there is $x^* \in \mathrm{sol}(P_0)$ with $\|x^*\|_0 = k > m$. Without loss of generality, let $x_i^* > 0$, $i = 1, \ldots, k$. We get

$$x_1^* a_1 + \cdots + x_k^* a_k = b. \tag{4}$$

Meanwhile, $\mathrm{rank}(A) = m < k$. Thus, $\{a_i \in \mathbb{R}^m \mid i = 1, \ldots, k\}$ must be linearly dependent; that is, there exist $d_1, \ldots, d_k$, not all zero, such that

$$d_1 a_1 + \cdots + d_k a_k = 0. \tag{5}$$

Assume that $d_1 > 0$. By denoting $d = (d_1, \ldots, d_k, 0, \ldots, 0)^T \in \mathbb{R}^n$, we get $Ad = 0$. Taking $\delta = \min\{x_i^*/d_i \mid d_i > 0, i = 1, \ldots, k\}$, it holds that

$$A(x^* - \delta d) = Ax^* - \delta Ad = b,$$
$$x^* - \delta d \geq 0, \tag{6}$$
$$\|x^* - \delta d\|_0 \leq k - 1.$$

This is a contradiction with $x^* \in \mathrm{sol}(P_0)$. We complete the proof. $\square$

To characterize property of the solution set $\mathrm{sol}(P_0)$, we need the next lemma. In particular, this brand new result will play a key role in proposing the sufficient condition of the uniqueness of $\mathrm{sol}(P_0)$ in Section 4.

Lemma 2. *Any two distinct solutions of (P_0) must have different support sets.*

Proof. Assume for contradiction that x^* and $\tilde{x}$ are two different solutions of (P_0), $x^* \neq \tilde{x}$. If $I(x^*) = I(\tilde{x})$, we have $x_i^* > 0$, $\tilde{x}_i > 0$, for all $i \in I(x^*)$. Set $\lambda = \min\{\min_{i \in I(x^*)}\{x_i^*/\tilde{x}_i\}, \min_{i \in I(x^*)}\{\tilde{x}_i/x_i^*\}\}$. Since $x^* \neq \tilde{x}$, it must hold $\lambda < 1$. When $\lambda = x_{i_0}^*/\tilde{x}_{i_0}$, take

$$\hat{x} = \frac{1}{1-\lambda} x^* - \frac{\lambda}{1-\lambda} \tilde{x}. \tag{7}$$

It is easy to see that

$$\hat{x}_i = \frac{1}{1-\lambda} x_i^* - \frac{\lambda}{1-\lambda} \tilde{x}_i \in \begin{cases} \{0\}, & i = i_0, \\ \mathbb{R}_+, & i \in I(x^*) \setminus \{i_0\}. \end{cases} \tag{8}$$

Thus, we have $A\hat{x} = b$, $\hat{x} \geq 0$ and $\|\hat{x}\|_0 \leq \|x^*\|_0 - 1$. This is a contradiction with the optimality of x^*. When $\lambda = \tilde{x}_{j_0}/x_{j_0}^*$, just taking $\hat{x} = (1/(1-\lambda))\tilde{x} - (\lambda/(1-\lambda))x^*$ instead, we get the contradiction by a similar way. The proof is completed. $\square$

Now we are in a position to give the main theorem in this section.

Theorem 3. *Any solution of (P_0) must be one of the extreme points of $\mathcal{S}$.*

Proof. Given any solution $x^* \in \mathrm{sol}(P_0)$ with representation

$$x^* = \sum_{i=1}^{t} \lambda_i^* \overline{x}^{(i)} + \sum_{j=1}^{q} \mu_j^* d^{(j)}, \tag{9}$$

where $\sum_{i=1}^{t} \lambda_i^* = 1$, $\lambda_i^* \geq 0$, $\overline{x}^{(i)} \geq 0$, $A\overline{x}^{(i)} = b$, for all $i = 1, \ldots, t$; $\mu_j^* \geq 0$, $d^{(j)} \geq 0$, $Ad^{(j)} = 0$, for all $j = 1, \ldots, q$. We only need to prove that

$$\sum_{i=1}^{p} \lambda_i^* = 1, \quad \lambda_i^* \in \{0, 1\}, \quad i = 1, \ldots, t,$$
$$\mu_j^* = 0, \quad j = 1, \ldots, q, \tag{10}$$

in (9). To this end, we have the following three steps.

Firstly, we claim that in (9),

$$I\left(\overline{x}^{(i)}\right) = I(x^*), \quad i \in \{i \mid \lambda_i^* > 0, \ i = 1, \ldots, t\},$$
$$I\left(d^{(j)}\right) \subset I(x^*), \quad j \in \left\{j \mid \mu_j^* > 0, \ j = 1, \ldots, q\right\}. \tag{11}$$

According to the fact that $\overline{x}^{(i)} \geq 0$, $i = 1, \ldots, t$, and $\mu_j^* \geq 0$, $d^{(j)} \geq 0$, $j = 1, \ldots, q$, one has for any $l \in I(\overline{x}^{(i)})$, $i \in \{i \mid \lambda_i^* > 0, i = 1, 2, \ldots, t\}$,

$$x_l^* = \sum_{i=1}^{t} \lambda_i^* \overline{x}_l^{(i)} + \sum_{j=1}^{q} \mu_j^* d_l^{(j)} > 0, \tag{12}$$

which means that $l \in I(x^*)$. This implies $I(\overline{x}^{(i)}) \subset I(x^*)$. Similarly, we get $I(d^{(j)}) \subset I(x^*)$. On the contrary, from the optimality of x^* and (9), we know that $I(x^*) \subset I(\overline{x}^{(i)})$.

Secondly, we will show that $\mu_j^* = 0$, $j = 1, \ldots, q$. If this is not true, there is an index j_0 such that $\mu_{j_0}^* > 0$, then $\mu_{j_0}^* d^{(j_0)}$ has at least one positive component. Denote $d = \sum_{j=1}^{q} \mu_j^* d^{(j)}$, so $Ad = 0$ and $\{d_l > 0 \mid l = 1, \ldots, n\} \neq \varnothing$. Noting that $\overline{x}^{(i)} \geq 0$, (11) implies the fact that $I(\sum_{i=1}^{t} \lambda_i^* \overline{x}^{(i)}) = I(x^*), I(d) \subset I(x^*)$. Take $\delta = \min\{(\sum_{i=1}^{t} \lambda_i^* \overline{x}^{(i)})_l/d_l \mid d_l > 0, l = 1, \ldots, n\}$, and

$$\tilde{x} = \sum_{i=1}^{t} \lambda_i^* \overline{x}^{(i)} - \delta d. \tag{13}$$

Without loss of generality, set $\delta = (\sum_{i=1}^{t} \lambda_i^* \overline{x}^{(i)})_{l_0}/d_{l_0}$. It is easy to verify that

$$\tilde{x}_j = \left(\sum_{i=1}^{t} \lambda_i^* \overline{x}^{(i)}\right)_j - \delta d_j \in \begin{cases} \{0\}, & j = l_0 \\ \mathbb{R}_+, & i \in I(x^*) \setminus \{l_0\}, \end{cases} \tag{14}$$

hence

$$A\tilde{x} = b,$$
$$\tilde{x} \geq 0, \tag{15}$$
$$\|\tilde{x}\|_0 \leq \|x^*\|_0 - 1,$$

which is a contradiction with the optimality of x^*.

Thirdly, we will prove that $\sum_{i=1}^{t} \lambda_i^* = 1$, $\lambda_i^* \in \{0, 1\}$, $i = 1, \ldots, t$. Suppose that there exist $\lambda_1^* > 0$, $\lambda_2^* > 0$, and $\lambda_1^* + \lambda_2^* = 1$ and two different extreme points of $\mathcal{S}$, say $\overline{x}^{(1)}, \overline{x}^{(2)}$, such that

$$x^* = \lambda_1^* \overline{x}^{(1)} + \lambda_2^* \overline{x}^{(2)}. \tag{16}$$

Based on (11), we have $I(\overline{x}^{(1)}) = I(\overline{x}^{(2)}) = I(x^*)$, hence $\overline{x}^{(1)} \in \text{sol}(P_0)$, $\overline{x}^{(2)} \in \text{sol}(P_0)$. Nevertheless, by Lemma 2, this is impossible. Hence, we show that

$$x^* \in \left\{ \overline{x}^{(1)}, \overline{x}^{(2)}, \ldots, \overline{x}^{(t)} \right\}. \tag{17}$$

We complete the argument. □

Theorem 3 tells us that each solution of (P_0) lies in the extreme point set of $\mathcal{S}$. Here is a concrete example.

Example 4. Let $A = \left(\begin{smallmatrix} 1 & 1 & -3 \\ -1 & 0 & 2 \end{smallmatrix} \right)$, $b = \left(\begin{smallmatrix} 3 \\ -2 \end{smallmatrix} \right)$. Obviously, the solution set and optimal value of (P_0) are

$$\text{sol}(P_0) = \{(2, 1, 0)\}, \qquad v(P_0) = 2, \tag{18}$$

respectively. In this case, $(2, 1, 0)$ is the only extreme point of $\mathcal{S}$. While the solution set and optimal value of $(P) := \min\{\|x\|_0 \mid Ax = b\}$ are

$$\text{sol}(P) = \{(0, 0, -1)\}, \qquad v(P) = 1, \tag{19}$$

respectively.

At the end of this section, we consider the l_p $(0 < p < 1)$ relaxation of (P_0)

$$
\begin{aligned}
\min \quad & \|x\|_p^p = \sum_{i=1}^{n} x_i^p, \\
\text{s.t.} \quad & Ax = b, \\
& x \geq 0.
\end{aligned}
\tag{P_p}
$$

Clearly, (P_p) is a concave relaxation of (P_0). For the program (P_p), Ge et al. [7] derived the useful result as the following.

Lemma 5. *The set of all extreme points of $\mathcal{S}$ is exactly the set of all local minimizers to (P_p).*

This lemma implies that any global solution of l_p relaxation must be one of its extreme points. From Theorem 3 and Lemma 5, we immediately draw a new proposition.

Proposition 6. *For any $p \in (0, 1)$, there exists an extreme point of $\mathcal{S}$ that is both an exact solution of (P_0) and a local minimizer of (P_p).*

This is different from the result of Fung and Mangasarian in [18], where they showed that for sufficiently small $\overline{p} \in (0, 1)$, there exists an extreme point $\overline{x}$ of the polyhedral set $\mathcal{T}$, obtained by lifting the set $\mathcal{S}$, such that $\overline{x}$ is an exact solution of (P_0) and a global solution of the $l_{\overline{p}}$ relaxation.

3. Linear Program Relaxation

Consider the linear program relaxation (P_1). Since the linear objective function $\langle e, x \rangle$ is bounded below over the feasible set, based on the Frank-Wolfe theorem, the minimum of (P_1) is attainable. Among all the extreme points of $\mathcal{S}$, $\{\overline{x}^{(i)} \in \mathbb{R}^n \mid i = 1, \ldots, t\}$, we call $\overline{x}^{(i)}$ an optimal extreme point if it also meets $\langle e, \overline{x}^{(i)} \rangle = v(P_1)$.

Proposition 7. *Any $x^* \in \text{sol}(P_1)$ could be stated as the convex combination of optimal extreme points of (P_1). Hence, sol (P_1) is bounded.*

Proof. Given any $x^* \in \text{sol}(P_0)$ with representation (9). If there is $i_0 \in \{1, \ldots, t\}$ such that $\overline{x}^{(i_0)}$ is not an optimal extreme point of (P_1) and $\lambda_{i_0}^* > 0$, we have $\langle e, \overline{x}^{(i_0)} \rangle > v(P_1)$, hence

$$
\begin{aligned}
v(P_1) &= \langle e, x^* \rangle \\
&= \sum_{i=1}^{t} \lambda_i^* \left\langle e, \overline{x}^{(i)} \right\rangle + \sum_{j=1}^{q} \mu_j^* \left\langle e, d^{(j)} \right\rangle \\
&\geq \sum_{i=1}^{t} \lambda_i^* \left\langle e, \overline{x}^{(i)} \right\rangle \\
&> \sum_{i=1}^{t} \lambda_i^* v(P_1) = v(P_1),
\end{aligned}
\tag{20}
$$

which is a contradiction.

Similarly, if there is $j_0 \in \{1, \ldots, q\}$ such that $\mu_{j_0}^* > 0$, we have $\mu_{j_0}^* \langle e, d^{(j_0)} \rangle > 0$, hence

$$
\begin{aligned}
v(P_1) &= \langle e, x^* \rangle \\
&= \sum_{i=1}^{t} \lambda_i^* \left\langle e, \overline{x}^{(i)} \right\rangle + \sum_{j=1}^{q} \mu_j^* \left\langle e, d^{(j)} \right\rangle \\
&> \sum_{i=1}^{t} \lambda_i^* \left\langle e, \overline{x}^{(i)} \right\rangle \\
&\geq \sum_{i=1}^{t} \lambda_i^* v(P_1) = v(P_1),
\end{aligned}
\tag{21}
$$

which is a contradiction. This completes the proof. □

From the above proposition, we know that linear program (P_1) has at least one optimal extreme point. Thus, we could use simplex method or interior point method to solve (P_1).

4. Nonnegative Restricted Property

In the framework of l_1 relaxation, a significant problem is how to guarantee the exact recovery of sparse image/signal via the l_1 relaxation. One of the most important qualifications is the restricted isometry property; see [15]. Recall that

the *k-restricted isometry constants* (RIC) δ_k is the smallest scalar satisfying

$$(1 - \delta_k)\|x\|_2^2 \le \|Ax\|_2^2 \le (1 + \delta_k)\|x\|_2^2, \quad \forall \|x\|_0 \le k. \tag{22}$$

Similarly, the k, k'-*restricted orthogonality constants* (ROC) $\theta_{k,k'}$ for $k + k' \le n$ is defined as the smallest scalar satisfying

$$|\langle Ax, Ay \rangle| \le \theta_{k,k'}\|x\|_2\|y\|_2, \tag{23}$$
$$\forall \|x\|_0 \le k; \ \forall \|y\|_0 \le k',$$

where x and y have disjoint support sets. The RIC δ_k and ROC $\theta_{k,k'}$ measure how close each submatrix of A with certain cardinality is behaving like an orthonormal system. Under some restricted isometry property, one can get the sparse recovery via its l_1 relaxation. Nevertheless, for the nonnegative case, the sparse recovery may maintain new characterizations. Above all, we define NRIC and NROC.

Definition 8. Let $A \in \mathbb{R}^{m \times n}$. We define the nonnegative k-restricted isometry constants (NRIC) δ_k^+ as the smallest number satisfying

$$\left(1 - \delta_k^+\right)\|x\|_2^2 \le \|Ax\|_2^2 \le \left(1 + \delta_k^+\right)\|x\|_2^2, \tag{24}$$
$$\forall x \ge 0, \ \|x\|_0 \le k.$$

Similarly, we define the nonnegative k, k'-restricted orthogonality constants (NROC) $\theta_{k,k'}^+$ for $k + k' \le n$ as the smallest number satisfying

$$|\langle Ax, Ay \rangle| \le \theta_{k,k'}^+\|x\|_2\|y\|_2, \tag{25}$$
$$\forall x \ge 0, \|x\|_0 \le k, \forall y \ge 0, \|y\|_0 \le k',$$

with $I(x)$ and $I(y)$ being disjoint sets.

Clearly,

$$\delta_k^+ \le \delta_k, \quad \theta_{k,k'}^+ \le \theta_{k,k'}. \tag{26}$$

Moreover, the numbers δ_k^+ and $\theta_{k,k'}^+$ are nondecreasing in k, k'.

By employing the projections of vectors in the null space of A to $\mathbb{R}_+^n$, we now provide a sufficient condition to determine a solution of (P_0).

Theorem 9. *Suppose that $k \ge 1$ is such that $\delta_k^+ + \theta_{k,k-1}^+ < 1$ and $x^* \in \mathcal{S}$ with $\|x^*\|_0 = k$. Then, x^* is a solution of (P_0).*

Proof. We complete the proof by contradiction. If this is not true, there exists $\bar{x}$ such that $A\bar{x} = b, \bar{x} \ge 0, \|\bar{x}\|_0 \le k - 1$. Set $h = x^* - \bar{x}$. Clearly, $Ah = 0$. Take $h = h^+ - h^-$ with h^+ and h^- being the projections of h and $-h$ to $\mathbb{R}_+^n$, respectively. We have $h^+ \ge 0, h^- \ge 0, \langle h^+, h^- \rangle = 0$. In particular,

$$(h^+)_i = \begin{cases} (x_i^* - \bar{x}_i)_+, & i \in I(x^*), \\ 0, & \text{else}, \end{cases}$$
$$\tag{27}$$
$$(h^-)_i = \begin{cases} (x_i^* - \bar{x}_i)_-, & i \in I(\bar{x}), \\ 0, & \text{else}. \end{cases}$$

Therefore,

$$1 \le \|h\|_0 \le \|x^*\|_0 + \|\bar{x}\|_0 \le 2k - 1,$$
$$1 \le \|h^+\|_0 \le k, \tag{28}$$
$$0 \le \|h^-\|_0 \le k - 1.$$

Thus, we get

$$0 = \|Ah\|_2^2$$
$$= \|Ah^+ - Ah^-\|_2^2$$
$$= \|Ah^+\|_2^2 + \|Ah^-\|_2^2 - 2\langle Ah^+, Ah^- \rangle$$
$$\ge \left(1 - \delta_k^+\right)\|h^+\|_2^2 + \left(1 - \delta_{k-1}^+\right)\|h^-\|_2^2$$
$$\quad - \theta_{k,k-1}^+\left(\|h^+\|_2^2 + \|h^-\|_2^2\right) \tag{29}$$
$$\ge \left(1 - \delta_k^+ - \theta_{k,k-1}^+\right)\|h^+\|_2^2$$
$$\quad + \left(1 - \delta_{k-1}^+ - \theta_{k,k-1}^+\right)\|h^-\|_2^2$$
$$> 0,$$

in which the first inequality is due to (24), (25), and the fact that $2ab \le a^2 + b^2$, and the last inequality is because of the assumption of $\delta_k^+ + \theta_{k,k-1}^+ < 1$ and the monotonicity of δ_k^+ in k. This is a contradiction. Therefore, $x^* \in \text{sol}(P_0)$. $\square$

With the special result that any two solutions of (P_0) have different support sets, we next derive a sufficient condition on the uniqueness of solution to (P_0).

Theorem 10. *Suppose that $k \ge 1$ is such that $\delta_k^+ + \theta_{k,k}^+ < 1$ and $x^* \in \mathcal{S}$ with $\|x^*\|_0 = k$. Then, x^* is the unique solution of (P_0).*

Proof. Since $\delta_k^+ + \theta_{k,k}^+ < 1$ implies $\delta_k^+ + \theta_{k,k-1}^+ < 1$, we know that $x^* \in \text{sol}(P_0)$. Now we just need to verify that x^* is the unique solution of (P_0). Assume that this is not true; that is, there is another solution $\bar{x} \ne x^*$. According to Lemma 2, it must hold $I(\bar{x}) \ne I(x^*)$. Take $h = x^* - \bar{x}$. By the argument similar to that in the proof of Theorem 9, we get

$$2 \le \|h\|_0 \le \|x^*\|_0 + \|\bar{x}\|_0 \le 2k,$$
$$1 \le \|h^+\|_0 \le k, \tag{30}$$
$$1 \le \|h^-\|_0 \le k,$$

and the contradiction. We conclude the proof. $\square$

Now we are ready to give the main result of this paper, which is called the nonnegative restricted property.

Theorem 11. *Assume that $k \ge 1$ is such that*

$$\delta_{2k}^+ + \left(\sqrt{2} + 1\right)\theta_{k,2k}^+ < 1, \tag{31}$$

and $x^* \in \mathcal{S}$ with $\|x^*\|_0 = k$. Then x^* is exactly the common unique minimizer of (P_0) and (P_1).

Proof. Since (31) implies $\delta_k^+ + \theta_{k,k}^+ < 1$, $\mathrm{sol}(P_0) = \{x^*\}$ by Theorem 10.

Suppose that $\bar{x}$ is a solution of (P_1). Take $h = \bar{x} - x^*$. To get $\mathrm{sol}(P_1) = \{x^*\}$, it suffices to verify that $h = 0$. The proof includes three steps, the first two steps are parallel to that in [19], in the third step, we utilize the technique of projecting the null space of A on $\mathbb{R}_+^n$; for details, see (42) and the argument around it.

Firstly, we introduce a partition of $\{1,\ldots,n\}$. Let T_0 be the support set of x^*, T_1 the index set including the first k large components of $\bar{x}$ in T_0^c, T_2 the index set including the next k large components of $\bar{x}$ in $T_0^c \setminus T_1$, and so on. Thus,

$$
\begin{aligned}
|T_j| &= k, \quad j = 0,1,\ldots,\left[\frac{n}{k}\right]-1, \\
|T_j| &\le k, \quad j = \left[\frac{n}{k}\right].
\end{aligned}
\tag{32}
$$

Moreover, for any $j = 0,1,\ldots,[n/k]$, we define

$$
\left(h_{T_j}\right)_i = \begin{cases} h_i, & i \in T_j, \\ 0, & \text{else}, \end{cases}
\tag{33}
$$

which is exactly that for any $j = 1,\ldots,[n/k]$

$$
\left(h_{T_j}\right)_i = \begin{cases} \bar{x}_i - x_i^*, & i \in T_0, \\ \bar{x}_i, & i \in T_j, \\ 0, & \text{else}. \end{cases}
\tag{34}
$$

Therefore, $Ah = 0, h = \sum_{j=0}^{[n/k]} h_{T_j}$, and

$$
\begin{aligned}
h_{T_j} &\ge 0, \quad j = 1,\ldots,\left[\frac{n}{k}\right], \\
\left\|h_{T_j}\right\|_0 &\le k, \quad j = 0,1,\ldots,\left[\frac{n}{k}\right].
\end{aligned}
\tag{35}
$$

Next, we show that $\|h_{(T_0 \cup T_1)^c}\|_2$ is bounded by $\|h_{T_0}\|_1$. Note that for each $j \ge 2$,

$$
\left\|h_{T_j}\right\|_2 \le \sqrt{k}\left\|h_{T_j}\right\|_\infty \le \frac{1}{\sqrt{k}}\left\|h_{T_{j-1}}\right\|_1,
\tag{36}
$$

where the second inequality is because of the monotonicity of h_i on T_0^c, and

$$
\sum_{j=2}^{[n/k]} \left\|h_{T_j}\right\|_2 \le \frac{1}{\sqrt{k}} \sum_{j=1}^{[n/k]-1} \left\|h_{T_j}\right\|_1 \le \frac{1}{\sqrt{k}}\left\|h_{T_0^c}\right\|_1.
\tag{37}
$$

This gives

$$
\left\|h_{(T_0 \cup T_1)^c}\right\|_2 = \left\|\sum_{j=2}^{[n/k]} h_{T_j}\right\|_2 \le \sum_{j=2}^{[n/k]} \left\|h_{T_j}\right\|_2 \le \frac{1}{\sqrt{k}}\left\|h_{T_0^c}\right\|_1.
\tag{38}
$$

In fact,

$$
\begin{aligned}
\left\|x_{T_0}^*\right\|_1 &= \|x^*\|_1 \ge \|\bar{x}\|_1 = \|x^* + h\|_1 \\
&= \left\|x_{T_0}^* + h_{T_0}\right\|_1 + \left\|x_{T_0^c}^* + h_{T_0^c}\right\|_1 \\
&\ge \left\|x_{T_0}^*\right\|_1 - \|h_{T_0}\|_1 + \left\|h_{T_0^c}\right\|_1 - \left\|x_{T_0^c}^*\right\|_1 \\
&= \left\|x_{T_0}^*\right\|_1 - \|h_{T_0}\|_1 + \left\|h_{T_0^c}\right\|_1,
\end{aligned}
\tag{39}
$$

which implies

$$
\left\|h_{T_0^c}\right\|_1 \le \|h_{T_0}\|_1.
\tag{40}
$$

By applying (38) and (40), we have

$$
\left\|h_{(T_0 \cup T_1)^c}\right\|_2 \le \frac{1}{\sqrt{k}}\left\|h_{T_0^c}\right\|_1 \le \frac{1}{\sqrt{k}}\|h_{T_0}\|_1.
\tag{41}
$$

Finally, we show that $\|h_{T_0 \cup T_1}\|_2 = 0$. By utilizing the projection $h = h^+ - h^-$, where h^+ and h^- are projections of h and $-h$ on $\mathbb{R}_+^n$, we have $h_{(T_0 \cup T_1)} = h_{(T_0 \cup T_1)}^+ - h_{(T_0 \cup T_1)}^-$. Moreover,

$$
\left(h_{(T_0 \cup T_1)}^+\right)_i = \begin{cases} (\bar{x}_i - x_i^*)_+, & i \in T_0, \\ \bar{x}_i, & i \in T_1, \\ 0, & \text{else}, \end{cases}
$$
$$
\left(h_{(T_0 \cup T_1)}^-\right)_i = \begin{cases} (\bar{x}_i - x_i^*)_-, & i \in T_0, \\ 0, & i \in T_1, \\ 0, & \text{else}. \end{cases}
\tag{42}
$$

It is easy to see

$$
\begin{aligned}
\left\|h_{(T_0 \cup T_1)}^+\right\|_0 &\le 2k, \\
\left\|h_{(T_0 \cup T_1)}^-\right\|_0 &\le k.
\end{aligned}
\tag{43}
$$

From $Ah = 0$, we compute

$$
\begin{aligned}
\|Ah_{(T_0 \cup T_1)}\|_2^2 &= \left\langle Ah_{(T_0 \cup T_1)}, Ah - Ah_{(T_0 \cup T_1)^c} \right\rangle \\
&= -\sum_{j=2}^{[n/k]} \left\langle Ah_{(T_0 \cup T_1)}, Ah_{T_j} \right\rangle.
\end{aligned}
\tag{44}
$$

On one hand, based on the definition of NROC, (35), and (43), for $j \ge 2$,

$$
\begin{aligned}
&-\left\langle Ah_{(T_0 \cup T_1)}, Ah_{T_j} \right\rangle \\
&= -\left\langle Ah_{(T_0 \cup T_1)}^+, Ah_{T_j} \right\rangle + \left\langle Ah_{(T_0 \cup T_1)}^-, Ah_{T_j} \right\rangle \\
&\le \left|\left\langle Ah_{(T_0 \cup T_1)}^+, Ah_{T_j} \right\rangle\right| + \left|\left\langle Ah_{(T_0 \cup T_1)}^-, Ah_{T_j} \right\rangle\right| \\
&\le \theta_{k,2k}^+ \left\|h_{(T_0 \cup T_1)}^+\right\|_2 \cdot \left\|h_{T_j}\right\|_2 \\
&\quad + \theta_{k,k}^+ \left\|h_{(T_0 \cup T_1)}^-\right\|_2 \cdot \left\|h_{T_j}\right\|_2 \\
&\le \theta_{k,2k}^+ \left\|h_{T_j}\right\|_2 \cdot \left(\left\|h_{(T_0 \cup T_1)}^+\right\|_2 + \left\|h_{(T_0 \cup T_1)}^-\right\|_2\right) \\
&\le \sqrt{2}\theta_{k,2k}^+ \left\|h_{T_j}\right\|_2 \cdot \|h_{(T_0 \cup T_1)}\|_2,
\end{aligned}
\tag{45}
$$

where the last inequality is by the fact that $(a + b)^2 \leq 2(a^2 + b^2)$. On the other hand, together with the definition of NRIC and (43), one has

$$
\begin{aligned}
&\left\|Ah_{(T_0 \cup T_1)}\right\|_2^2 \\
&= \left\|Ah_{(T_0 \cup T_1)}^+ - Ah_{(T_0 \cup T_1)}^-\right\|_2^2 \\
&= \left\|Ah_{(T_0 \cup T_1)}^+\right\|_2^2 + \left\|Ah_{(T_0 \cup T_1)}^-\right\|_2^2 \\
&\quad - 2\left\langle Ah_{(T_0 \cup T_1)}^+, Ah_{(T_0 \cup T_1)}^- \right\rangle \\
&\geq \left(1 - \delta_{2k}^+\right)\left\|h_{(T_0 \cup T_1)}^+\right\|_2^2 + \left(1 - \delta_k^+\right)\left\|h_{(T_0 \cup T_1)}^-\right\|_2^2 \\
&\quad - 2\theta_{k,2k}^+\left\|h_{(T_0 \cup T_1)}^+\right\|_2 \cdot \left\|h_{(T_0 \cup T_1)}^-\right\|_2 \\
&\geq \left(1 - \delta_{2k}^+\right)\left\|h_{(T_0 \cup T_1)}^+\right\|_2^2 + \left(1 - \delta_k^+\right)\left\|h_{(T_0 \cup T_1)}^-\right\|_2^2 \\
&\quad - \theta_{k,2k}^+\left(\left\|h_{(T_0 \cup T_1)}^+\right\|_2^2 + \left\|h_{(T_0 \cup T_1)}^-\right\|_2^2\right) \\
&= \left(1 - \delta_{2k}^+ - \theta_{k,2k}^+\right)\left\|h_{(T_0 \cup T_1)}^+\right\|_2^2 \\
&\quad + \left(1 - \delta_k^+ - \theta_{k,2k}^+\right)\left\|h_{(T_0 \cup T_1)}^-\right\|_2^2 \\
&\geq \left(1 - \delta_{2k}^+ - \theta_{k,2k}^+\right)\|h_{(T_0 \cup T_1)}\|_2^2.
\end{aligned}
\tag{46}
$$

Therefore, we compute

$$
\begin{aligned}
&\left(1 - \delta_{2k}^+ - \theta_{k,2k}^+\right)\|h_{(T_0 \cup T_1)}\|_2^2 \\
&\leq \sqrt{2}\theta_{k,2k}^+\|h_{(T_0 \cup T_1)}\|_2 \\
&\quad \cdot \sum_{j=2}^{[n/k]} \left\|h_{T_j}\right\|_2 \quad \text{(by (44), (45), and (46))} \\
&\leq \sqrt{2}\theta_{k,2k}^+\|h_{(T_0 \cup T_1)}\|_2 \cdot \frac{1}{\sqrt{k}}\left\|h_{T_0^c}\right\|_1 \quad \text{(by (37))} \\
&\leq \sqrt{2}\theta_{k,2k}^+\|h_{(T_0 \cup T_1)}\|_2 \cdot \frac{1}{\sqrt{k}}\left\|h_{T_0}\right\|_1 \quad \text{(by (40))} \\
&\leq \sqrt{2}\theta_{k,2k}^+\|h_{(T_0 \cup T_1)}\|_2 \cdot \left\|h_{T_0}\right\|_2 \\
&\leq \sqrt{2}\theta_{k,2k}^+\|h_{(T_0 \cup T_1)}\|_2^2,
\end{aligned}
\tag{47}
$$

where the forth inequality is from the fact that $\|h_{T_0}\|_1^2 \leq k\|h_{T_0}\|_2^2$. Then the assumption $\delta_{2k}^+ + (\sqrt{2} + 1)\theta_{k,2k}^+ < 1$ forces

$$
\|h_{(T_0 \cup T_1)}\|_2^2 = 0.
\tag{48}
$$

Thus,

$$
h_{(T_0 \cup T_1)} = 0.
\tag{49}
$$

Therefore, we get $h_{(T_0 \cup T_1)^c} = 0$ by (41), hence $\|h\| = 0$. This is exactly what we want. We complete the proof. $\square$

5. Conclusion

In this paper, we have derived a nonnegative restricted property condition, which ensures the exact recovery of sparse nonnegative image/signal via the linear program relaxation. Since the NRIC and NROC are defined in $\mathbb{R}_+^n$, there may be more types of measurement matrices satisfying the nonnegative restricted property than that in the case of RIP, regardless of random matrices or deterministic matrices. As a byproduct of the main result, we have investigated the solution property of the sparse nonnegative recovery and shown that any solution of (P_0) must be one of the extreme points of its feasible set. However, it is not clear whether a given extreme point of the feasible set is a solution to (P_0). This can serve as a target for future work.

Acknowledgments

References

[1] W. Dai, M. A. Sheikh, O. Milenkovic, and R. G. Baraniuk, "Compressive sensing DNA microarrays," *Eurasip Journal on Bioinformatics and Systems Biology*, vol. 2009, Article ID 162824, 12 pages, 2009.

[2] M. A. Sheikh, S. Sarvotham, O. Milenkovic, and R. G. Baraniuk, "DNA array decoding from nonlinear measurements by belief propagation," in *Proceedings of the IEEE/SP 14th WorkShoP on Statistical Signal Processing (SSP '07)*, pp. 215–219, IEEE, Washington, DC, USA, August 2007.

[3] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[4] E. Candès, "Compressive sampling," in *Proceedings of the Congress of Mathematics*, vol. 3, pp. 1433–1452, Madrid, Spain, 2006.

[5] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[7] D. Ge, X. Jiang, and Y. Ye, "A note on the complexity of l_p minimization," *Mathematical Programming*, vol. 129, no. 2, pp. 285–299, 2011.

[8] A. M. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 4813–4820, 2008.

[9] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9446–9451, 2005.

[10] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete and Computational Geometry*, vol. 43, no. 3, pp. 522–541, 2010.

[11] A. Juditsky, F. KIlInç Karzan, and A. Nemirovski, "Verifiable conditions of ℓ 1-recovery for sparse signals with sign

restrictions," *Mathematical Programming B*, vol. 127, no. 1, pp. 89–122, 2011.

[12] M. A. Khajehnejad, A. G. Dimakis, W. Xu, and B. Hassibi, "Sparse recovery of nonnegative signals with minimal expansion," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 196–208, 2011.

[13] Y. Zhang, "A simple proof for the recoverability of l_1-minimization (II): the nonnegative case," Tech. Rep. TR05-10, Department of Computational and Applied Mathematical, Rice University, Houston, Tex, USA, 2005.

[14] M. Wang, W. Xu, and A. Tang, "A unique "nonnegative" solution to an underdetermined system: from vectors to matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1007–1016, 2011.

[15] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[16] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[17] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[18] G. M. Fung and O. L. Mangasarian, "Equivalence of minimal l_0- and l_p-norm solutions of linear equalities, inequalities and linear programs for sufficiently small p," *Journal of Optimization Theory and Applications*, vol. 151, no. 1, pp. 1–10, 2011.

[19] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

Research Article

Using the K-Nearest Neighbor Algorithm for the Classification of Lymph Node Metastasis in Gastric Cancer

Chao Li,[1] Shuheng Zhang,[1] Huan Zhang,[2] Lifang Pang,[2] Kinman Lam,[3] Chun Hui,[1] and Su Zhang[1]

[1] *School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China*
[2] *Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China*
[3] *Centre for Signal Processing, Department of Electronic and Information Engineering,*
  The Hong Kong Polytechnic University, Hong Kong

Correspondence should be addressed to Su Zhang, suzhang@sjtu.edu.cn

Accurate tumor, node, and metastasis (TNM) staging, especially N staging in gastric cancer or the metastasis on lymph node diagnosis, is a popular issue in clinical medical image analysis in which gemstone spectral imaging (GSI) can provide more information to doctors than conventional computed tomography (CT) does. In this paper, we apply machine learning methods on the GSI analysis of lymph node metastasis in gastric cancer. First, we use some feature selection or metric learning methods to reduce data dimension and feature space. We then employ the K-nearest neighbor classifier to distinguish lymph node metastasis from nonlymph node metastasis. The experiment involved 38 lymph node samples in gastric cancer, showing an overall accuracy of 96.33%. Compared with that of traditional diagnostic methods, such as helical CT (sensitivity 75.2% and specificity 41.8%) and multidetector computed tomography (82.09%), the diagnostic accuracy of lymph node metastasis is high. GSI-CT can then be the optimal choice for the preoperative diagnosis of patients with gastric cancer in the N staging.

1. Introduction

According to the global cancer statistics in 2011, an estimated 989,600 new stomach cancer cases and 738,000 deaths occurred in 2008, which account for 8% of the total cases and 10% of the total deaths. Over 70% of the new cases and deaths were recorded in developing countries [1, 2]. The most commonly used staging system is the American Joint Committee on Cancer Tumor, Node, and Metastasis (TNM) [3–5]. The two most important factors that influence survival among patients with resectable gastric cancer are the depth of cancer invasion from the gastric wall and the number of lymph nodes present. In areas not screened for gastric cancer, late diagnosis reveals a high frequency of nodal involvement. Even in early gastric cancer, the incidence of lymph node metastasis exceeds 10%. The overall incidence was reported to be 14.1% and 4.8% to 23.6% depending on cancer depth [6]. The lymph node status must be preoperatively evaluated for proper treatment. However, the

various modalities could not obtain sufficient results. The lymph node status is one of the most important prognostic indicators of poor survival [7, 8].

Preoperative examinations, endoscopy, and barium meal examinations are routinely used to evaluate cancerous lesions in the stomach. Abdominal ultrasound, computed tomography (CT) examination, and magnetic resonance imaging (MRI) are commonly used to examine the presence of invasion to other organs and metastatic lesions. However, their diagnostic accuracy is limited. Endoscopic ultrasound has been the most reliable nonsurgical method in the evaluation of the primary tumor with 65% to 77% accuracy of N staging due to the limited penetration ability of the ultrasound for lymph node distant metastasis. In spite of the higher image quality and dynamic contrast-enhanced imaging, MRI only has an N staging accuracy of 65% to 70%. The multidetector row computed tomography (MDCT) [9] scanner enables for thinner collimation and faster scanning, which markedly improves imaging resolution and enable rapid handling of

image reconstruction. Moreover, intravenous bolus administration of contrast material permits precise evaluation of carcinoma enhancement, and the water-filling method enables negative contrast to enhance the gastric wall. Thus, MDCT has a higher N staging accuracy of up to 82% and has become a main examination method for preoperative staging of gastric cancer [10]. Fukuya et al. [11] showed in their study for lymph nodes of at least 5 mm that sensitivity for detecting metastasis positive nodes was 75.2% and specificity for detecting metastasis negative nodes was 41.8%. A large-scale Chinese study [10] conducted by Ruijin Hospital showed that the overall diagnostic sensitivity, specificity, and accuracy of MDCT for determining lymph node metastasis was 86.26%, 76.17%, and 82.09%, respectively. However, with clinically valuable scanning protocols of the spectral CT imaging technology, we can obtain more information with gemstone spectral imaging (GSI) than with any conventional CT (e.g., MDCT).

In conventional CT imaging, we measure the attenuation of the X-ray beam through an object. We commonly define the X-ray beam quality in terms of its kilo voltage peak (kVp) that denotes the maximum photon energy, as the X-ray beam comprises a mixture of X-ray photon energies. GSI [12] with spectral CT, and conventional attenuation data may be transformed into effective material densities, that enhance the tissue characterization capabilities of CT. Furthermore, through the monochromatic representation of the spectral CT, the beam-hardening artifacts can be substantially reduced, which is a step toward quantitative imaging with more consistent image measurements for examinations, patients, and scanners.

In this paper, we intend to use the machine learning method to handle the large amount information provided by GSI and to improve the accuracy for the determination of lymph node metastasis in gastric cancer.

The paper is arranged as follows, Section 2 describes the details of the methods used in this paper, Section 3 presents the experimental framework and the results, and Section 4 concludes the present study and discusses potential future research.

2. Methodology

Figure 1 shows a flow chart illustrating the whole framework of the classification on lymph node metastasis in gastric cancer.

2.1. Pre-Processing. GSI-CT examination was performed among patients using the GE Discovery CT750 HD (GE-Healthcare) scanner [13]. Each patient received an intramuscular administration of 20 mg of anisodamine to decrease peristaltic bowel movement and drank 1,000 to 1,200 mL tap water for gastric filling 5 to 10 min before the scan. Patients were in a supine position. After obtaining the localizer CT radiographs (e.g., anterior-posterior and/or lateral), we captured the unenhanced scan of the upper abdomen and then employed the enhanced GSI scan in two phases. An 80 mL to 100 mL bolus of nonionic iodine contrast agent

was administered to the ante-cubital vein at a flow rate of 2 mL/sec to 3 mL/sec through a 20-gauge needle using an automatic injector. CT acquisitions were performed in the arterial phase (start delay of 40 s) and in the portal venous phase (start delay of 70 sec). The arterial phase scans the whole stomach and the portal venous phase examines from the top of the stomach diaphragm to the abdominal aortic bifurcation plane. The GSI-CT scanning parameters are as follows: scan mode of spectral imaging with fast tube-voltage switching between 80 kVp and 140 kVp, the currents of 220 mA to 640 mA, slice thickness of 5 mm, rotation speed of 1.6 s to 0.8 s, and pitch ratio of 0.984 : 1.

2.2. Feature Extraction. Lymph node regions of interest (ROIs) were delineated by experienced doctors. Not all the lymph nodes could be captured in the images because of the node size or location. Figure 2 shows lymph node and aortic in the arterial phase and venous phase under 70 keV monochromatic energy. The lymph node on Figure 2(b) is difficult to find for its small size. The monochromatic values (Hu) and the mean of material basis pairs (μg/cm^3) were calculated. The features used in this paper are monochromatic CT values (40 keV to 140 keV) and material basis pairs (Calcium-Iodine, Calcium-Water, Iodine-Calcium, Iodine-Water, Water-Calcium, Water-Iodine, Effective-Z).

During the image acquisition process, variations on the injection speed, dose of the contrast agents and their circulation inside the body of patients can cause differences in the CT numerical values. To eliminate discrepancies, the arterial CT value of the same slice was recorded at mean time, and then normalization work was conducted by using the following formula:

$$\text{Norm} = \frac{\text{ROI mean CT value}}{\text{Aortic mean CT value}}. \tag{1}$$

2.3. Feature Selection

2.3.1. mRMR Algorithm. Minimal redundancy maximal relevance (mRMR) is a feature-selection scheme proposed by [14] mRMR that uses the information theory as a standard with better generalization and efficiency and accuracy for feature selection. Each feature can be ranked based on its relevance to the target variable, and the ranking process considers the redundancy of these features. An effective feature is defined as one that has the best trade-off between minimum redundancy within the features and maximum relevance to the target variable [15]. Mutual information (MI), which measures the mutual dependence of two variables, is used to quantify both relevance and redundancy in this method [16]. The two most used mRMR criteria are mutual information difference (MID) and mutual information quotient (MIQ),

$$\text{MID} : \max_{i \in \Omega_s} \left[H(i,c) - \frac{1}{|S|} \sum_{j \in S} H(i,j) \right],$$

$$\text{MIQ} : \max_{i \in \Omega_s} \left\{ \frac{H(i,c)}{\left[(1/|S|) \sum_{j \in S} H(i,j) \right]} \right\}, \tag{2}$$
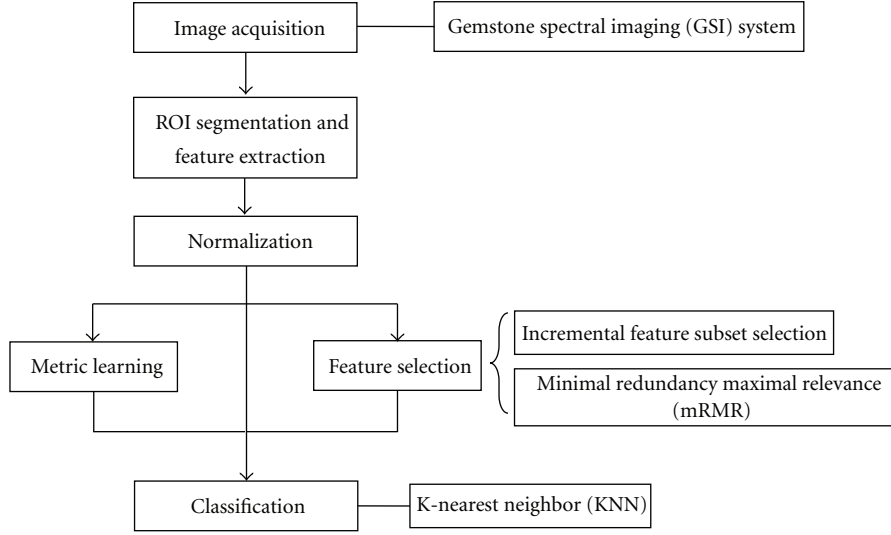
FIGURE 1: Flow chart classification on lymph node metastasis in gastric cancer.

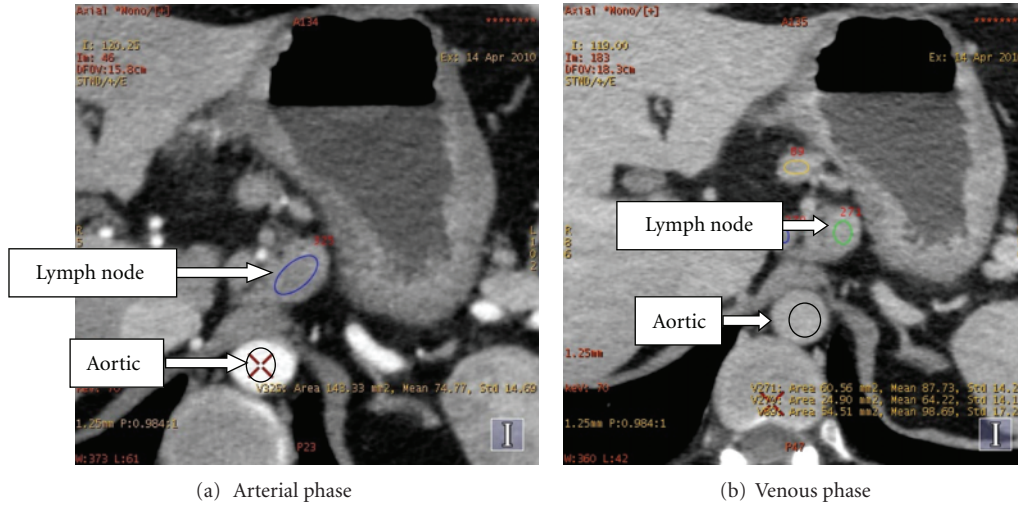

(a) Arterial phase

(b) Venous phase

FIGURE 2: Gastric lymph node at 70 keV energy.

where $H(i, c)$ is the MI between feature i and classification c, $H(i, j)$ is MI between features i and j, S is the current feature set, and $|S|$ is the length of the feature set.

2.3.2. SFS Algorithm. Sequential forward selection (SFS) is a traditional heuristic feature selection algorithm [17, 18]. SFS starts with an empty feature subset S_i. In each iteration only one feature is added to the feature subset. To determine which feature to add, the algorithm tentatively adds an unselected feature to the candidate feature subset and tests the accuracy of the classifier built on the tentative feature subset. The feature that exhibits the highest accuracy is finally added to the feature subset. The process stops after an iteration in which no features can be added, resulting in an improvement in accuracy.

2.4. Metric Learning Algorithm. Learning good distance metrics in feature space is crucial to many machine learning

works (e.g., classification). A lot of existing works has shown that properly designed distance metrics can greatly improve the KNN classification accuracy compared to the standard Euclidean distance. Depending on the feasibility of the training samples, distance metric learning algorithms can be divided into two categories: supervised distance metric learning and unsupervised distance metric learning. Table 1 shows the several distance metric learning algorithms. Among them, principal component analysis (PCA) is the most commonly used algorithm for the problem of dimensionality reduction of large datasets like in the application of face recognition [19], image retrieval [20].

2.5. Classification. The K-nearest neighbor (KNN) [21, 26] algorithm is among the simplest of all machine algorithms. In this algorithm, an object is classified by a majority vote of its neighbors. The object is consequently assigned to the class that is most common among its KNN, where K is a positive

TABLE 1: Distance metric learning methods used in this work.

Unsupervised distance metric learning method		Principal component analysis (PCA) [19, 21]
	Global	Fisher discriminative analysis (FDA) [21]
		Relevant component analysis (RCA) [22]
Supervised distance metric learning method		Neighborhood component analysis (NCA) [23]
	Local	Local fisher discriminative analysis (LFDA) [24]
		Large margin nearest neighborhood (LMNN) [25]

integer that is typically small. If $K = 1$, then the object is simply assigned to the class of its nearest neighbor.

The KNN algorithm is first implemented by introducing some notations $S = (x_i, y_i)$, $i = 1, 2, \ldots N$ is considered the training set, where x_i is the d-dimensional feature vector, and $y_i \in \{+1, -1\}$ is associated with the observed class labels. For simplicity, we consider a binary classification. We generally suppose that all training data are iid samples of random variables (X, Y) with unknown distribution.

With previously labeled samples as the training set S, the KNN algorithm constructs a local subregion $R(x) \subseteq \mathfrak{R}^d$ of the input space, which is situated at the estimation point x. The predicting region $R(x)$ contains the closest training points to x, which is written as follows:

$$R(x) = \{\hat{x} \mid D(x, \hat{x}) \leq d_{(k)}\}, \tag{3}$$

where $d_{(k)}$ is the kth order statistic of $\{D(x, \hat{x})\}_1^N$, and $D(x, \hat{x})$ is the distance metric. $k[y]$ denotes the number of samples in region $R(x)$, which is labeled y. The KNN algorithm is statistically designed for the estimation of posterior probability $p(y \mid x)$ of the observation point x:

$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x)} \cong \frac{k[y]}{k}. \tag{4}$$

For a given observation x, the decision $g(x)$ is formulated by evaluating the values of $k[y]$ and selecting the class that has the highest $k[y]$ value

$$g(x) = \begin{cases} 1, & k[y = 1] \geq k[y = -1], \\ -1, & k[y = -1] \geq k[y = 1]. \end{cases} \tag{5}$$

Thus, the decision that maximizes the associated posterior probability is employed in the KNN algorithm. For a binary classification problem in which $y_i \in \{+1, -1\}$, the KNN algorithm produces the following decision rule:

$$g(x) = \text{sgn}\left(\text{ave}_{x_i \in R(x)} y_i\right). \tag{6}$$

3. Experimental Results and Discussion

3.1. Experiments. The image data used in our work were acquired from GE Healthcare equipment in Ruijin Hospital on April 2010. We collected got 38 gastric lymph node datasets. Among the datasets were 27 lymph node metastasis (positive) and 11 nonlymph node metastasis (negative). All the lymph node data were pathology results obtained after lymph node dissection (lymphadenectomy) in patients.

3.1.1. Univariate Analysis. In this study, we conduct univariate analysis by exploring variables (features) one by one. We analyze each feature by calculating its relevance to lymph node metastasis. Here, we use the following measurements:

(i) *Two-Tailed t-test*: The two-tailed test is a statistical test used in inference, in which a given statistical hypothesis, H0 (the null hypothesis), is rejected when the value of the test statistic is either sufficiently small or sufficiently large.

(ii) *Point Biserial Correlation Coefficient* (r_{pb}):

$$r_{pb} = \frac{\text{Avg}_p - \text{Avg}_q}{\text{Std}_y} \sqrt{pq}. \tag{7}$$

In regard to the p, q notation formula Avg_p is the mean for nondichotomous values in connection with the variable coded 1, and Avg_q is the mean for the non-dichotomous values for the same variable-coded 0. Std_y is the standard deviation for all non-dichotomous entries, and p and q are the proportions of the dichotomous variable-coded 1 and 0, respectively.

(iii) *Information Gain (IG)*: IG is calculated by the entropy of the feature X, $H(X)$ minus the conditional entropy of Y given X, $H(X \mid Y)$

$$\text{IG}(X \mid Y) = H(X) - H(X \mid Y). \tag{8}$$

(iv) *Area Under Curve (AUC)*.

(v) *Symmetrical Uncertainty (SU)*: SU is the normalization of IG within $[0, 1]$, where the higher value of SU shows a higher relevance for feature X and class Y (as a measure of correlation between the features and the concept target)

$$\text{SU}(X, Y) = 2\left[\frac{IG(X \mid Y)}{H(X) + H(Y)}\right]. \tag{9}$$

The experimental results of the univariate analysis are shown in Tables 2 and 3. Based on the table, the Iodine-Water, Iodine-Calcium, Calcium-Iodine, and Effective-Z features show high relevance to lymph node metastasis. Among these features, high relevance to lymph node metastasis was clinically confirmed for Iodine-Water and Effective-Z features. Both Iodine-Water and Iodine-Calcium features reflect the concentration of the iodinated contrast media uptake by the surrounding tissue, and thus they are related

TABLE 2: Univariate analyses of the features of gastric lymph node metastasis arterial phase.

No.	Feature	Mean ± Standard		$P1$	$P2$	r_{pb}	AUC	SU	IG
		Negative	Positive						
1	40 keV	114.97 ± 29.84	177.79 ± 46.25	0.000	0.000	0.569	0.875	0.174	0.186
2	50 keV	85.55 ± 18.81	123.69 ± 30.44	0.000	0.000	0.540	0.869	0.174	0.186
3	60 keV	67.49 ± 13.36	90.63 ± 21.15	0.000	0.002	0.488	0.845	0.186	0.208
4	70 keV	56.74 ± 11.53	68.93 ± 15.63	0.000	0.025	0.362	0.774	0.106	0.104
5	80 keV	49.93 ± 11.53	54.84 ± 13.68	0.001	0.302	0.172	0.596	0.070	0.071
6	90 keV	45.30 ± 12.05	45.27 ± 13.55	0.025	0.994	−0.001	0.502	0.001	0.001
7	100 keV	42.01 ± 12.18	39.08 ± 13.46	0.114	0.537	−0.103	0.552	0.013	0.015
8	110 keV	39.71 ± 12.37	34.68 ± 13.54	0.272	0.295	−0.174	0.599	0.014	0.015
9	120 keV	38.13 ± 12.56	31.62 ± 13.68	0.434	0.182	−0.221	0.623	0.025	0.027
10	130 keV	36.83 ± 12.73	29.25 ± 13.86	0.570	0.127	−0.252	0.653	0.079	0.086
11	140 keV	35.89 ± 12.86	27.41 ± 14.00	0.673	0.092	−0.277	0.660	0.079	0.086
12	Effective-Z	8.18 ± 0.26	8.71 ± 0.35	0.000	0.000	0.601	0.896	0.317	0.336
13	Calcium-Iodine	819.69 ± 10.39	810.02 ± 10.70	0.284	0.015	−0.391	0.754	0.126	0.127
14	Calcium-Water	14.05 ± 5.77	27.26 ± 9.12	0.000	0.000	0.594	0.899	0.315	0.343
15	Iodine-Calcium	−579.62 ± 8.65	−568.30 ± 9.31	0.000	0.001	0.500	0.822	0.174	0.165
16	Iodine-Water	10.10 ± 4.11	19.20 ± 6.22	0.000	0.000	0.596	0.896	0.315	0.343
17	Water-Calcium	1021.57 ± 15.74	1000.17 ± 18.21	0.000	0.002	−0.494	0.818	0.174	0.165
18	Water-Iodine	1030.55 ± 13.74	1017.24 ± 15.20	0.291	0.017	−0.386	0.734	0.174	0.165

TABLE 3: Univariate analyses of the features of gastric lymph node metastasis venous phase.

No.	Feature	Mean ± Standard		$P1$	$P2$	r_{pb}	AUC	SU	IG
		Negative	Positive						
19	40 keV	168.56 ± 45.67	199.95 ± 51.33	0.000	0.087	0.282	0.684	0.070	0.072
20	50 keV	117.94 ± 29.61	137.13 ± 32.85	0.000	0.102	0.269	0.673	0.070	0.072
21	60 keV	86.91 ± 20.03	98.71 ± 22.14	0.000	0.135	0.247	0.653	0.086	0.092
22	70 keV	67.54 ± 13.52	73.94 ± 14.14	0.000	0.209	0.209	0.620	0.106	0.104
23	80 keV	55.09 ± 11.10	57.96 ± 11.31	0.000	0.481	0.118	0.559	0.110	0.122
24	90 keV	46.76 ± 10.95	47.02 ± 11.50	0.000	0.949	0.011	0.535	0.018	0.018
25	100 keV	41.08 ± 11.04	39.90 ± 12.06	0.000	0.781	−0.047	0.562	0.018	0.020
26	110 keV	37.08 ± 11.29	34.81 ± 12.73	0.003	0.611	−0.085	0.599	0.011	0.011
27	120 keV	34.25 ± 11.56	31.27 ± 13.30	0.012	0.521	−0.107	0.613	0.011	0.011
28	130 keV	32.10 ± 11.86	28.56 ± 13.79	0.028	0.461	−0.123	0.613	0.011	0.011
29	140 keV	30.37 ± 12.09	26.42 ± 14.19	0.052	0.423	−0.134	0.626	0.018	0.019
30	Effective-Z	8.61 ± 0.38	8.87 ± 0.42	0.000	0.081	0.286	0.680	0.087	0.088
31	Calcium-Iodine	812.48 ± 9.36	807.83 ± 11.88	0.651	0.254	−0.190	0.643	0.025	0.026
32	Calcium-Water	24.65 ± 9.57	31.44 ± 11.52	0.000	0.093	0.276	0.673	0.086	0.087
33	Iodine-Calcium	−570.78 ± 8.89	−565.23 ± 11.43	0.005	0.159	0.233	0.650	0.037	0.042
34	Iodine-Water	17.56 ± 6.43	22.25 ± 7.71	0.000	0.084	0.284	0.667	0.074	0.073
35	Water-Calcium	1005.60 ± 17.73	994.85 ± 22.22	0.011	0.162	−0.231	0.657	0.037	0.042
36	Water-Iodine	1021.14 ± 13.95	1014.62 ± 17.03	0.631	0.270	−0.184	0.636	0.011	0.011

to lymph node metastasis. The Calcium-Iodine feature indicates tissue calcification, which rarely exists in lymph nodes. However, experimental results show that the Calcium-Iodine feature is highly related to lymph node metastasis, which must be further verified by clinical results.

Based on the statistical results of r_{pb}, AUC, SU, and IG, compared with high monochromatic energy, low-energy features have higher relevance to lymph node metastasis according to clinical results. As shown in Figure 3, low-energy images display a large difference between lymph node metastasis (positive) and non-lymph node metastasis (negative), as monochromatic energy is associated with higher energies that yield less contrast between materials and more contrast with low energies. However, low-energy images

(a) Raw data in the arterial phase

(b) Raw data in the venous phase

(c) Normalized data in the arterial phase
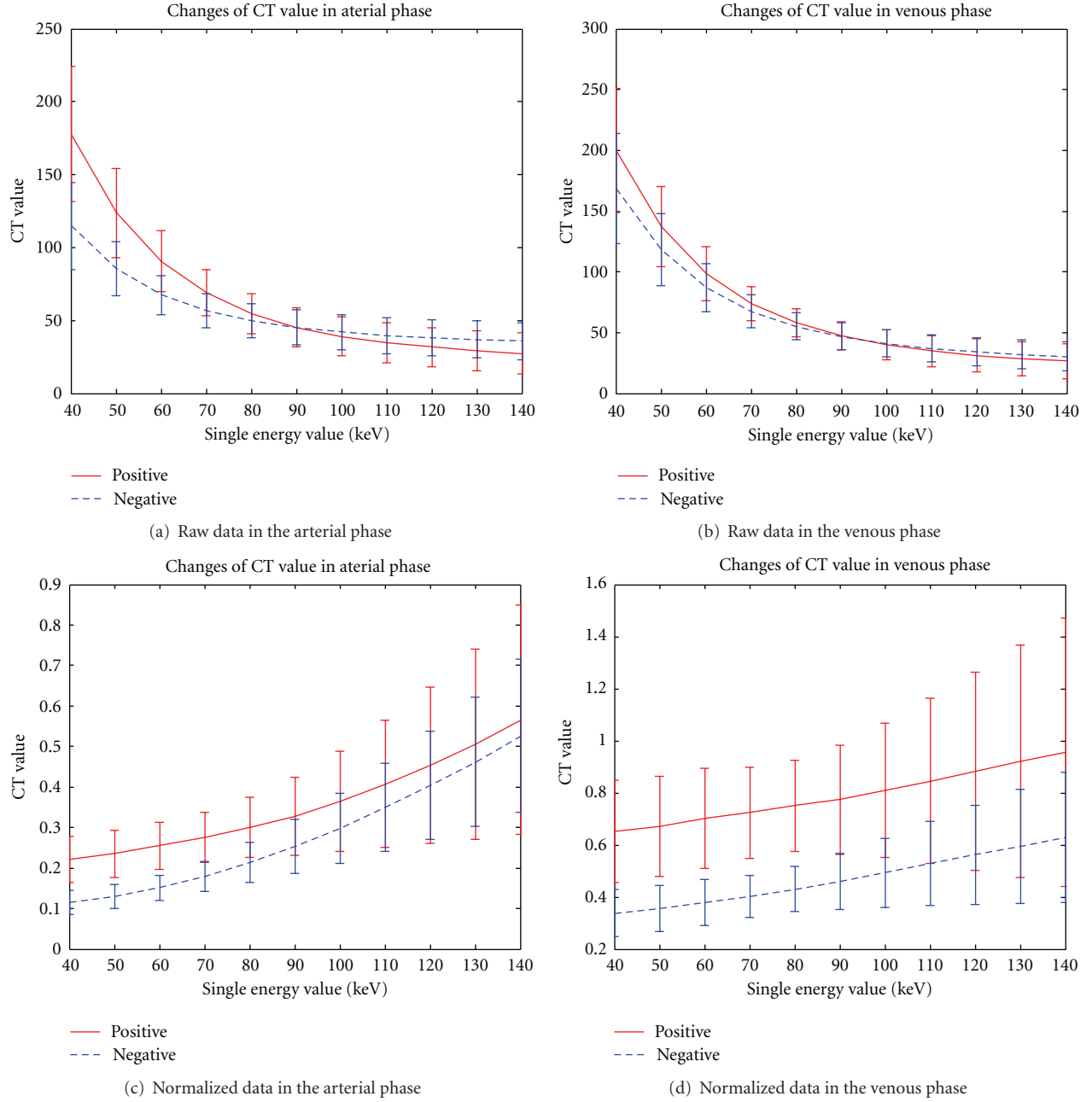
(d) Normalized data in the venous phase

FIGURE 3: Monochromatic energy CT value in the arterial and venous phases of gastric lymph node metastasis.

TABLE 4: Classification performance of the SFS-KNN algorithm with different neighborhood sizes.

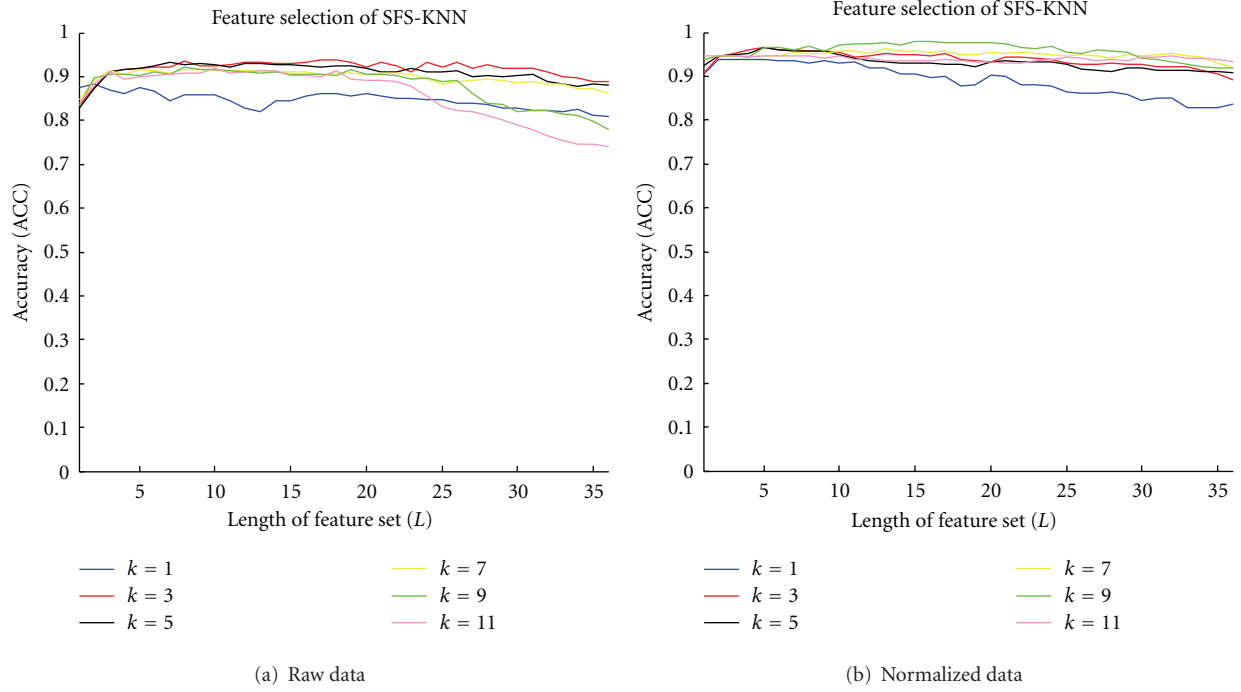|	Neighborhood size	$K = 1$	$K = 3$	$K = 5$	$K = 7$	$K = 9$
Prenorm	Selected features	14, 16	14, 31, 5, 15, 26, 4, 27, 21, 24, 9, 32, 2, 25, 8, 28, 3, 16	14, 31, 10, 36, 3, 25, 2	12, 31, 8, 29, 3, 15, 33, 1	12, 31, 23, 26, 3, 24, 30, 16
	Accuracy	88.29%	93.68%	93.29%	91.71%	92.24%
Norm	Selected features	12, 30	20, 15, 11, 30, 5	12, 30, 31, 33, 14	12, 19, 20, 30, 5, 18, 25, 17, 34, 3, 32, 15, 24	12, 19, 29, 30, 8, 34, 33, 25, 15, 6, 24, 7, 10, 20, 17
	Accuracy	93.95%	96.45%	96.58%	96.18%	97.89%

(a) Raw data

(b) Normalized data

FIGURE 4: SFS-KNN feature selection procedures on raw and normalized data.

TABLE 5: Classification performance of mRMR-KNN (MIQ) with different neighborhood sizes.

	Neighborhood size	$K = 1$	$K = 3$	$K = 5$	$K = 7$	$K = 9$
Prenorm	Sequence	14, 19, 5, 17, 23, 12, 3, 16, 18, 22, 1, 15, 4, 2, 30, 13, 21, 32, 10, 33, 11, 34, 20, 35, 31, 25, 9, 29, 24, 8, 7, 26, 36, 27, 28, 6				
	Length	1	28	28	35	1
	Accuracy	87.50%	89.74%	89.08%	87.24%	81.71%
Norm	Sequence	15, 21, 3, 30, 17, 24, 12, 14, 23, 5, 16, 22, 2, 18, 27, 1, 20, 4, 33, 25, 13, 19, 6, 28, 35, 26, 32, 7, 29, 34, 8, 31, 9, 11, 10, 36				
	Length	4	2	2	2	10
	Accuracy	90.00%	94.87%	94.87%	94.74%	95.66%

bring more noise with higher contrast. Therefore, doctors usually select 70 keV as a tradeoff for clinical diagnosis.

3.1.2. SFS-KNN Results. Figure 4 and Table 4 present the classification accuracy (ACC) of the KNN algorithm with different neighborhood sizes and the SFS algorithm with increasing lengths of the feature set. ACC first increases with the increasing length of the feature set, and then decreases. After application of the SFS algorithm, the feature set becomes shorter, whereas accuracy becomes higher compared with the original feature set that explains the effectiveness of SFS. From Table 4, we can examine ACC with different neighborhood sizes and selected features. When $K = 5$, the performance remains stable before and after data normalization, and ACC reaches 96.58% after normalization and finally selects 12 (effective-Z in the arterial phase), 30 (effective-Z in the venous phase), 31 (Calcium-Iodine in the venous phase), 33 (Iodine-Calcium in the venous phase), and 14 (Calcium-Water in the arterial phase) feature sets. These

selected features are highly related to the classification results (lymph node metastasis). Among which the 12 (effective-Z in the arterial phase), 30 (effective-Z in the venous phase), 33 (Iodine-Calcium in the venous phase) feature sets are consistent with the pathology theory and clinical experience of doctors. As for the other feature sets, their effectiveness need to be further verified by studies. However, the SFS-KNN algorithm is not a global optimized solution, and it may lead to overfitting problems, which explain the decrease in ACC. In our experiments, the amount of the samples is not sufficient, so the large neighborhood size fails to reflect the local characteristics of the KNN classifier. Therefore, $K = 9$ is not selected as the optimal size.

3.1.3. mRMR-KNN Results. Figure 5 shows two feature selection procedures with different mRMR criteria. Tables 5 and 6 reveal the classification performance of mRMR-KNN (MIQ and MID) with different neighborhood sizes [27]. We can see form the two tables that the two criteria

TABLE 6: Classification performance of mRMR-KNN (MID) with different neighborhood sizes.

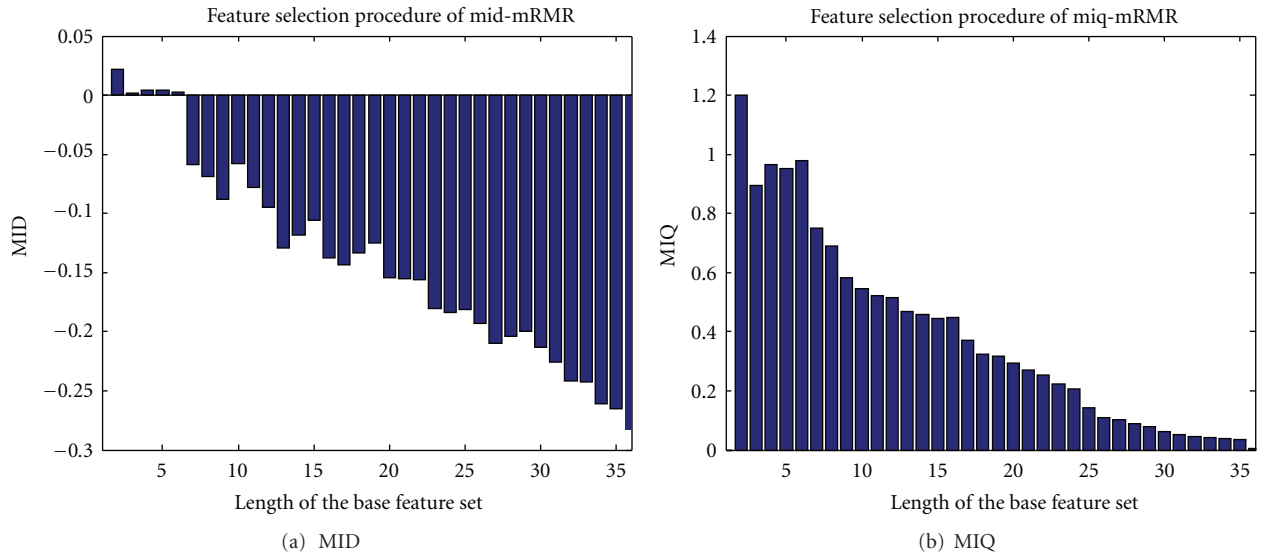| Neighborhood size		$K = 1$	$K = 3$	$K = 5$	$K = 7$	$K = 9$
	Sequence	12, 26, 22, 18, 3, 30, 14, 6, 19, 16, 36, 2, 17, 5, 1, 24, 35, 15, 23, 4, 34, 13, 29, 21, 7, 31, 11, 32, 25, 20, 9, 28, 33, 10, 8, 27				
Prenorm	Length	1	26	26	26	20
	Accuracy	87.50%	90.39%	89.34%	87.11%	82.50%
	Sequence	15, 21, 2, 30, 24, 17, 5, 14, 23, 12, 18, 22, 27, 4, 16, 33, 7, 1, 20, 25, 13, 3, 29, 19, 6, 35, 28, 31, 8, 32, 26, 11, 34, 36, 9, 10				
Norm	Length	4	2	2	16	8
	Accuracy	89.74%	94.87%	94.87%	95.66%	95.26%

(a) MID

(b) MIQ

FIGURE 5: Feature selection procedures with different mRMR criteria.

of MIQ and MID acquire almost the same performances. After normalization, the accuracy with all different K are highly increased, thus demonstrate the positive effect of data normalization. Among the feature sets, we can conclude from the table that 15 (Iodine-Calcium in the arterial phase), 21 (60 keV in the venous phase), 30 (Effective-Z in the venous phase), and 3 (60 keV in arterial phase) are closely related to the lymph node metastasis, which highly agree with the pathology theory and clinical experience of doctors. With $K = 5$, the classification performance remains stable before and after normalization, which further verifies the optimal K (neighborhood size) value.

3.1.4. Metric Learning Results.
Figure 6 shows 2D visualized results of 6 different distance metric learning methods in one validation. In the two-dimensional projection space, the classes are better separated by the LDA transformation than by other distance metrics. However, the result of KNN with single distance metric is not very satisfying, that's why we consider combination.

Table 7 shows the classification accuracy of KNN algorithm with different-distance metric learning methods. Apparently, these results show that the data normalization

TABLE 7: Classification performance of the KNN algorithm with metric learning methods.

Data (length of feature set)	Prenorm (4)	Norm 01 (5)
KNN	80.79%	83.68%
without PCA	80.79%	83.68%
PCA	82.11%	81.84%
PCA + LDA	77.89%	96.33%
PCA + RCA	77.63%	96.33%
PCA + LFDA	76.97%	96.33%
PCA + NCA	76.58%	86.32%
PCA + LMNN	76.84%	96.33%

helps a lot on classification. Moreover, PCA is a popular algorithm for data dimensionality reduction and operates in an unsupervised setting without using the class labels of the training data to derive informative linear projections. However, PCA can still have useful properties as linear preprocessing for KNN classification. By combining PCA with other supervised distance metric learning methods (e.g., LDA, RCA), we can obtain greatly improved performance. The accuracy of KNN classification depends significantly

(a) PCA

(b) LDA

(c) RCA

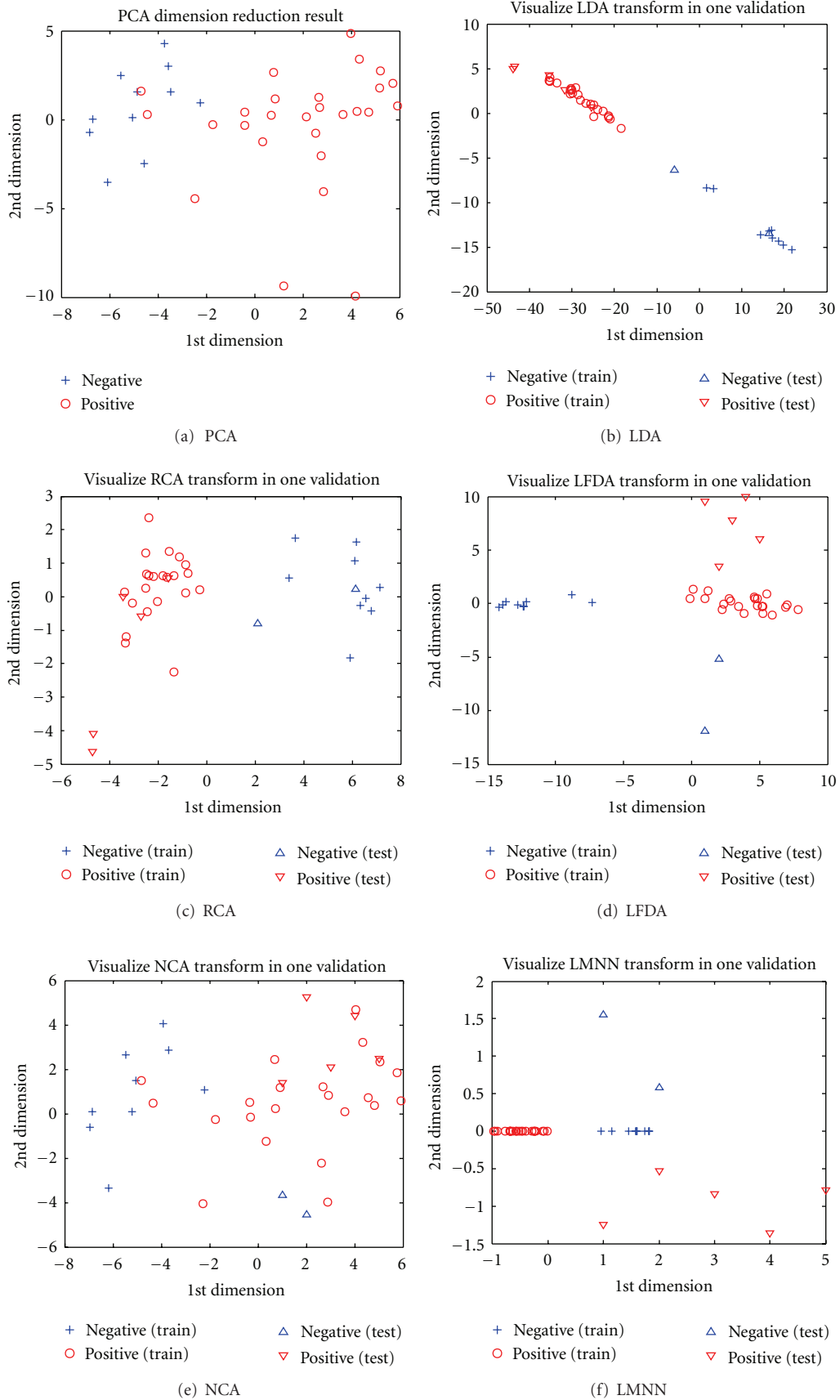
(d) LFDA

(e) NCA

(f) LMNN

FIGURE 6: Dimension reduction results of different metric learning methods in one validation.

on the metric used to compute distances between different samples.

3.2. Discussion. Based on the experimental results, the use of machine learning methods can improve the accuracy of clinical lymph node metastasis in gastric cancer. In our study, we mainly used the KNN algorithm for classification, which shows high efficiency. To improve effectiveness and classification accuracy, we first employed several feature-selection algorithms, such as mRMR and SFS methods, which both show an increase in accuracy. We obtained the highly related features of lymph node metastasis in accordance with the validated results of clinical pathology. Another way to improve accuracy is the use of distance metric learning for the input space of the data from a given collection of similar/dissimilar points that preserve the distance relation among the training data, and the application of the KNN algorithm in the new data patterns. Some schemes used in our experiments attained the overall accuracy of 96.33%.

4. Conclusions

The main contribution of our study is to prove the feasibility and the effectiveness of machine learning methods for computer-aided diagnosis (CAD) of lymph node metastasis in gastric cancer using clinical GSI data. In this paper, we employed a simple and classic algorithm called KNN that combines several feature selection algorithms and metric learning methods. The experimental results show that our scheme outperforms traditional diagnostic means (e.g., EUS and MDCT).

One limitation of our research is the insufficient number of clinical cases. Thus, in our future work, we will conduct more experiments on clinical data to improve further the efficiency of the proposed scheme and to explore more useful and powerful machine learning methods for CAD in clinical.

Acknowledgments

References

[1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[2] "Cancer Facts and Figures," The American Cancer Society, 2012.

[3] M. H. Lee, D. Choi, M. J. Park, and M. W. Lee, "Gastric cancer: imaging and staging with MDCT based on the 7th AJCC guidelines," *Abdominal Imaging*, vol. 37, no. 4, pp. 531–540, 2011.

[4] M. M. Ozmen, F. Ozmen, and B. Zulfikaroglu, "Lymph nodes in gastric cancer," *Journal of Surgical Oncology*, vol. 98, no. 6, pp. 476–481, 2008.

[5] P. Aurello, F. D'Angelo, S. Rossi et al., "Classification of lymph node metastases from gastric cancer: comparison between N-site and N-number systems. Our experience and review of the literature," *American Surgeon*, vol. 73, no. 4, pp. 359–366, 2007.

[6] T. Akagi, N. Shiraishi, and S. Kitano, "Lymph node metastasis of gastric cancer," *Cancers*, vol. 3, no. 2, pp. 2141–2159, 2011.

[7] H. Saito, Y. Fukumoto, T. Osaki et al., "Prognostic significance of the ratio between metastatic and dissected lymph nodes (n ratio) in patients with advanced gastric cancer," *Journal of Surgical Oncology*, vol. 97, no. 2, pp. 132–135, 2008.

[8] F. Espin, A. Bianchi, S. Llorca et al., "Metastatic lymph node ratio versus number of metastatic lymph nodes as a prognostic factor in gastric cancer," *European Journal of Surgical Oncology*, vol. 38, pp. 497–502, 2012.

[9] M. Karcaaltincaba and A. Aktas, "Dual-energy CT revisited with multidetector CT: review of principles and clinical applications," *Diagnostic and Interventional Radiology*, vol. 17, pp. 181–194, 2011.

[10] C. Yan, Z. G. Zhu, M. Yan et al., "Value of multidetector-row computed tomography in the preoperative T and N staging of gastric carcinoma: a large-scale Chinese study," *Journal of Surgical Oncology*, vol. 100, no. 3, pp. 205–214, 2009.

[11] T. Fukuya, H. Honda, T. Hayashi et al., "Lymph-node metastases: efficacy of detection with helical CT in patients with gastric cancer," *Radiology*, vol. 197, no. 3, pp. 705–711, 1995.

[12] N. Chandra and D. A. Langan, "Gemstone detector: dual energy imaging via fast kvp switching," in *Dual Energy CT in Clinical Practice*, T. F. Johnson, T. F. C, S. O. Schönberg, and M. F. Reiser, Eds., Springer, Berlin, Germany, 2011.

[13] D. Zhang, X. Li, and B. Liu, "Objective characterization of GE discovery CT750 HD scanner: gemstone spectral imaging mode," *Medical Physics*, vol. 38, no. 3, pp. 1178–1188, 2011.

[14] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[15] Y. Cai, T. Huang, L. Hu, X. Shi, L. Xie, and Y. Li, "Prediction of lysine ubiquitination with mRMR feature selection and analysis," *Amino Acids*, vol. 42, pp. 1387–1395, 2012.

[16] F. Amiri, M. R. Yousefi, C. Lucas, A. Shakery, and N. Yazdani, "Mutual information-based feature selection for intrusion detection systems," *Journal of Network and Computer Applications*, vol. 34, no. 4, pp. 1184–1199, 2011.

[17] D. Ververidis and C. Kotropoulos, "Sequential forward feature selection with low computational cost," in *Proceedings of the 8th European Signal Processing Conference*, Antalya, Turkey, 2005.

[18] L. Wang, A. Ngom, and L. Rueda, "Sequential forward selection approach to the non-unique oligonucleotide probe selection problem," in *Proceedings of the 3rd IAPR International Conference on Pattern Recognition in Bioinformatics*, pp. 262–275, 2008.

[19] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 625–632, 2011.

[20] Z. Huang, H. T. Shen, J. Shao, S. Rüger, and X. Zhou, "Locality condensation: a new dimensionality reduction method for image retrieval," in *Proceedings of the 16th ACM International Conference on Multimedia, (MM '08)*, pp. 219–228, can, October 2008.

[21] Liu Yang and Rong Jin, "Distance metric learning: a comprehensive survey," Tech. Rep., Department of Computer Science and Engineering, Michigan State University, 2006.

[22] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning distance functions using equivalence relations," in *Proceedings of the 20th International Conference on Machine Learning*, pp. 11–18, August 2003.

[23] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," in *Proceedings of the Conference on Neural Information Processing Systems (NIPS '04)*, 2004.

[24] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

[25] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, New York, NY, USA, 2007.

[27] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, no. 2, pp. 185–205, 2005.

Research Article

A New Particle Swarm Optimization-Based Method for Phase Unwrapping of MRI Data

Wei He,[1] Yiyuan Cheng,[1] Ling Xia,[1] and Feng Liu[2]

[1] *Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China*
[2] *School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia*

Correspondence should be addressed to Ling Xia, xialing@zju.edu.cn

A new method based on discrete particle swarm optimization (dPSO) algorithm is proposed to solve the branch-cut phase unwrapping problem of MRI data. In this method, the optimal order of matching the positive residues with the negative residues is first identified by the dPSO algorithm, then the branch cuts are placed to join each pair of the opposite polarity residues, and in the last step phases are unwrapped by flood-fill algorithm. The performance of the proposed algorithm was tested on both simulated phase image and MRI wrapped phase data sets. The results demonstrated that, compared with conventionally used branch-cut phase unwrapping algorithms, the dPSO algorithm is rather robust and effective.

1. Introduction

In magnetic resonance imaging (MRI), the complex signal contains both the magnitude and phase parts. Usually the magnitude of the MRI signal has been mainly considered. However, the phase of MRI signal offers very important information on the velocity of the moving spins, and can also be used to deduce useful information about the main B_0 field inhomogeneity and the magnetic susceptibility variations [1]. In MRI, the phase information $\psi_{i,j}$ is usually obtained from a complex MRI dataset $I_{i,j} = |I_{i,j}| \exp(\psi_{i,j})$ through some mathematical operations, and the value always lies in the principal interval of $(-\pi, \pi]$, consequently producing a wrapped phase $\varphi_{i,j}$. This relationship can be described by $\varphi_{i,j} = W(\psi_{i,j}) = \psi_{i,j} \pm 2k_{i,j}\pi$, where $k_{i,j}$ is an integer and W defines a wrapping operator that forces all values of its argument into the range $(-\pi, \pi]$ by adding or subtracting an integral multiple of 2π radians from its argument. Phase unwrapping is the process of estimating the true phase $\psi_{i,j}$ from the wrapped phase $\varphi_{i,j}$. As an important tool, it can not only be used for the three-point Dixon water and fat separation, but also be applied to increase the dynamic range of phase contrast MR velocity measurements [2]. If the true phase gradients (i.e., the differences of $\psi_{i,j}$) between contiguous pixels are less than π radians in magnitude

in the entire space, the true phase can be unwrapped in a straightforward manner by just integrating the wrapped phase gradients [3]. However, the presence of the noise, undersampling, and/or object discontinuities often makes this condition unavailable. Therefore, the problem of phase unwrapping becomes complex in practice and difficult to solve, although significant amount of research effort has been devoted to date. In the literature, there are quite a few existing phase unwrapping algorithms [4], which can be grouped into two categories: path-following and minimum-norm methods [5].

The branch-cut phase unwrapping method is one kind of the path-following methods. Unlike the minimum-norm methods, the branch-cut phase unwrapping technique offers correct phase unwrapping with no solution approximations [4]. In the operation, it locates residues, joins the residues by branch cuts, and then unwraps all the pixels avoiding those branch cuts. In the algorithm, residues are defined as local inconsistencies, which mark the starting and end of 2π discontinuities. Corresponding to formula (1), that is, the value n is 1 or -1 in a 2×2 closed-loop of the wrapped phase gradients [4], as shown in Figure 1:

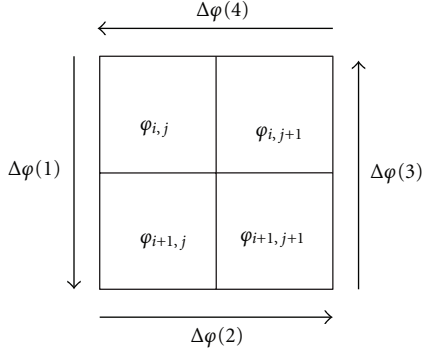$$\sum_{i=1}^{4} \Delta\varphi(i) = 2\pi n. \tag{1}$$

FIGURE 1: Residue calculation.

These wrapped phase gradients are computed counter-clockwise:

$$
\begin{aligned}
\Delta\varphi(1) &= W\left(\varphi_{i+1,j} - \varphi_{i,j}\right), \\
\Delta\varphi(2) &= W\left(\varphi_{i+1,j+1} - \varphi_{i+1,j}\right), \\
\Delta\varphi(3) &= W\left(\varphi_{i,j+1} - \varphi_{i+1,j+1}\right), \\
\Delta\varphi(4) &= W\left(\varphi_{i,j} - \varphi_{i,j+1}\right).
\end{aligned}
\tag{2}
$$

In formula (1), when $n = 1$, we label it a residue with a positive polarity. And when $n = -1$, the residue is labeled with a negative polarity. Otherwise it indicates there is no residue when $n = 0$. In this algorithm, the total length of branch cuts must been minimized, resulting a decrease of the amount of good pixels used as the branch cuts. This provides more unwrapping paths in dense residue areas, leading to a smoother result [5].

To achieve a minimum total length of branch cuts, various techniques have been developed for the branch-cut phase unwrapping method. These techniques include, for example, the Goldstein's algorithm [5, 6], the nearest-neighbor algorithm [7], the minimum-cost matching (MCM) algorithm [8], and the hybrid genetic algorithm (HGA) [4].

The nearest-neighbor algorithm is very efficient, but it utilizes local heuristics, thus causing some long branch cuts embedded in the phase image. Therefore, the distribution of the branch cuts does not achieve the optimum, and the resultant phase image is lack of the smoothness. The MCM is a graph theory-based algorithm, which uses the Hungarian algorithm to find the minimum total length of branch cuts. Although powerful, it is computationally expensive. The HGA employs a combination of global and local search methods, and its solution is usually good. However, the complexity of this algorithm tends to be a problem with the increase of residues.

Compared with these three methods, which place the branch cuts to connect pairs of residues of opposite polarity (called dipoles), the Goldstein's method joins the residues in clumps instead of pairs [5]. The Goldstein's algorithm is very efficient, but often forms some isolated patches.

In this paper, we propose a new discrete particle-swarm-optimization- (dPSO-) based branch-cut algorithm

for phase unwrapping. The new dPSO algorithm is used to find the best way in which the negative polarity residues match with the positive ones, so that the overall length of the branch cuts is minimized. The performance of the new dPSO algorithm is compared with the Goldstein's and MCM algorithms.

2. Phase Unwrapping Using the Proposed Algorithm

2.1. Overview of the Basic PSO Algorithm. As an artificial intelligent algorithm, the particle swarm optimization (PSO) [9, 10], is easy for implementation (only few parameters to be adjusted) and converges fast.

In the PSO algorithm, the swarm consists of several particles, and each contains N elements. Then each particle is viewed as a point in an N-dimensional space. The ith particle of swam is represented as $U_i = \{u_{i1}, u_{i2}, \ldots, u_{iN}\}$. All the particles share their information and move to find the global optima. $P_i = \{p_{i1}, p_{i2}, \ldots, p_{iN}\}$ represents that the local best position that the ith particle has reached, and $P_g = \{p_{g1}, p_{g2}, \ldots, p_{gN}\}$ is the global best position. The velocity of the ith particle is $V_i = \{v_{i1}, v_{i2}, \ldots, v_{iN}\}$. Each particle of the swarm updates its velocity and position using the following formulas:

$$
\begin{aligned}
V_i^{t+1} &= w * V_i^t + c_1 * \mathrm{rand}() * \left(P_i^t - U_i^t\right) \\
&\quad + c_2 * \mathrm{Rand}() * \left(P_g^t - U_i^t\right),
\end{aligned}
\tag{3}
$$

$$
U_i^{t+1} = U_i^t + V_i^{t+1},
\tag{4}
$$

where t denotes the iteration number, c_1 and c_2 are learning factors (nonnegative constants), controlling (or regulating) the influence of P_i and P_g, the function rand() and Rand() generate a random number ($[0{\sim}1]$), and w is the inertia weight factor.

A problem-specific fitness function (symbolized by f) is employed to measure the performance of each particle. Thereby, for a minimization problem, P_i and P_g can be found in current iteration as follows:

$$
P_i^{t+1} = \begin{cases} P_i^t & \text{if } f(P_i^t) \le f(U_i^t) \\ U_i^t & \text{if } f(P_i^t) > f(U_i^t), \end{cases}
\tag{5}
$$

$$
P_g^t = \arg\min_{P_i^t}\left[f(P_i^t)\right].
\tag{6}
$$

To date the PSO technique has been well developed for the continuous problem, but not in discrete domain.

2.2. The dPSO Algorithm for Phase Unwrapping

2.2.1. Particle and Swarm Initialization. Any problem adopting the PSO algorithm has to be interpreted into PSO particle form. For phase unwrapping, every particle should be composed of some elements corresponding to the indexes of residues. If we calculate all the residues in the wrapped phase image at one stroke, sometimes the size of the particle may be too large and the swarm size required may be

enlarged accordingly. This may result in a poor solution and/or extending the convergence time. To avoid this, the whole image is divided into sub regions and therefore the residues are set into different small groups.

The process can be described as follows.

(1) The image is partitioned on the basis of its phase derivative variance map. The phase derivative variance is defined as follows [5]:

$$Z_{m,n} = \frac{\sqrt{\sum \left(\Delta\varphi^x_{i,j} - \overline{\Delta\varphi^x_{m,n}} \right)^2} + \sqrt{\sum \left(\Delta\varphi^y_{i,j} - \overline{\Delta\varphi^y_{m,n}} \right)^2}}{l^2}, \quad (7)$$

where for each sum the indexes (i, j) cover over the $l \times l$ window centered at the pixel (m, n). The terms $\Delta\varphi^x_{i,j}$ and $\Delta\varphi^y_{i,j}$ are the wrapped phase gradients in the $l \times l$ windows, and $\overline{\Delta\varphi^x_{m,n}}$ and $\overline{\Delta\varphi^y_{m,n}}$ are the averages of these wrapped phase gradients. In this paper $l = 3$.

(2) Based on an appropriate threshold, the phase derivative variance map can be converted into a binary one, where the low phase derivative variance values turns out to be 0 and the high ones becomes to be 1. In this way the image is divided into separate areas. To obtain a proper threshold, a classical approach called Otsu's method [11] has been adopted in this study.

(3) It is observed that the majority of residues cluster in the patches of value 1. Thus the residues are grouped by taking the above steps.

In each region the indexes of the residues are inserted in two arrays regardless of the order. One is a positive polarity residue array, and the other is a negative polarity residue array. Provided there are M positive polarity residues and N negative polarity residues in one region, respectively, the positive residue array and the negative one are accordingly denoted as $\{s_1, s_2, \ldots, s_M\}$ and $U_i = \{u_{i1}, u_{i2}, \ldots, u_{iN}\}$. The former will be fixed throughout all generations and serves as a reference. And the later acts as a particle U_1 of the initial swarm. The rest particles of the swarm are generated via arranging the order of the elements in U_1 in a random manner.

2.2.2. Fitness Estimate.
From the abovementioned, it can be easily seen that the dPSO algorithm is used to find out the best matched order of the elements in the particle with the reference array.

In dPSO, the quality of the current solution is judged by the fitness function. Since the branch-cut phase unwrapping must minimize the total length of branch cuts, the corresponding fitness function is obviously for calculating the total length of branch cuts in the wrapped phase image. Here we employ Euclidian distance to assess the total cut length:

$$f = \sum_{j}^{\min(M,N)} \sqrt{\left(x_{s_j} - x_{u_{ij}} \right)^2 + \left(y_{s_j} - y_{u_{ij}} \right)^2}, \quad (8)$$

where x and y denote the residue's x-coordinate and y-coordinate.

2.2.3. Velocity Update.
Owing to the attribute of dPSO algorithm for branch-cut phase unwrapping, the iterative velocity in formula (3) should be a set of permutation operators rather than a usual vector. Various permutation operators have been introduced for discrete particle swarm optimization. Here we choose the adjustment operator [12], not the swap operator [13, 14], as the permutation operator. This is because compared with the swap operator, the adjustment operator avoids returning to the previous position [12].

In addition, during all the iterations, w is set to be linearly varied ([0.9~0.4]) [10]. This setting considers the global searching capacity and convergence rate of the optimization:

$$w = 0.9 - 0.5 \times \left(\frac{t}{T} \right), \quad (9)$$

where T is the maximal iteration times.

To make formula (3) suitable for the dPSO operation, some concepts are given as follows.

Definition 1. The adjustment operator $AO(k, l)$ is defined as deleting the element in the kth position and popping it in the lth position in the array.

For example, $AO(5, 3)$ acting on the array $U_i = \{5, 1, 4, 2, 3\}$ gets a result $U'_i = \{5, 1, 3, 4, 2\}$.

Definition 2. One or more adjustment operators make up an adjustment sequence (AS). That is $AS = \{AO(k_1, l_1), AO(k_2, l_2), \ldots, AO(k_n, l_n)\}$.

Acting an AS on an array means that every adjustment operator of the sequence acts on the array in turn. Therefore, in the AS, it is critical to properly arrange the order of adjustment operators.

Definition 3. The plus sign "+" between ASs has its new meaning. It is defined as forming a new longer AS by putting the latter behind the former.

For example, $AO(2, 4) + AO(5, 4) = \{AO(2, 4), AO(5, 4)\}$, $\{AO(2, 4), AO(5, 4)\} + AO(3, 1) = \{AO(2, 4), AO(5, 4), AO(3, 1)\}$, $\{AO(2, 4), AO(5, 4)\} + \{AO(1, 3), AO(5, 4)\} = \{AO(2, 4), AO(5, 4), AO(1, 3), AO(5, 4)\}$. This operation does not satisfy the commutative law.

Definition 4. The minus sign "−" between two arrays means constructing an AS which can act on the array after minus sign to obtain the array before.

Supposing there are two arrays, according to the array before minus sign from left to right, this AS can be obtained by adjusting the order of the array after. For example, in $W - R$, $W = \{1, 4, 3, 2, 5\}$ and $R = \{1, 5, 4, 2, 3\}$. $W(2) = R(3)$, so the first adjustment operator of AS is $AO(3, 2)$, and the first new array $R' = \{1, 4, 5, 2, 3\}$. Then $W(3) = R'(5)$, so the second adjustment operator is $AO(5, 3)$, and the second new array is $R'' = \{1, 4, 3, 5, 2\}$. $W(4) = R''(5)$, so the third adjustment operator is $AO(5, 4)$, and we finally obtain the same array as W. Thus the calculation ends and $AS = W - R = \{AO(3, 2), AO(5, 3), AO(5, 4)\}$.

Definition 5. The multiplication sign "∗" between two real numbers is simply the operation of multiplication.

Definition 6. The sign "∗" between a real number and an AS means reserving a certain amount of the adjustment operators in the AS in turn when the real number is in the range $(0, 1)$.

Given that the number of adjustment operators in the AS denotes as $||AS||$, the number of retaining adjustment operators is $[b||AS||]$, which rounds the product of b and $||AS||$ to the nearest integer.

2.2.4. Position Update. Similarly, the formula (4) has to be redeclared to fit the requirement of the dPSO.

Definition 7. The plus sign "+" between an array and an AS means acting the adjustment operators of the AS on the array in order. For example, given $U_i = \{5, 1, 3, 2, 4\}$ and AS = $\{AO(2, 1), AO(2, 5)\}$, $U_i + AS = \{1, 3, 2, 4, 5\}$.

It is obvious that the operations defined in Definitions 4 and 7 are inverse to each other.

2.2.5. The Procedure Description of dPSO Algorithm for Phase Unwrapping. The process of using the dPSO algorithm for phase unwrapping is summarized as follows.

(1) According to Section 2.2.1, the residues in the image are divided into several clusters. Set $h = 1$.

(2) Do the following steps in the hth group of residues.

(3) Set values to the parameters of dPSO, which include the learning factors (c_1, c_2), the maximal iteration times (T), and the number of particles in swarm, and also the termination criteria.

(4) Initialize the swarm according to Section 2.2.1. Each particle has its random velocity, that is, AS. Set $t = 1$.

(5) Evaluate the fitness of every particle according to formula (8) and find the current P_i P_g by formula (5), (6), respectively. Calculate the current inertia weight factor according to formula (9).

(6) Set $t = t + 1$. Use formula (3) to get the new velocity V_i. Then calculate the new position U_i according to formula (4).

(7) Repeat (5)-(6) until $t = T$ or meeting the termination criteria.

(8) P_g is the best indexes order of the negative polarity residues matched with the positive ones in this group.

(9) Set $h = h + 1$. Repeat (2)–(8) until h equals the number of the residues groups plus 1.

2.3. Branch Cuts and Unwrapping. Once the best match in every group has been found by dPSO, each pair of two matching opposite polarity residues are connected by the branch cuts. It is worth mentioning that, owing to that the number of opposite polarity residues is not always equal to each other, there are usually one or more residues left in each group. Then the nearest-neighbor algorithm [7] is employed to place branch cuts to balance these remaining residues. So far all the residues have been balanced by branch cuts.

Finally, the phase data can be unwrapped by flood-fill algorithm [15, 16], without crossing the branch cuts as follows:

(1) Choose a start pixel, whose phase value is stored as an unwrapped phase value in the solution matrix. The four neighboring pixels are unwrapped next and their unwrapped phase values are placed in the solution matrix. These four pixels are inserted in the unwrapped list.

(2) Pick (and then eliminate) a pixel from the unwrapped list. Unwrap the phase values of its four neighboring pixels, avoiding pixels that have been unwrapped. Insert these pixels in the unwrapped list and put their unwrapped phase values in the solution matrix.

(3) Repeat (2) until the unwrapped list becomes empty.

In fact, it does not always mean that all the pixels have been unwrapped when the unwrapped list becomes empty. Because sometimes there are some pixels in the image encircled by the branch cuts, they cannot be unwrapped if not crossing the branch cuts.

2.4. Weighted L^0 Measure. Weighted L^0 measure, the most general/practical error measure to consider [5], is used to evaluate the quality of an unwrapped solution:

$$\varepsilon = \frac{1}{rc}\left[\sum_{i=1}^{r-1}\sum_{j=1}^{c} w_{i,j}^x \left\|\psi(i+1, j) - \psi(i, j) - \Delta\varphi_{i,j}^x\right\|_0 \right.$$
$$\left. + \sum_{i=1}^{r}\sum_{j=1}^{c-1} w_{i,j}^y \left\|\psi(i, j+1) - \psi(i, j) - \Delta\varphi_{i,j}^y\right\|_0\right], \tag{10}$$

where r and c are the number of rows and columns, $w_{i,j}^x$ and $w_{i,j}^y$ are user-defined weights, and the L^0 norm measures a count of the number of pixels at which the gradients of the unwrapped solution mismatch the wrapped phase gradients. In this paper the weights adopted are derived from the quality map mentioned above, not just omitted (i.e., equal to 1).

3. Results and Discussion

We have tested the performance of the proposed algorithm on both simulated and MRI phase data on a PC (Intel 2 Quad CPU 2.39 GHz, MATLAB). We set both learning factors (c_1, c_2) used in dPSO to be 2. The results were compared with the well-known Goldstein's and the MCM algorithms.

3.1. Simulation Results. The proposed algorithm was implemented on a simulated wrapped phase image with salt and pepper noise (the signal-to-noise ratio is 7.58 dB), which has 2460 residues. Figure 2(a) shows the simulated wrapped phase image. And its residue distribution is shown in Figure 2(b), where the positive and negative residues are marked as white and black pixels, respectively.

The resultant unwrapped phase image in Figure 3(a) was achieved by dPSO using a swarm of 300 particles and
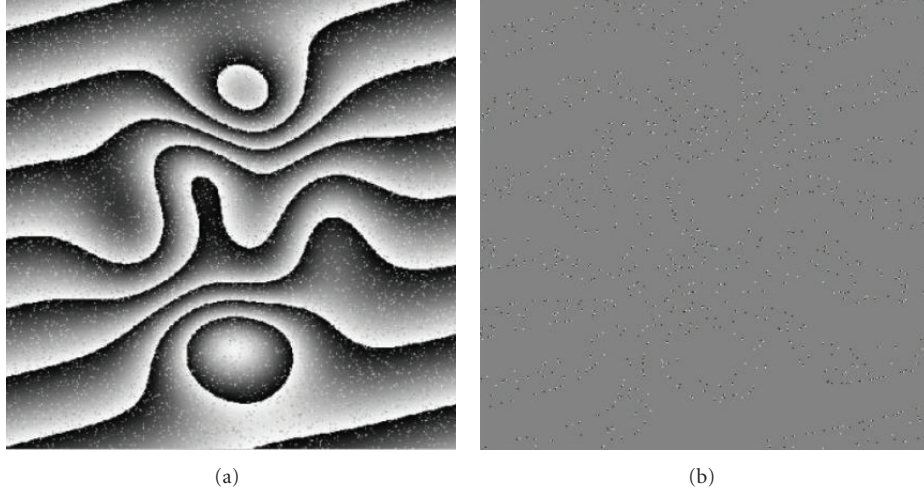
(a)

(b)

FIGURE 2: (a) A 512×512 simulated wrapped phase image, (b) its residue distribution including 2460 residues, 1231 positive polarity residues (white pixels), and 1229 negative polarity residues (black pixels).

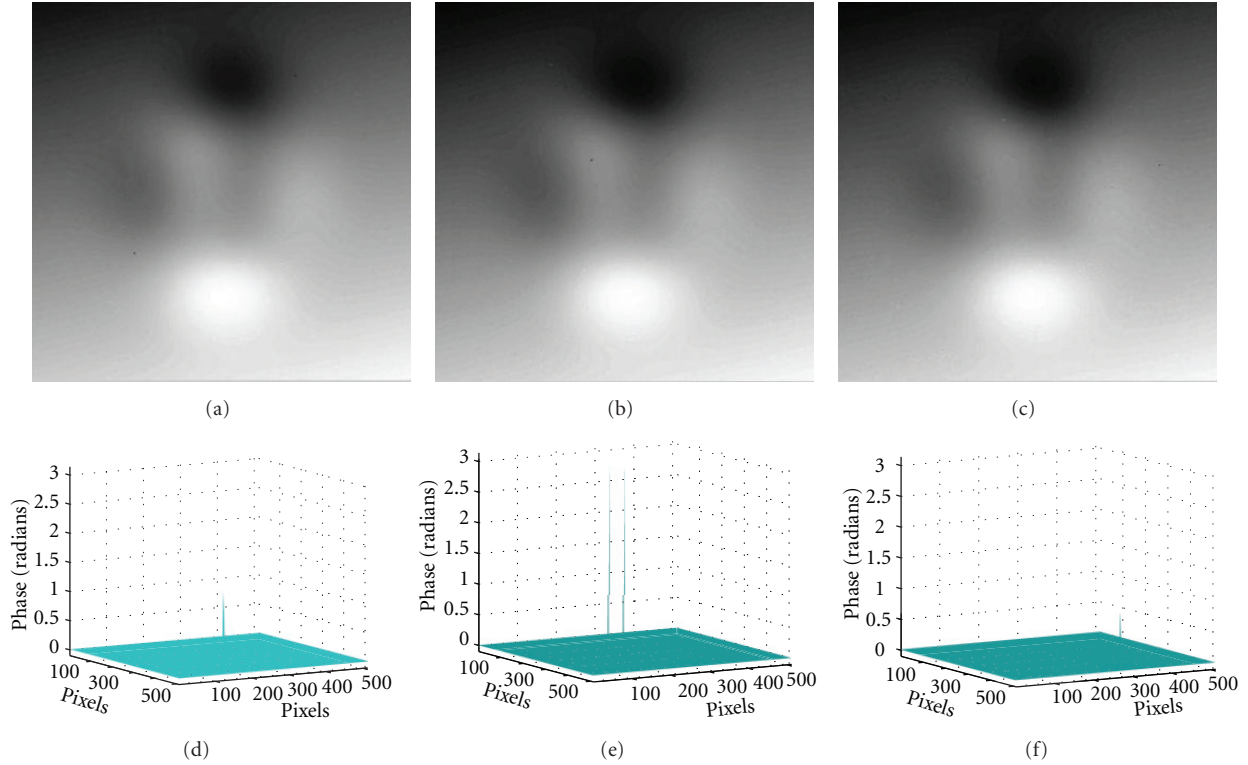(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 3: The top row is the unwrapped phase image for the simulated wrapped phase map in Figure 2(a) achieved by (a) dPSO, (b) Goldstein's, and (c) MCM algorithms. The bottom is the corresponding difference map got by (d) dPSO, (e) Goldstein's, and (f) MCM algorithms.

$T = 1000$. Figures 3(b) and 3(c) depict the corresponding unwrapped phase images obtained by Goldstein's and MCM algorithms, respectively. We rewrapped these resultant solutions and subtracted the original wrapped phase data from them. The corresponding difference maps, which are plotted in 3D visualization, are shown in Figures 3(d)–3(f). As mentioned in Section 1, the difference between the wrapped phase and the true phase is an integer multiples of

2π. As a result, these curve surfaces of difference maps give a visual representation of the deviations between the true phases and the results, which can illustrate the accuracy of the unwrapped solutions directly. The average value of the difference map is called *average difference*.

The performance of dPSO with respect to the other two algorithms can clearly be seen in Table 1. In terms of weighted L^0 measure, average difference, and total cuts length,

TABLE 1: Comparing dPSO with other algorithms for the simulated phase image in Figure 2(a) in terms of weighted L^0 measure, average difference, total cuts length, and execution time.

Algorithm	Weighted L^0 measure	Average difference (radian)	Total cuts length	Execution time (s)
dPSO	0.001301	$0.873e-5$	1310	1328
Goldstein's	0.001384	$3.77e-5$	1313	79
MCM	0.001286	$0.365e-5$	1298	3513

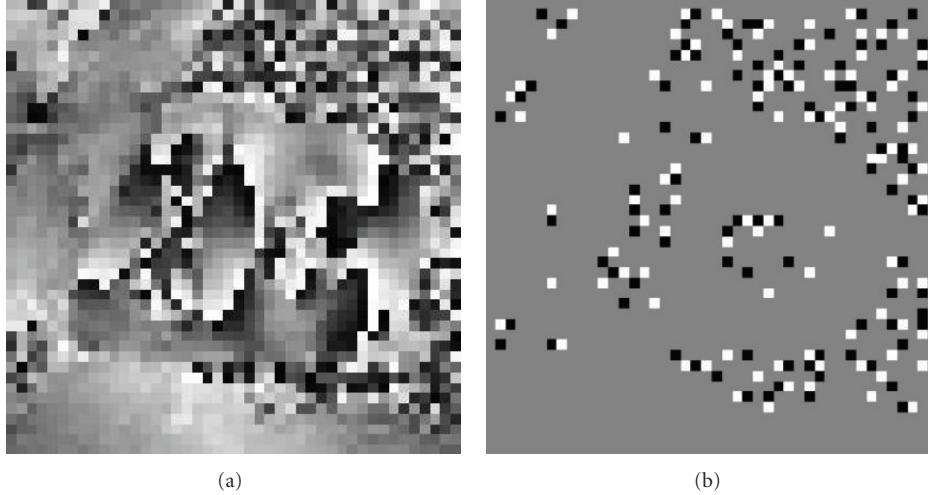(a)   (b)

FIGURE 4: (a) A 44×44 displacement encoded MRI heart phase image, (b) its corresponding residue distribution involving 396 residues, 198 positive polarity residues, and 198 negative polarity residues.

the dPSO algorithm is better than the Goldstein's, but not as good as the MCM. On the other hand, the execution time of dPSO is much less than that of the MCM, but not comparable to that of the Goldstein's.

3.2. Results of MRI Data. The proposed algorithm was also executed on a displacement encoded MRI heart phase data set [17] with 396 residues. The wrapped phase image and its corresponding residue distribution are shown in Figure 4.

Figures 5(a)–5(c) depict the branch-cut distribution achieved by the three algorithms, where the black pixels mark the branch cuts. The dPSO result in Figure 5(d) was obtained by using a swarm of 300 particles. And the results of the other algorithms are shown in Figures 5(e) and 5(f), respectively. In Figure 5(e) several patches are isolated, two large ones in the upper part, two small ones in the middle, and a very large one in the lower right part. Compared with Figures 4(b) and 5(b), it is easy to observe that these patches are completely isolated by branch cuts, which would lead to an incorrect unwrapping. In addition, the isolated areas tend to arise in the regions with dense residues, because in such regions the branch cuts are often close to each other and have more possibility to encircle some pixels. However, in Figures 5(d) and 5(f) the isolated patches are much smaller and less. The dipole branch-cut methods appear to be less likely to isolate regions in the phase image by branch cuts, since the branch cuts balance the residues in pairs not in clumps. The

difference maps of the three approaches are then generated like in Section 3.1, shown in Figures 5(g)–5(i). Intuitively, both the dPSO and MCM produce more desirable results than the Goldstein's.

As shown in Table 2, the Goldstein's algorithm is extremely fast. Neither dPSO nor MCM is comparable to it. But the proposed approach has the smallest weighted L^0 measure and average difference. In addition, its total cuts length is the shortest.

Another example is the MRI head phase data. The wrapped phase image is shown in Figure 6(a). Figure 6(b) depicts its residues distribution. The unwrapped phase images achieved by the dPSO, Goldstein's, and MCM algorithms are displayed in Figures 7(a)–7(c), respectively. Obviously, in Figure 7(b) the surrounding area was not unwrapped at all. According to the previous analysis, the Goldstein's method isolates this area by branch cuts. In comparison to Figure 6(a), the surrounding part roughly is the back ground with dense noise. However, due to the property of dipole branch-cut phase unwrapping method the branch cuts could hardly enclose the surrounding area, which certainly caused unwrapping phases in this area. The same inferences can be also made according to the difference maps of three methods shown in Figures 7(d)–7(f).

The weighted L^0 measure, average difference and total cuts length are calculated over the whole image whether the pixel is inside the region of interest (ROI) or background.
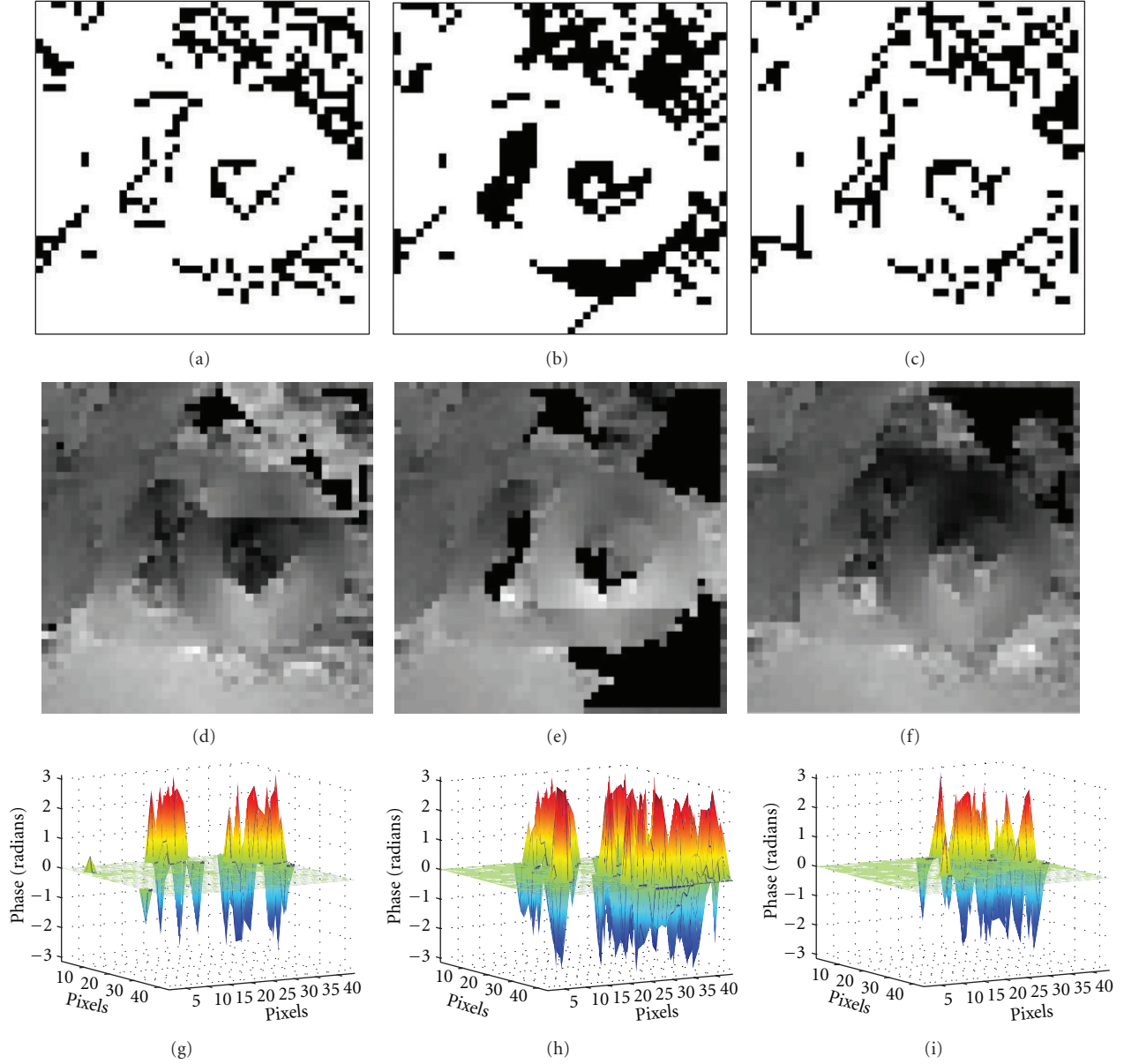
Figure 5: The top row is the branch-cut distribution for the MRI heart phase image in **Figure 4(a)** achieved using (a) dPSO, (b) Goldstein's, and (c) MCM algorithms. The middle is the corresponding unwrapped phase result of (d) dPSO, (e) Goldstein's, and (f) MCM algorithms. The bottom is the corresponding difference map of (g) dPSO, (h) Goldstein's, and (i) MCM algorithms.

Table 2: Comparing dPSO with other algorithms for the displacement encoded MRI heart phase map in **Figure 4(a)** in terms of weighted L^0 measure, average difference, total cuts length, and execution time.

Algorithm	Weighted L^0 measure	Average difference (radian)	Total cuts length	Execution time (s)
dPSO	0.052223	0.056251	310	142
Goldstein's	0.109667	0.294767	468	8
MCM	0.052490	0.076967	317	203

Thereby in these three respects, as shown in Table 3, the dPSO and MCM algorithms are much better than the Goldstein's. Though the dPSO method does not get a better solution than that of the MCM, there is little difference between them. That is, dPSO is comparable to MCM.

Furthermore the former converges nearly 68% faster than the latter.

Viewing dPSO's performance on these three examples, we can find that the dPSO algorithm takes more time to achieve an optimal solution when plenty of residues
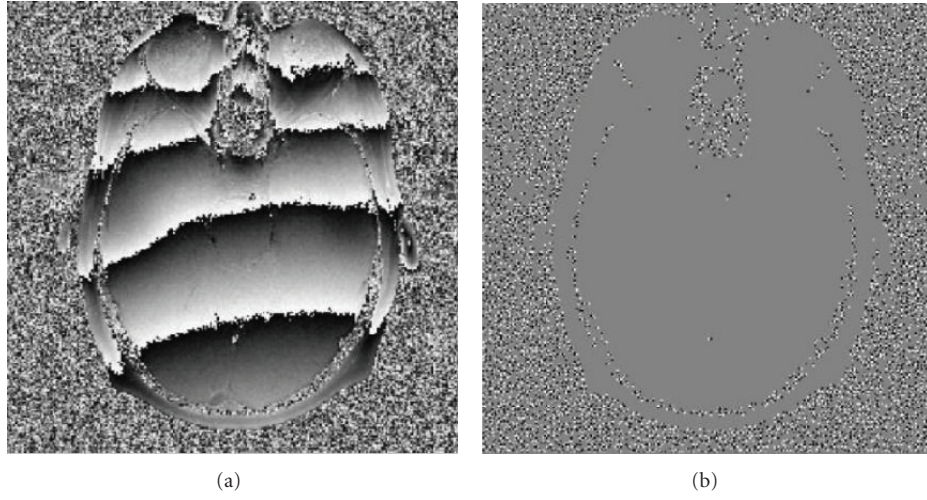
(a)

(b)

FIGURE 6: (a) A 256 × 256 MRI head phase image, (b) its corresponding residue distribution containing 9795 residues, 4904 positive polarity residues, and 4891 negative polarity residues.

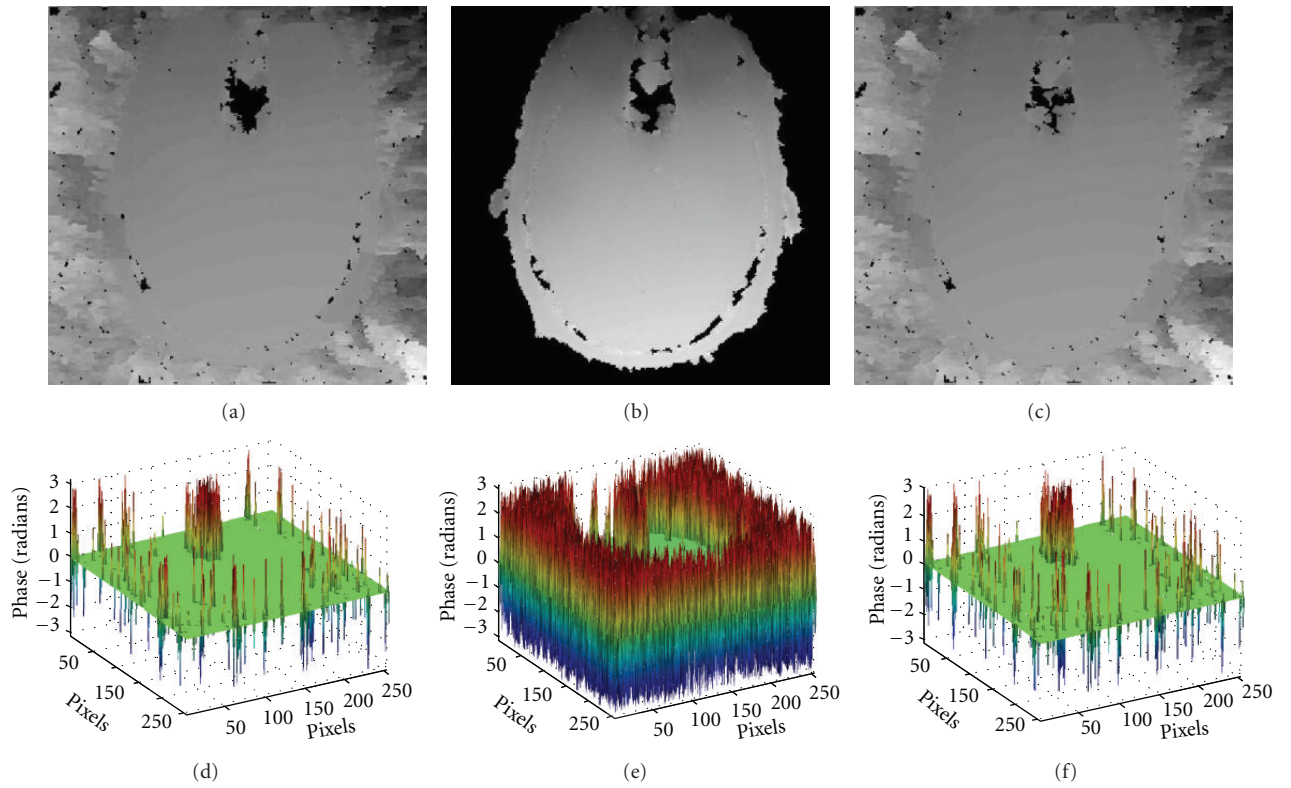(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 7: The top row is the unwrapped phase image for the MRI head phase image in Figure 6(a) achieved by (a) dPSO, (b) Goldstein's, and (c) MCM algorithms. The bottom is the corresponding difference map got by (d) dPSO, (e) Goldstein's, and (f) MCM algorithms.

TABLE 3: Comparing dPSO with other algorithms for the MRI head phase map in Figure 6(a) in terms of weighted L^0 measure, average difference, total cuts length, and execution time

Algorithm	Weighted L^0 measure	Average difference (radian)	Total cuts length	Execution time (s)
dPSO	0.064380	0.025642	9871	979
Goldstein's	0.156827	0.665523	25533	41
MCM	0.063179	0.021957	9846	3052

uniformly scatter throughout large areas. This is because the pixels in each of these areas often have similar quality and then the residues in each area can hardly be separated into more than one group, which results in the increase of the particle size for every group.

4. Conclusions

We have presented a new branch-cut phase unwrapping method based on dPSO algorithm in this paper. Both simulated and real wrapped phase data were used to test the performance of the proposed algorithm. The results of dPSO were compared with the Goldstein's and the MCM algorithms. It was found that the dPSO method is better than Goldstein's algorithm in terms of weighted L^0 measure, average difference and total branch cuts length. Moreover, the dPSO is much faster than the MCM algorithm in getting a global optimum solution while it is comparable to the latter in terms of weighted L^0 measure, average difference and total branch cuts length. Generally speaking, it has been demonstrated to be robust, effective for the phase unwrapping application.

In addition, it is capable of dealing with large branch-cut problem with thousands of residues. The complexity of the dPSO algorithm increases when the number of residues in a group increases, as the length of the particle extends which requires a larger swarm size. Future research will make this algorithm to be more efficiently operated for the phase unwrapping study.

Acknowledgment

References

[1] I. R. Young and G. M. Bydder, "Phase imaging," in *Magnetic Resonance Imaging*, D. D. Stark and W. G. Bradley, Eds., Mosby Year Book, St. Louis, Mo, USA, 2nd edition, 1992.

[2] J. Rydell et al., "Phase sensitive reconstruction for water/fat separation in MR imaging using inverse gradient," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '07)*, Brisbane, Australia, October 2007.

[3] K. Itoh, "Analysis of the phase unwrapping problem," *Applied Optics*, vol. 21, no. 14, article 2470, p. 2470, 1982.

[4] S. A. Karout, M. A. Gdeisat, D. R. Burton, and M. J. Lalor, "Two-dimensional phase unwrapping using a hybrid genetic algorithm," *Applied Optics*, vol. 46, no. 5, pp. 730–743, 2007.

[5] D. C. Ghiglia and M. D. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*, John Wiley & Sons, New York, NY, USA, 1998.

[6] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: two-dimensional phase unwrapping," *Radio Science*, vol. 23, no. 4, pp. 713–720, 1988.

[7] R. Cusack, J. M. Huntley, and H. T. Goldrein, "Improved noise-immune phase-unwrapping algorithm," *Applied Optics*, vol. 34, no. 5, pp. 781–789, 1995.

[8] J. R. Buckland, J. M. Huntley, and J. M. Turner, "Unwrapping noisy phase maps by use of a minimum-cost-matching algorithm," *Applied Optics*, vol. 34, no. 23, pp. 5100–5108, 1995.

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

[10] Y. Shi and R. Eberhart, "Modified particle swarm optimizer," in *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation (ICEC '98)*, pp. 1945–1950, May 1998.

[11] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[12] C. Wang, J. Zhang, J. Yang, C. Hu, and J. Liu, "A modified particle swarm optimization algorithm and its application for solving traveling salesman problem," in *Proceedings of the International Conference on Neural Networks and Brain Proceedings (ICNNB '05)*, pp. 689–694, October 2005.

[13] M. Clerc, "Discrete particle swarm optimization," 2000, http://clerc.maurice.free.fr/pso/pso_tsp/Discrete_PSO_TSP.htm.

[14] K. P. Wang, L. Huang, C. G. Zhou, and W. Pang, "Particle swarm optimization for traveling salesman problem," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 1583–1585, November 2003.

[15] Wikipedia, "Flood fill," http://en.wikipedia.org/wiki/Flood_fill.

[16] K. Chen, J. Xi, Y. Yu, and J. F. Chicharo, "Fast quality-guided flood-fill phase unwrapping algorithm for three-dimensional fringe pattern profilometry," in *Optical Metrology and Inspection for Industrial Applications*, vol. 7855, October 2010.

[17] B. Spottiswoode, "2D phase unwrapping algorithms," http://www.mathworks.com/matlabcentral/fileexchange/22504.

Research Article

Quantitative Measurements in 3-Dimensional Datasets of Mouse Lymph Nodes Resolve Organ-Wide Functional Dependencies

Jürgen Mayer,[1] Jim Swoger,[1] Aleksandra J. Ozga,[2] Jens V. Stein,[2] and James Sharpe[1]

[1] *EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG) and UPF, Dr. Aiguader 88,
08003 Barcelona, Spain*
[2] *Theodor Kocher Institut, Universität Bern, Freiestrasse 1, 3012 Bern, Switzerland*

Correspondence should be addressed to Jürgen Mayer, juergen.mayer@crg.es

Received 1 June 2012; Accepted 13 August 2012

Academic Editor: Huafeng Liu

Deep tissue imaging has become state of the art in biology, but now the problem is to quantify spatial information in a global, organ-wide context. Although access to the raw data is no longer a limitation, the computational tools to extract biologically useful information out of these large data sets is still catching up. In many cases, to understand the mechanism behind a biological process, where molecules or cells interact with each other, it is mandatory to know their mutual positions. We illustrate this principle here with the immune system. Although the general functions of lymph nodes as immune sentinels are well described, many cellular and molecular details governing the interactions of lymphocytes and dendritic cells remain unclear to date and prevent an in-depth mechanistic understanding of the immune system. We imaged *ex vivo* lymph nodes isolated from both wild-type and transgenic mice lacking key factors for dendritic cell positioning and used software written in MATLAB to determine the spatial distances between the dendritic cells and the internal high endothelial vascular network. This allowed us to quantify the spatial localization of the dendritic cells in the lymph node, which is a critical parameter determining the effectiveness of an adaptive immune response.

1. Introduction

In recent years, 3D optical imaging techniques have become popular (confocal microscopy, two-photon microscopy, SPIM, OPT, and multifocel microscopy) [1–4]. Most of them have become standard imaging techniques. There has also been a concurrent trend towards high-resolution imaging (SIM, STED, PALM, and STORM) [5–8]. These are all capable of producing huge amounts of data in form of 3D data sets. But as the questions asked and the data sets generated to address them are often very diverse, there are few general interpretation tools available. There are a number of standard image processing programs, both commercial (Imaris, Volocity, Huygens, etc.) [9–11], freeware (MevisLab, Drishti, etc.) [12, 13], as well as open source (Fiji, ImageJ, ICY, BioImageXD, etc.) [14–17]; however, they are mostly designed for visualization and some common analysis. Often the tools or algorithms required to analyze a given data set

are not available off-the-shelf and must be implemented in-house by the user. Quantification and spatial localization in 3D data sets can tell us about interactions of the structures imaged, which can be essential for deducing the mechanisms underlying features visible in the images. Pure data visualization in itself is often extremely informative, but quantification and statistical analysis of the data can indicate features or trends that are not readily apparent by visual inspection.

During adaptive immune responses, specialized cells located in the dermis called dendritic cells (DCs), use lymphatic vessels to transport pathogen-derived material to draining peripheral lymph nodes (PLNs), where they activate T lymphocytes located close to the center of lymphoid tissue (the T-cell zone or paracortex). DC migration from skin to PLN requires expression of the promigratory receptor CCR7, which responds to secreted protein ligands CCL19 and CCL21 expressed both in lymphatic vessels as well as in

the T-cell zone [18]. During their journey, dermal DCs need to cross the endothelial layer forming the lymphatic vessel lumen in the dermis. After a largely passive transport to the most proximal draining PLN, DCs cross a second barrier in the outer rim of PLN, the subcapsular sinus (SCS) [19]. CCR7 is thought to be involved in both the entry of dermal DCs into lymphatic vessels as well as their transmigration through the SCS into the deep T-cell zone. Experimentally, this process can be recapitulated by subcutaneous injection of *in vitro* activated DCs, which then migrate within the following 12–36 hours towards the nearest draining PLNs. Although such experiments uncovered an absolute requirement for CCR7 during this process, the precise contribution of CCR7 and its ligands for either of the two processes— lymphatic entry versus SCS egress—has thus far not been thoroughly investigated.

We use selective plane illumination microscopy (SPIM) to generate 3D voxel data sets of fluorescently labeled DCs and high endothelial venules (HEV) in mouse lymph nodes. To analyze this data, which has sufficient resolution to resolve single DCs, we have implemented an algorithm in MATLAB to quantify the spatial distributions of DCs and HEV within the intact lymph node. We have used this algorithm to compare the spatial distributions between wildtype mice and the plt/plt mutant, which lacks the ligands for the CCR7 receptor. This allows us to quantify not only the number of DCs in each PLN—a measure of their ability to migrate from the dermis to the lymph node—but also their distribution within the organ, which allows us to determine how effective the DCs are at finding their way to the T-cell zone.

2. Materials and Methods

To achieve high-resolution imaging deep within intact organs (millimeter-sized objects) or tissues [20], SPIM is the most convenient imaging technique [2]. This technique is based on illuminating fluorescent markers in the following way: a laser light sheet penetrates the whole sample, illuminating one plane in the imaged tissue. The emission signal of the fluorophores is detected orthogonal to the incoming light sheet. As only fluorophores located in the plane of the light sheet are excited, we attain optical sectioning. Scanning the sample through the light sheet results in a 3D voxel data stack. An important prerequisite for this kind of imaging is to have a transparent sample; as is the case for the lymph nodes investigated in this study, if the tissue is opaque in its native state, it has to be cleared chemically, enforcing *ex vivo* imaging. For more detailed information about SPIM, see [2]. Labeling distinct features in the sample with different fluorescent markers generates 3D datasets with multiple channels. We applied the SPIM technology to locate and quantify leukocyte subpopulations and vascular networks in intact murine lymphoid tissue in order to examine the usefulness of our imaging approach to a biologically relevant issue. In particular, we tested the relevance of CCR7 ligands in lymphatic vessels and in the deep T-cell area for efficient DC accumulation.

To quantify DC accumulation and their relation to the HEV, the distances to the center of the PLN and the distance to the closest neighboring HEV are calculated. The nearest neighbor calculation necessary for finding the closest HEV is done by creating and comparing voxel lists, lists of the coordinates of the centers of mass (CoMs) of the segmented DCs, and of all the segmented voxels within the HEV. After segmenting and filtering for intensity and size, for every object p of CH1 (DCs) the coordinates of its CoM is compared to the coordinates of all signal-containing voxels i of CH2 (HEV), and their respective distances are calculated as ($d_{p,i} = \sqrt{(x_p - x_i)^2 + (y_p - y_i)^2 + (z_p - z_i)^2}$ is the Euclidian distance between the CoM of object p (CoM$_p$) and voxel i of CH2). The minimum distance of every object p to a CH2 voxel, $\min_i(d_{p,i})$, is then used to construct the histogram. Note, that the method presented is fast only for sparse datasets (which applies here) but is not the most efficient regarding nearest neighbor search (NNS) in general (e.g., an Octree approach would be faster in denser cases). For details, see Section 3 . The calculation of the distance to the center of the lymph node works similarly: we assume the center of the PLN to be the CoM of the HEV (CoM$_{HEV}$) and then calculate dist (CoM$_p$, CoM$_{HEV}$) for every object p. CH2 effectively becomes a one-element voxel list, that one element being CoM$_{HEV}$.

Generation of dendritic cells from bone marrow (BMDCs) and PLN sample preparation was done as follows: DCs were generated by culturing bone marrow cells from tibias and femurs of C57BL/6 mice for 7-8 days in RPMI 1640/10% fetal calf serum (FCS) supplemented with FLT3 ligand as DC maturation agent. During the final 24 hours of culture BMDCs were stimulated with lipopolysaccharides (LPS) ($1 \mu g/mL$; Sigma). Activated BMDCs were harvested from the plates by vigorous pipetting and washing with phosphate-buffered saline (PBS) w/o Ca^{2+} and Mg^{2+}. Activation and phenotype of BMDCs were confirmed by flow cytometry staining with the DC markers CD11c (HL3), CD11b (M1/70), and CD86 (GL1) mAbs. Harvested BMDCs in RPMI1640/10% FCS were labeled with CellTracker Orange (CMTMR; $5 \mu M$ for 30 min), washed and injected subcutaneously (1×10^6 in $20 \mu L$ RPMI 1640/10% FCS) into the hind footpads of sex-matched recipient C57BL/6 mice or C57BL/6$^{plt/plt}$ mice. One day after BMDCs transfer, 10–$15 \mu g$ of Alexa594-coupled MECA-79 mAb was injected intravenously into recipient mice in order to visualize the HEV network in lymph nodes. Twenty minutes later, mice were sacrificed, the footpad-draining popliteal PLNs were excised and fixed in 0.4% PFA/PBS over night. Fixed PLNs were transferred to ice-cold PBS and cleaned carefully from the surrounding fat tissue under a stereomicroscope. Cleaned and fixed PLNs were stored in 0.1% sodium azide/PBS at 4°C before dehydration in methanol and clearing in BABB as described [21].

3. Results and Discussion

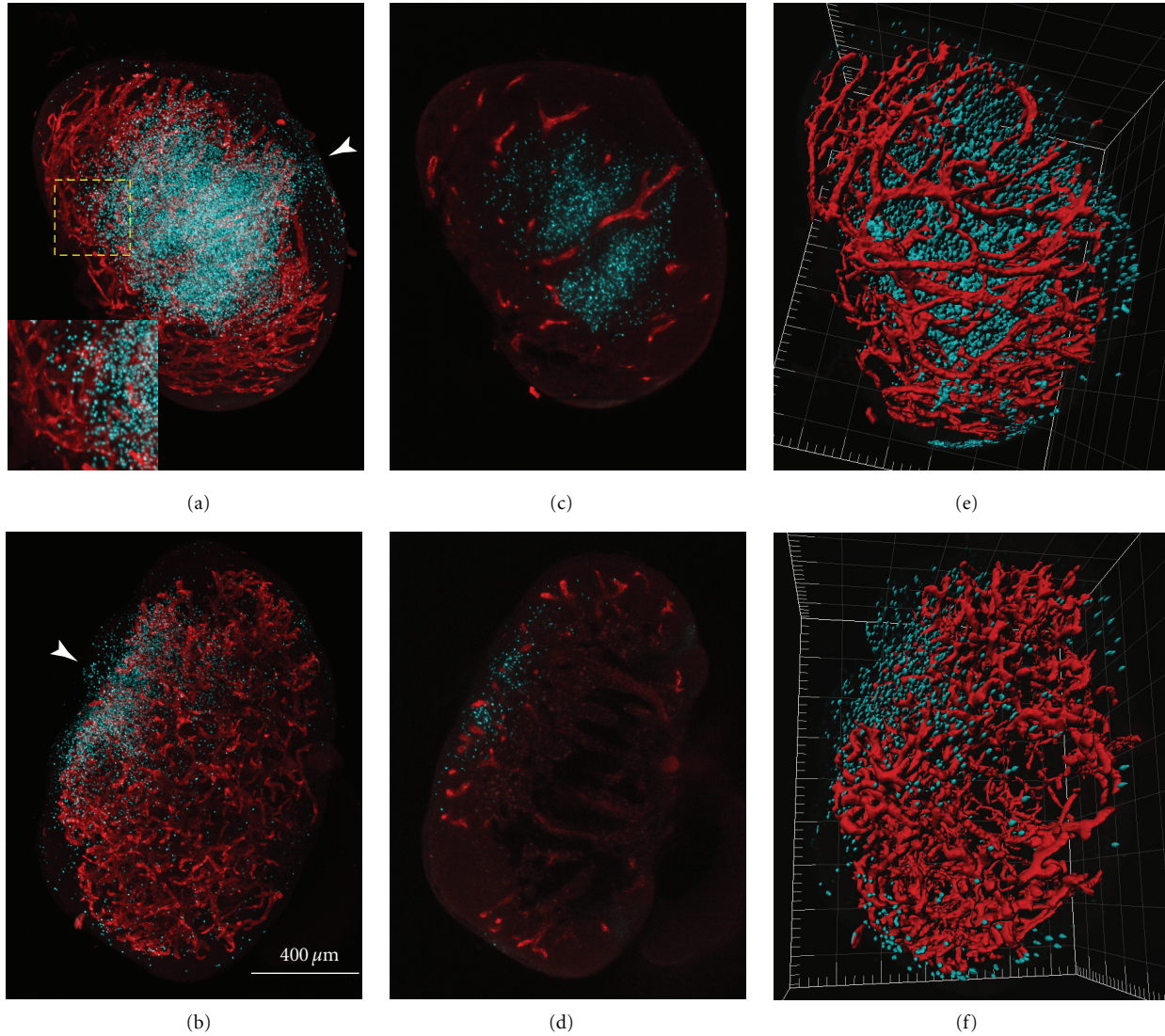We imaged entire draining PLNs isolated from mice that had received preactivated and fluorescently labeled DCs 18 h

FIGURE 1: Images of PLNs from wildtype (a), (c), and (e) and plt/plt mutant mice (b) and (d), (f). HEV, red; DCs, green. (a), (b) Maximum-value projections through the voxel data sets. The inset in (a) shows a higher magnification of the region indicated. White arrowheads show the approximate positions of the afferent lymphatic vessels through which the DCs enter the PLNs. (c) and (d) Representative optical slices through the central regions of the lymph nodes shown in (a) and (b). (e) and (f); surface renderings of the data in (a) and (b).

before the mice were sacrificed (our first channel: CH1). In addition, we imaged the PLN-specific extensive postcapillary network of HEVs as an anatomical landmark (our second channel: CH2). As seen in Figure 1, SPIM reconstructions clearly identified single DCs, which were mostly located towards the center of wildtype mouse PLNs. We were also able to detect a streak of DCs from the central accumulation to a location near the edge of the PLN, where the SCS is located. This streak most likely reflects the path that central DCs have taken after their arrival from the SCS towards the central T-cell zone.

Manual inspection of the single slices gives an idea of the relative distribution of the channels in this particular slice, but it is difficult to capture the 3D context (Figures 1(c) and 1(d)). A maximum value projection creates an image that shows more of the sample, but it is impossible to tell anything about the third dimension (Figures 1(a) and 1(b)). We developed an algorithm to quantify the distance of adoptively transferred DCs to the nearest point in the HEV network, as well as to the center of the PLN as a surrogate marker for accumulation in the deep T-cell area. To process the data, the algorithm pipeline (Figure 2) uses standard image processing methods, like 26-connectivity for segmentation of objects and CoM or Volume (V) calculations for the segmented objects. Intrinsic knowledge about the object size is used to discard signal which survived the intensity thresholding process. The choice of the threshold level influences the accuracy and precision of CoM estimation [22], but will be discussed later. In the left path of the processing pipeline (CH1, where the data of the DCs is processed), small objects

FIGURE 2: Flow diagram of the quantification algorithm. In the left path (green) dendritic cells are processed, in the right path (red), the vasculature data. The actual nearest neighbor search is done in the unified path (yellow). Abbreviations: center of mass (CoM), volume (V), peripheral lymph node (PLN).

either originate from pixel errors, tissue autofluorescence, or other artefacts, whereas large objects can be fluorescent impurities (e.g., dust) that have been embedded during the sample mounting procedure [23]. Both cases have to be excluded, leaving a window for objects in the size regime of DCs. We have been conservative by specifying a window of $3.2\,\mu m$ to $18.5\,\mu m$ for the DC diameter counting as a valid signal, as the average DC diameter is about $12 \pm 2\,\mu m$ [24]. In the right path in Figure 2 (CH2, HEV), there should be ideally only one object, because it is a network of connected tubes penetrating the whole PLN. As a result, small objects can be excluded in the downstream calculations for CH2.

To explore the impact of the receptors for CCR7 on immune function within the lymph node, we took advantage of the C57BL6$^{plt/plt}$ (plt/plt) mice, a naturally occurring mutant mouse strain, which lacks CCL19 and CCL21 protein in the T-cell zone in PLN, while maintaining expression of a second CCL21 isoform in lymphatic vessels. Although it has been previously reported that subcutaneously injected DCs accumulated with lower efficiency in the T-cell zone of plt/plt PLNs as compared to wild type PLNs [25], it remains unclear to what extent the decreased accumulation of DCs in plt/plt PLN was due to impaired entry into dermal lymphatic vessels versus decreased migration through the SCS barrier. Therefore, we performed experiments where equal numbers of DCs as used in wild type mice were injected into the footpads of plt/plt mice, followed animal sacrifice 18 h later and subsequent SPIM analysis.

Our whole-organ data revealed a significantly overall reduced number of DCs in plt/plt PLN, suggesting that migration of DCs into dermal lymphatic vessels was reduced in the plt/plt mice (in our case we have 5700 and 7934 identified cells for the wt/wt mice compared to 2318 and 427 cells in the plt/plt mice). More importantly, the 3D spatial distribution of DCs within plt/plt PLNs differed dramatically from wild types, with most DCs retained in the outer area in close proximity to or within the SCS. The distinct distributions for the wild type and plt/plt PLNs in the histograms in Figure 3 clearly show this effect. For the wild type (Figures 3(a) and 3(c)), the mean distance of the DCs from the center of the PLN is $330 \pm 130\,\mu m$, whereas for the plt/plt (Figures 3(b) and 3(d)) it is $500 \pm 150\,\mu m$. This result is clear despite the fact that plt/plt PLNs are generally smaller than those in wild type mice. Note that the specification of the center of the PLN is a rough estimate. It does not have to be precise in terms of specifying a real spatial center, so that first our assumption to take the HEV for calculations holds and secondly aiming for higher accuracy becomes obsolete.

To further quantify the difference between the behaviors of DCs in wild type and mutant PLNs, we assessed the distribution the DCs have concerning their minimal distance to the next HEV, so that we have a NNS-problem in a 3D dataset. The chosen method (comparing voxel coordinates) is a tradeoff between computational cost and implementation time. It is a type of linear searching and is the simplest approach to the NNS, as it has no spatial dependencies to consider (we operate in a one-dimensional space where only distances are considered). For the PLNs used in this study, generation of the samples and their preparation and

scanning were a process requiring several days, whereas the computational analysis performed with our algorithm typically required tens of minutes per sample. Thus, calculation time is not the limiting factor in throughput in our system, so a linear search algorithm is completely satisfactory. If applications arise requiring reduced computational time, a more efficient algorithm regarding computation time could easily be implemented such as the use of octrees [26], or, as we look for Euclidian distances, a method similar to the vp-tree algorithm [27, 28].

In the wild type, DCs are expected to migrate from their point of entry to paracortex (or T-cell zone), where they will interact with T-cells to enable an adaptive immune response. However, this migration is dependent on DC response to the chemoattractant CCL21, which is generated in the paracortex of wildtype PLN but absent in the plt/plt mutant. We observe that the DCs in plt/plt mice, which remain located close to their entry point—the peripheral SCS—have both a smaller mean distance to the HEV (Figures 4(b) and 4(d)) and a larger distance from the center of the lymph node (Figures 3(b) and 3(d)). They do not migrate away from the peripheral region with its high HEV density (it can be observed in Figure 1(d) that the HEV is very dense in the periphery of the plt/plt PLN). With $39 \pm 31\,\mu m$, the plt/plt mean distance is much smaller compared to the wt/wt mean distance of $66 \pm 31\,\mu m$. The broader distribution in the wt/wt (compare Figures 4(a) and 4(c) to Figures 4(b) and 4(d)) is caused by DCs located both in the deep T-cell zone in the core of the PLN (where the HEV network is not as dense as in the periphery) and near the SCS (where the network is very dense).

To validate our algorithm, we digitally dissected regions of interest from the PLOs that were small enough for manual counting of the DCs by visual inspection. Comparison of the manual counts with the results of our algorithm yielded agreement to within 2-3% (data not shown). We attribute the discrepancies to statistical variations in the DC signal levels and to the difficulty in distinguishing DCs that are very close together. In any event, these small errors in DC identification are not expected to have a significant effect on the results of our analysis. The error induced by the CoM calculation to determine positions of the DCs is small as we have an almost spherical shape and a highly specific staining with a pronounced signal. The thresholding process influences the discretization bias and the averaging bias [22], but the induced error is small compared to the measured distances and does not affect our conclusions. For the same reason, we did not aim for higher precision in localization using more sophisticated localization methods.

4. Conclusions

In order to draw conclusions about the mechanisms behind biological behaviors of interest, it is important to quantify cell distributions comprehensively in an organ-wide context. Quantification of spatial localization can be crucial for proper understanding of the mechanisms of interactions
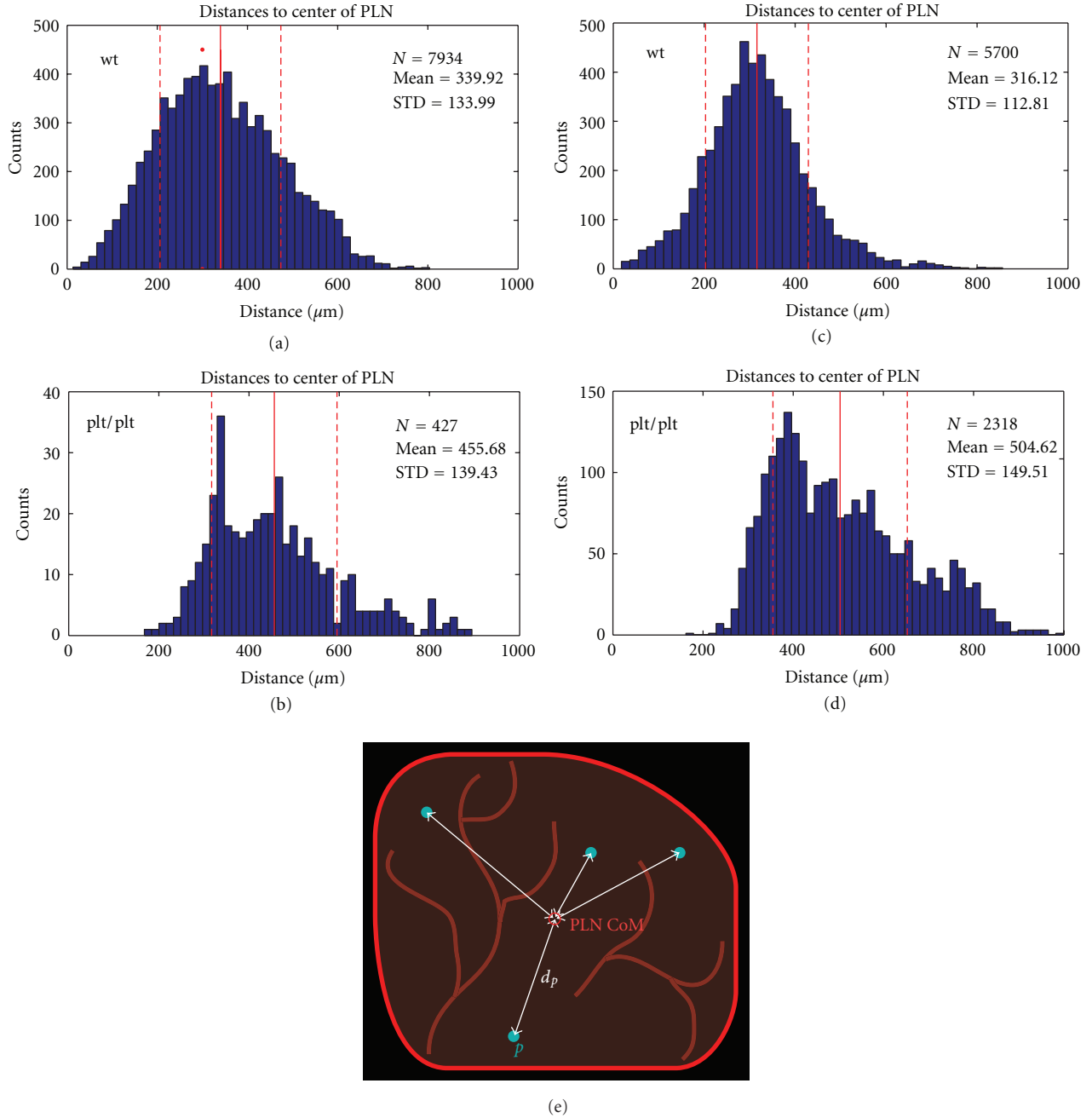
(a)

(c)

(b)

(d)

(e)

FIGURE 3: Histograms of the distances from the DCs to the center of the PLN. (a) and (c) two representative wt-PLNs, (b) and (d) plt-PLNs. The mean distance for the DCs in the wild type is smaller, indicating their location around the center of the PLN. In the plt/plt mutant, DCs remain close to their entry point in the periphery of the PLN. The red vertical line marks the mean distance, the dashed lines mark the standard deviation. (e) Schematic 2D depiction of the extent of the PLN (red) and DCs (cyan). The distances between the CoMs of the DCs and that of the PLN are indicated by white arrows.

between distinct components of an organ. We have demonstrated the application of quantitative computational analysis to an immunological question, which could be used to study 3D cell-based mechanisms in general. Fast and reliable statistics on 3D properties like the mutual distances of interacting members of distinct sets of cell types and/or morphological features in large data sets acquired in spatially extend organs

or tissues can help to reveal underlying mechanisms and their functional significance. Our data quantifies the differences in the behavior of DCs in wild type and plt/plt mouse PLNs. It indicates that both the transport of DCs to PLNs (as shown by the approximately 80% reduction of DCs found in plt/plt PLNs), and their subsequent migration within these organs (as indicated by the quantification of
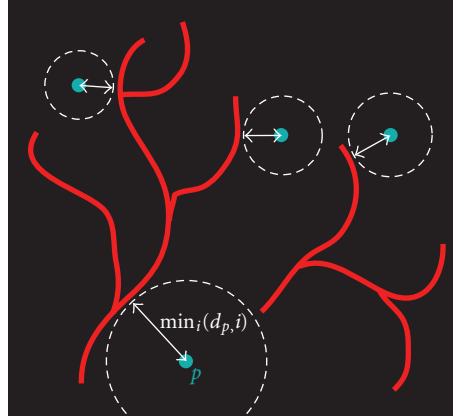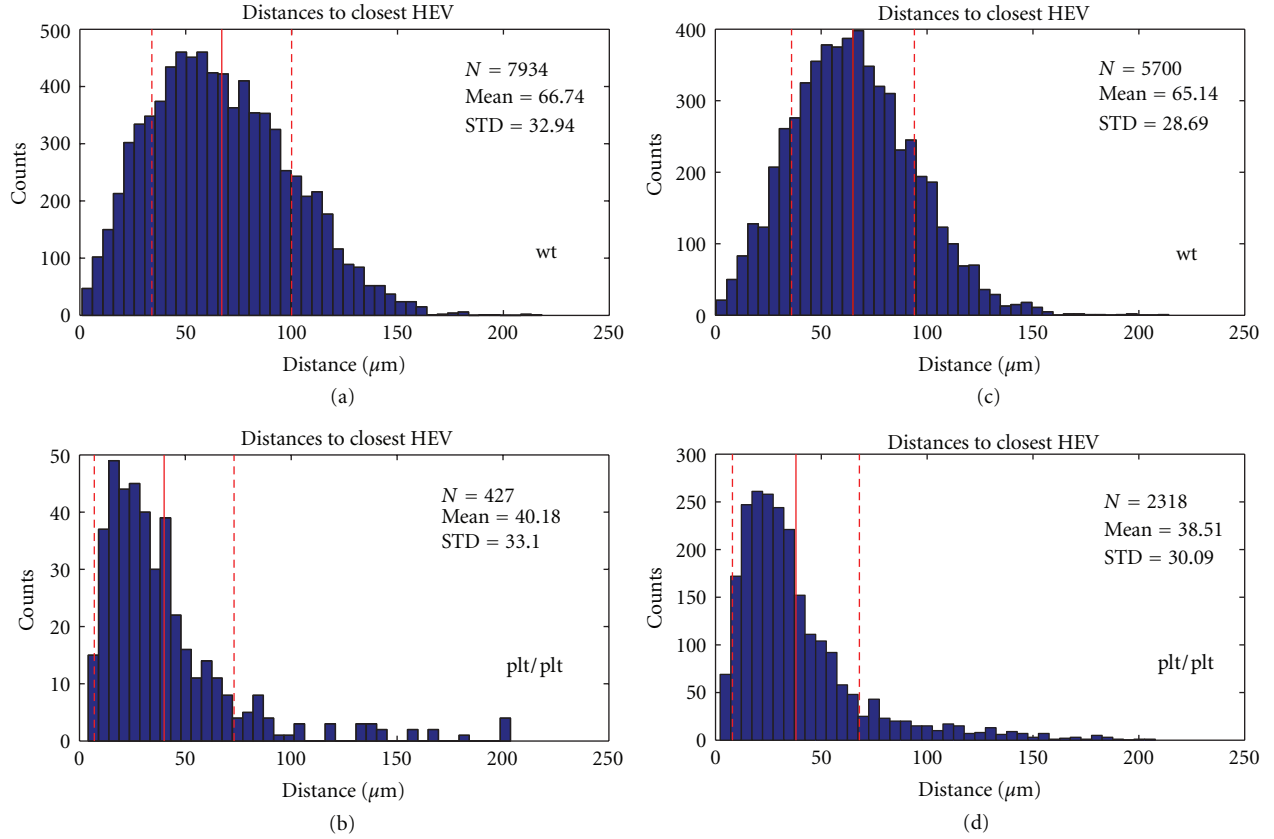
FIGURE 4: Histograms of the distances from the DCs to closest HEV, for the same PLNs as Figures 3(a) and 3(c) wt-PLN, (b) and (d) plt-PLN. Mean distance for DCs in wt-PLN is much larger than for the plt/plt mutant, reflecting the high density of the HEV in the periphery. The red vertical line marks the mean distance, the dashed lines mark the standard deviation. (e) Schematic 2D depiction of the HEV (red) and DCs (cyan). The minimum distances between the CoMs of the DCs and the HEV are indicated by white arrows and dashed circles.

the distribution of the DCs, in particular the approximately 50% increased distance from the center of the PLN) is compromised in the mutant mice. The approach can be applied to a wide range of biological and medical questions in context of disease mechanisms, especially considering that data from novel deep tissue imaging techniques such as SPIM are now becoming accessible. In particular, our results now make previously challenging venues of research feasible, such as the assessment of T-lymphocyte—DC interaction frequencies with physiologically low numbers of cells.

Acknowledgments

References

[1] J. B. Pawley, *Handbook of Biological Confocal Microscopy*, Springer, Singapore, 3rd edition, 2006.

[2] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. K. Stelzer, "Optical sectioning deep inside live embryos by selective plane illumination microscopy," *Science*, vol. 305, no. 5686, pp. 1007–1009, 2004.

[3] J. Sharpe, "Optical projection tomography," *Annual Review of Biomedical Engineering*, vol. 6, pp. 209–228, 2004.

[4] T. Zamofing and H. Hügli, "Applied multifocus 3D microscopy," in *Two- and Three-Dimensional Vision Systems for Inspection, Control, and Metrology*, Procceddings of SPIE, pp. 134–144, October 2003.

[5] L. Shao, P. Kner, E. H. Rego, and M. G. L. Gustafsson, "Super-resolution 3D microscopy of live whole cells using structured illumination," *Nature Methods*, vol. 8, no. 12, pp. 1044–1046, 2011.

[6] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, "Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 15, pp. 8206–8210, 2000.

[7] E. Betzig, G. H. Patterson, R. Sougrat et al., "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[8] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, pp. 793–795, 2006.

[9] Bitplane, Imaris 3D and 4D Real-Time Interactive Data Visualization, http://www.bitplane.com/.

[10] PerkinElmer, Volocity 3D Image Analysis Software, http://www.perkinelmer.com/pages/020/cellularimaging/products/volocity.xhtml.

[11] SVI, Huygens Scientific Volume Imaging, http://www.svi.nl/HomePage.

[12] Fraunhofer, MeVisLab, http://www.mevislab.de.

[13] A. Limaye, Drishti—Volume Exploration and Presentation Tool, 2006, http://sf.anu.edu.au/Vizlab/drishti/.

[14] Dev. FIJI, http://fiji.sc/wiki/index.php/Fiji.

[15] W. Rasband, ImageJ, http://rsbweb.nih.gov/ij/.

[16] Dev. ICY, http://icy.bioimageanalysis.org/.

[17] Dev. BioImageXD, http://www.bioimagexd.net/.

[18] U. H. Von Andrian and T. R. Mempel, "Homing and cellular traffic in lymph nodes," *Nature Reviews Immunology*, vol. 3, no. 11, pp. 867–878, 2003.

[19] R. Forster, A. Braun, and T. Worbs, "Lymph node homing of T cells and dendritic cells via afferent lymphatics," *Trends in Immunology*, vol. 33, no. 6, pp. 271–280, 2012.

[20] F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," *Nature Methods*, vol. 2, no. 12, pp. 932–940, 2005.

[21] J. Sharpe, U. Ahlgren, P. Perry et al., "Optical projection tomography as a tool for 3D microscopy and gene expression studies," *Science*, vol. 296, no. 5567, pp. 541–545, 2002.

[22] H. C. Van Assen, M. Egmont-Petersen, and J. H. C. Reiber, "Accurate object localization in gray level images using the center of gravity measure: accuracy versus precision," *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1379–1384, 2002.

[23] L. Quintana and J. Sharpe, "Preparation of mouse embryos for optical projection tomography imaging," *Cold Spring Harbor Protocols*, vol. 6, no. 6, pp. 664–669, 2011.

[24] D. J. Davidson, A. J. Currie, G. S. D. Reid et al., "The cationic antimicrobial peptide LL-37 modulates dendritic cell differentiation and dendritic cell-induced T cell polarization," *Journal of Immunology*, vol. 172, no. 2, pp. 1146–1156, 2004.

[25] M. D. Gunn, S. Kyuwa, C. Tam et al., "Mice lacking expression of secondary lymphoid organ chemokine have defects in lymphocyte homing and dendritic cell localization," *Journal of Experimental Medicine*, vol. 189, no. 3, pp. 451–460, 1999.

[26] R. Szeliski, "Rapid octree construction from image sequences," *CVGIP: Image Understanding*, vol. 58, no. 1, pp. 23–32, 1993.

[27] H. Xu and D. K. Agrafiotis, "Nearest neighbor search in general metric spaces using a tree data structure with a simple heuristic," *Journal of Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1933–1941, 2003.

[28] N. Kumar, L. Zhang, and S. K. Nayar, "What is a good nearest neighbors algorithm for finding similar patches in images?" in *Proceedings of the European Conference on Computer Vision (ECCV '08)*, pp. 364–378, Springer, 2008.

Research Article

Reliable and Efficient Approach of BOLD Signal with Dual Kalman Filtering

Cong Liu and Zhenghui Hu

State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China

Correspondence should be addressed to Zhenghui Hu, zhenghui@zju.edu.cn

By introducing the conflicting effects of dynamic changes in blood flow, volume, and blood oxygenation, Balloon model provides a biomechanical compelling interpretation of the BOLD signal. In order to obtain optimal estimates for both the states and parameters involved in this model, a joint filtering (estimate) method has been widely used. However, it is flawed in several aspects (i) Correlation or interaction between the states and parameters is incorporated despite its nonexistence in biophysical reality. (ii) A joint representation for states and parameters necessarily means the large dimension of state space and will in turn lead to huge numerical cost in implementation. Given this knowledge, a dual filtering approach is proposed and demonstrated in this paper as a highly competent alternative, which can not only provide more reliable estimates, but also in a more efficient way. The two approaches in our discussion will be based on unscented Kalman filter, which has become the algorithm of choice in numerous nonlinear estimation and machine learning applications.

1. Introduction

A thorough understanding of the dynamic relationship between cerebral blood flow (CBF), cerebral blood volume (CBV), and the blood oxygenation level dependent (BOLD) signal is essential for the physiological interpretation of fMRI activation data. The Balloon model described by Buxton et al. (1998) [1] is the first biomechanical plausible model to expound this relationship: increasing the flow (or perfusion rate) generally leads to dilution of venous deoxyhemoglobin (dHb), reducing the tendency of the blood to attenuate the magnetic resonance signal. The resultant increase in signal intensity is referred to as the BOLD response [2]. It is by extending this model to cover the dynamic coupling between CBF and synaptic activity, more sophisticated physiological realities are incorporated, for example, oxygen metabolism dynamics, both intra- and extravascular signal [3, 4], and more intricate models obtained.

The Balloon model is an input-state-output model with three state variables: blood flow, and volume, deoxyhemoglobin content and several biologically reasonable parameters. The problems of state estimation and parameter estimation (sometimes referred to as system identification or machine learning) associated with Balloon model are often formulated in a state-space representation, where Balloon model serves as a set of continuous-time system equations to describe the hemodynamic process. The equations are nonlinear, corresponding to the fact that Balloon model is one of the numerous nonlinear approaches to characterizing evoked hemodynamic response in fMRI.

Several work has utilized the Balloon model or its enhanced versions in the analysis of fMRI response. Some approaches, including expectation maximization (EM) [5, 6] and maximum likelihood [7], model the BOLD observation as deterministic hemodynamic process. However, a limitation of these methods is that they can only deal with measurement noise. Many promising approaches to the dual estimation problem belong to filtering algorithm that is able to account for both the physiological and measurement noise. Riera et al. [8] addressed the data assimilation problem in an extended Kalman filter (EKF) strategy. As EKF might lead to the problem of divergence due to linearized approximation, Johnston et al. proposed particle filter [9, 10] to avoid the flaw of linearization. Moreover, Hu et al. [11–13] employed unscented Kalman filter (UKF) that also outperforms EKF in terms of estimation error but with

roughly the same computational cost. Most recently, Friston et al. described variational filtering to optimize the approximation of posterior density on hidden model variables, while accumulating sufficient statistics to optimize the conditional densities of parameters and precision [14, 15].

The approaches mentioned above have greatly improved our ability to explore, and above all, to quantify the physiological mechanism involved in neural activation. However, they still have palpable defects. It is noteworthy that many of them actually are in the spirit of joint filtering, in which the underlying states and parameters are concatenated into a single higher dimensional joint state space, a filter runs for estimating both the states and parameters. Despite its straightforwardness in theory and convenience in implementation, the weakness of joint filtering is obvious. The objective of this paper is to introduce and develop an estimator equally concise but with higher performance—dual filtering. We will demonstrate the advantage of dual filter from two aspects, by the example of dual UKF versus joint UKF. (1) In terms of Balloon model, there is no inherent biophysical correlation between the states and parameters. By treating them separately, dual filtering can avoid undesired transaction between them. (2) Larger dimension of state-space vector implies much more computational expense. Specifically, computational complexity for general state-space problems is $\mathcal{O}(L^3)$ [16]. Although the frequency of predict-update cycle required by dual filter is the twice of that required by joint filter, dual estimate is much more computational efficient.

2. Materials and Methods

2.1. Hemodynamic Model. Balloon model describes the coupled kinetic changes from synaptic activity to the fMRI BOLD signal at a given region. This model has been extended by Friston et al. (2000) [5] to include the effects of external inputs to an autoregulated vasodilatory signal, assuming that the relationship between evoked neural activity and blood flow is linear. The subsequent work added different variations to this model, several of them were reviewed and integrated in Stephan et al. (2004) [18] and Buxton (2004) [19]. Based on fundamental physiology, rather than empirical approaches, these enhanced models are able to unify existing literature and provide insight into how the underlying physiological mechanisms result in stable or/and transit BOLD response. However, the original model proposed by Buxton et al. and completed by Friston et al. is sufficient to account for the nonlinear behaviors observed in real-time series [5]. Too many state variables and parameters will not serve our purpose here better.

The dynamic intertwinement between multiple physiological variables, the cerebral blood flow (CBF) f, blood venous volume v, and veins deoxyghemoglobin content q, can be given as a set of nonlinear nondimensional differential equations [1, 20]:

$$\ddot{f} = \epsilon u(t) - \frac{\dot{f}}{\tau_s} - \frac{f-1}{\tau_f},$$

$$\dot{v} = \frac{1}{\tau_0}\left(f - v^{1/\alpha}\right),$$

$$\dot{q} = \frac{1}{\tau_0}\left(f\frac{1-(1-E_0)^{1/f}}{E_0} - v^{1/\alpha}\frac{q}{v}\right), \quad (1)$$

where ϵ is neuronal efficacy, reflecting the significance of neuronal activity evoked by experimental event, hence it varies with trial event; τ_s and τ_f represent time constant for signal decay and autoregulatory feedback from blood flow, respectively. The existence of feedback term can be inferred from the poststimulus undershoots in CBF [21]. The degree of nonlinearity of the BOLD signal is largely determined by the stiffness parameter α, which characterizes the balloon-like capacity of the venous compartment to expel blood at a greater rate when distended [22]. E_0 is resting net oxygen extraction fraction. All variables are expressed in normalized form, relative to resting values.

Noticing that the first equation has a second-order time derivative, so we can write this input-state-output system as a set of first-order ordinary differential equations by introducing another variable $s = \dot{f}$. By defining the state vector as $\mathbf{x}(t) = [f, s, q, v]^T$, the system dynamic equation can be constructed from (1):

$$\dot{x} = f(\mathbf{x}, \theta, \mathbf{u}, \mathbf{v}) \quad v \sim N(0, \mathbf{Q}), \quad (2)$$

where $\theta = \{\epsilon, \tau_s, \tau_f, \tau_0, \alpha, E_0, V_0\} \in \mathbb{R}^l$ is system parameter, the neuronal input $\mathbf{u}$ represents system input, and $\mathbf{v}$ is to account for the process noise.

The observed signal can be taken as a nonlinear function of volume v and deoxyghemoglobin q that comprises a volume-weighted sum of intra- and extravascular signal:

$$y(t) = V_0\left(k_1(1-q) + k_2\left(1 - \frac{q}{v}\right) + k_3(1-v)\right),$$
$$k_1 = 7E_0, \qquad k_2 = 2, \qquad k_3 = 2E_0 - 0.2, \quad (3)$$

appropriate for a 1.5 tesla magnet [1]. V_0 is the resting blood volume fraction, which generally varies across brain regions and subjects. All parameters are independent of each other. Their physiological definitions and probability distributions are given in Table 1 [23].

The actual observation $\mathbf{y}$ is then composed of a deterministic part $h(x, \theta, t)$ and a stochastic part $\mathbf{w}$:

$$\mathbf{y} = h(\mathbf{x}, \beta, \mathbf{w}) \quad w \sim N(0, \mathbf{R}), \quad (4)$$

where $\mathbf{y}$ is the observation vector, $\mathbf{w}$ is measurement noise, and β consists of k_1, k_2, and k_3. Simultaneous estimation of V_0 and other parameters would be impossible, since their product is settled for each sampled measurement y_k. The stiffness parameter α is a nominal factor to BOLD contrast, it can be fixed to any value with its reasonable range in system identification [17].

Equation (2) describes a continuous-time hemodynamic process, and (4) models fMRI measurement as discrete sampling of the continuous system states, together they have formed a standard state-space representation for fMRI data assimilation. Given y_k, the physiological states $\mathbf{x}$ and the

TABLE 1: Hemodynamic model parameters and their probability distribution.

Notation	Definition	Distribution
ϵ	Neuronal efficacy	$\epsilon \sim N(0.54, 0.1^2)$
τ_s	Signal decay	$\tau_s \sim N(1.54, 0.25^2)$
τ_f	Autoregulation	$\tau_f \sim N(2.46, 0.25^2)$
τ_0	Transit time	$\tau_0 \sim N(0.98, 0.25^2)$
α	Stiffness parameter	$\alpha \sim N(0.33, 0.045^2)$
E_0	Resting oxygen extraction	$E_0 \sim N(0.34, 0.1^2)$
V_0	Resting blood volume fraction	$V_0 \sim N(0.02, 0.005^2)$

optimal parameters for a certain voxel can be estimated by use of UKF—a recursive minimum mean-square-error (MMSE) estimator.

2.2. Dual UKF and Joint UKF. The unscented Kalman filter has been applied extensively to the field of nonlinear estimation for both states and parameters. The basic framework of UKF involves estimation of the states of a discrete-time nonlinear dynamic system:

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k),$$
$$\mathbf{y}_k = \mathbf{H}(\mathbf{x}_k, \mathbf{n}_k), \tag{5}$$

where $\mathbf{x}_k$ represents the unknown system states, the system is driven by a known exogenous input $\mathbf{u}_k$ and process noise $\mathbf{v}_k$. The observation noise is given by $\mathbf{n}_k$.

The UKF generally involves recursive utilization of a deterministic "sampling" approach. The sampled points (sigma points) completely capture the true mean and covariance of the variables, and when propagated through the nonlinear system ($\mathbf{F}$ in this case), they are able to capture the posterior mean and covariance accurately to the 2nd order of Taylor series expansion [16].

Parameter estimation, or machine learning, on the other hand, involves determining a nonlinear mapping:

$$\mathbf{y}_k = \mathbf{G}(\mathbf{x}_k, \mathbf{w}), \tag{6}$$

where $\mathbf{x}_k$ is the input, $\mathbf{y}_k$ is the output, and the nonlinear map $\mathbf{G}(\cdot)$ is parameterized by the vector $\mathbf{w}$. Typically, a training set is provided by sample pairs consisting of known input and desired output, $\{\mathbf{x}_k, \mathbf{d}_k\}$. The goal of the learning can be expressed to some degree as solving for $\mathbf{w}$ which minimizes the error of the machine: $\mathbf{e}_k = \mathbf{d}_k - \mathbf{G}(\mathbf{x}_k, \mathbf{w})$. In order to estimate the parameters by utilizing UKF, a new state-space representation can be written:

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mathbf{r}_k,$$
$$\mathbf{d}_k = \mathbf{G}(\mathbf{x}_k, \mathbf{w}_k) + \mathbf{e}_k, \tag{7}$$

where the parameters $\mathbf{w}_k$ correspond to a stationary process with identity state transition matrix, driven by process noise $\mathbf{r}_k$.

Given that BOLD signal is the only output and observation of the system in terms of Balloon model, the dual estimation problem, in which the system states and model parameters must be estimated simultaneously, can be given as follows:

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, \mathbf{w}_k),$$
$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mathbf{r}_k, \tag{8}$$
$$\mathbf{y}_k = \mathbf{G}(\mathbf{x}_k, \mathbf{w}_k) + \mathbf{e}_k.$$

Since standard UKF cannot be applied to this system immediately, dual UKF and joint UKF have been proposed as two alternatives. In the dual filtering method, two UKFs—one for state estimation, the other for parameter estimation—run in an alternate way. At each time step, the current estimate of parameters $\hat{\mathbf{w}}_k$ is used in the state filter as given input, and likewise the current states estimate $\hat{\mathbf{x}}_k$ is used in the parameter filter. On the contrary, a single UKF runs for both state and parameter estimation in the joint filtering. A higher dimensional joint state vector is defined: $\tilde{\mathbf{x}}_k = [\mathbf{x}_k^T \mathbf{w}_k^T]^T$, and the state-space model is reformed as follows:

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{F}}(\hat{\mathbf{x}}_k, \mathbf{u}_k, \hat{\mathbf{v}}_k),$$
$$\mathbf{y}_k = \mathbf{G}(\hat{\mathbf{x}}_k, \mathbf{n}_k). \tag{9}$$

The dual UKF and joint UKF approaches are illustrated in Figure 1.

In this section, the framework of UKF is briefly reviewed, a dual estimation problem with two approaches have been presented. In the next section, we will focus on examining the different performances of dual UKF and joint UKF, and all of our discussion will be in the context of Balloon model.

3. Results and Discussion

3.1. Biophysical Interpretation. One of the most prominent bifurcations between dual estimate and joint estimate is whether to incorporate interaction or correlation between states and parameters into filtering. As discussed earlier, the joint filter concatenates the state and parameter random variables into a single augmented state $(\hat{\tilde{x}}_k)$, so the cross-covariance between states and parameters is effectively modeled, that is,

$$E\left[\left(\hat{\tilde{x}} - E\left[\hat{\tilde{x}}\right]\right) \left(\hat{\tilde{x}} - E\left[\hat{\tilde{x}}\right]\right)^T \right] = \begin{bmatrix} \mathbf{P}_{x_k x_k} & \mathbf{P}_{x_k w_k} \\ \mathbf{P}_{w_k x_k} & \mathbf{P}_{w_k w_k} \end{bmatrix}. \tag{10}$$

Dual filtering, on the other hand, decouples (in a statistical sense) the dual estimation problem by treating states and parameters separately, which means $\mathbf{P}_{x_k w_k} = \mathbf{P}_{w_k x_k} = 0$. For states and parameters involved in Balloon model, no dynamic interaction or biophysical correlation between them has been observed (they are uncorrelated variables), therefore, it is reasonable to expect dual filtering to exhibit more biophysical accuracy. Thus the fMRI experiments substantiated our assumption.

The real fMRI data was acquired from 8 health subjects. 136 acquisitions in total were made (RT = 2s), in block of 8, giving 16 16-second blocks. The condition for successive
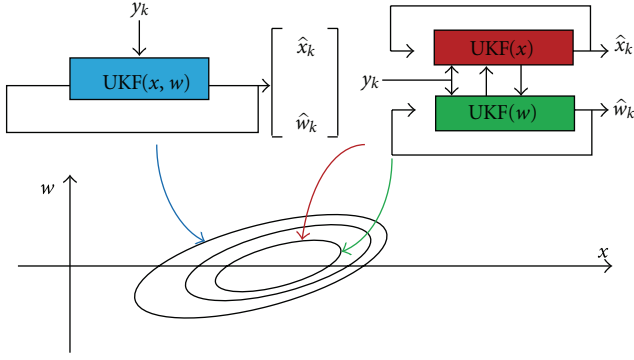
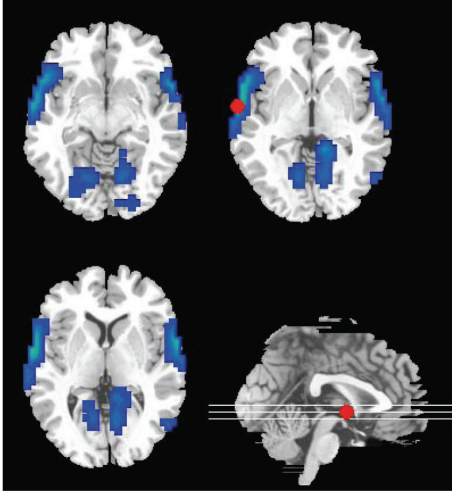FIGURE 1: Schematic diagrams of joint filter (left) and dual filter (right).

Algorithm	Dimension of state vector	Total flops
Joint UKF	$L = L_x + L_w = 9$	285
Dual UKF	$L_x = 4, L_w = 5$	170

overshoot, followed by gradual return to reduced plateau, and ending with a strong poststimulus undershoot. On the other hand, joint UKF fails to some degree to reconstruct a clean BOLD signal: the plateau is missing, neither does the evolving pattern of the signal show itself in a stable way in each block. The overestimated transit time (τ_0) leads to a reduction in amplitude of the BOLD peak; the less intense poststimulus undershoot can be explained by the underestimated signal decay (τ_s). Given the fact that dual UKF and joint UKF have very similar performance for state estimation, which is made clear in Figure 3, it is safe to attribute the failure to the undesired fluctuations of parameters (especially E_0, which affects the estimated signal directly by (3)), or more precisely, the undesired interaction between parameters and states.

FIGURE 2: The greatest activated area of the group in the superior temporal gyrus (GT) for data assimilation.

3.2. Computational Interpretation. One of the most computationally expensive operations in UKF corresponds to calculating the new set of sigma points at each time update. This requires taking a matrix squareroot of the state covariance matrix, $\mathbf{P} \in \mathbb{R}^{L \times L}$, given by $\mathbf{SS}^T = \mathbf{P}$ [16]. An efficient implementation using a Cholesky factorization requires in general $\mathcal{O}(L^3/6)$ computations [26]. Therefore enlarging the dimension of state-space vector will dramatically increase the computational complexity (also can be referred to as time complexity). In this subsection we will introduce two criteria for evaluating the property of dual UKF and joint UKF in time complexity.

Number of Floating-Point Operations (flops). In computing, floating point can be thought of as a computer realization of scientific notation, which is able to represent a wide range of values. Flops number required by a given algorithm or computer program is independent of the computing platform, although its precise value may differ under different counting rules. MATLAB (version before 6.0) has provided us with a useful function *flops* to specify the cumulative number of flops. For instance, if A and B are N-by-N matrixes, then the output of $flops(A + B)$ and $flops(AB)$ will approximately be N^2 and $2N^3$.

Since Cholesky factorization is the only operation within UKF whose time complexity is proportional to N^3, it is appropriate to consider that the flops number of Cholesky factorization is sufficient to determine the flops of the whole algorithm. For Balloon model, the dual estimation problem is about determining four state variables ($L_x = 4$) and five parameters (all the parameters except V_0 and α, $L_w = 5$) at each time step. Table 2 shows the total flops number of Cholesky factorization involved in each predict-update cycle of UKF.

blocks alternated between rest and auditory stimulation, starting with rest. Auditory stimulation was emotionally neutral words presented at a rate of 60 per minute. We selected the largest activated voxels in superior temporal gyrus (GT) to implement data assimilation [24] (Figure 2). Bias correction was performed using the method in [25]. The two algorithms were initialized in identical way on experimental data and parameters.

Figure 3 shows the hemodynamic states given by joint UKF and dual UKF. The lower peak of blood flow inferred from joint UKF corresponds to the smaller neuronal efficacy (ϵ) in Figure 4.

Parameters estimated are shown in Figure 4. Signal decay, autoregulation et al. remain unchanged (almost) during dual filtering. While for joint UKF, the parameters do not converge to their final values until the 4th $\sim$ 5th block (60 $\sim$ 80 s after the first stimulation). This phenomenon is a strong indicator for the introduced interaction between states and parameters.

Real and estimated fMRI signals are plotted in Figure 5. The simulated BOLD signal given by dual UKF shows a slight
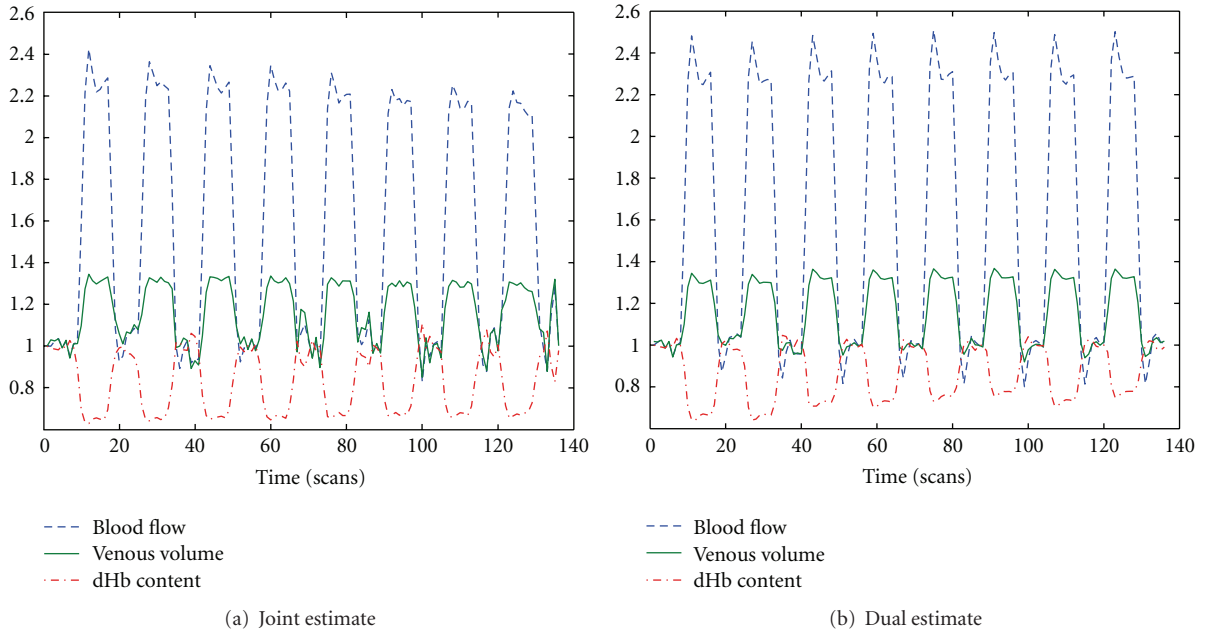
(a) Joint estimate

(b) Dual estimate

FIGURE 3: States estimated by dual and joint UKF. The dotted line corresponds to change of blood flow, the solid line shows venous volume and dashed line depicts dHb content.
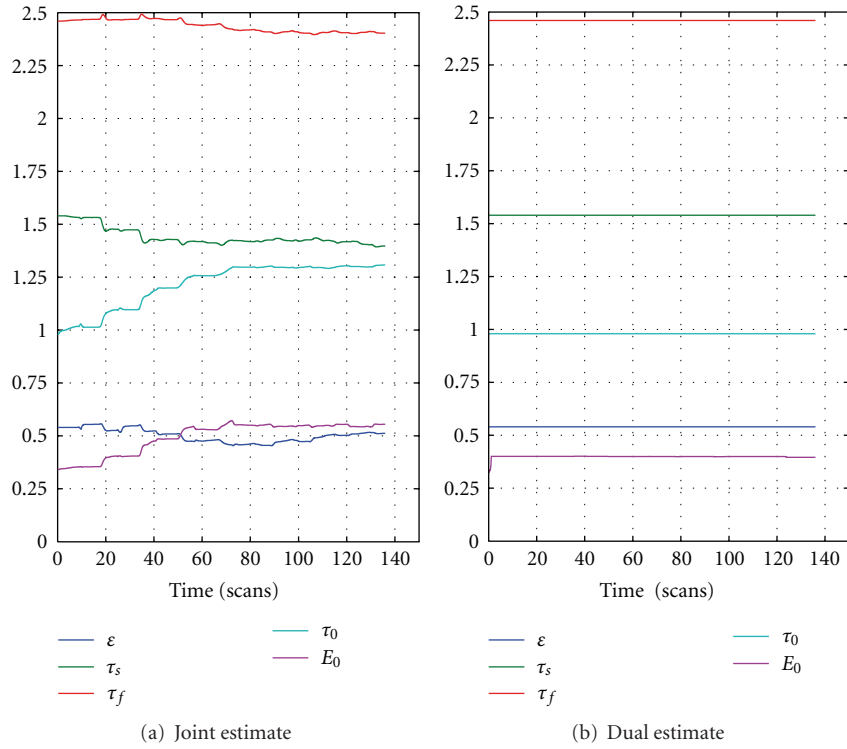

(a) Joint estimate

(b) Dual estimate

FIGURE 4: Parameters estimated by joint UKF and dual UKF. The mean values in Table 1 are used as initial values in our simulation. Apart from E_0, which grows sharply at the beginning and does not change afterwards, all the parameters obtained from dual UKF can be seen as constants. Resting oxygen extraction V_0 is the most important parameter in driving the model uncertainty [17], but simultaneous estimation of V_0 and other parameters would be impossible, as stated earlier. α is a nominal mechanism to BOLD signal. Therefore V_0 and α does not enter filtering.
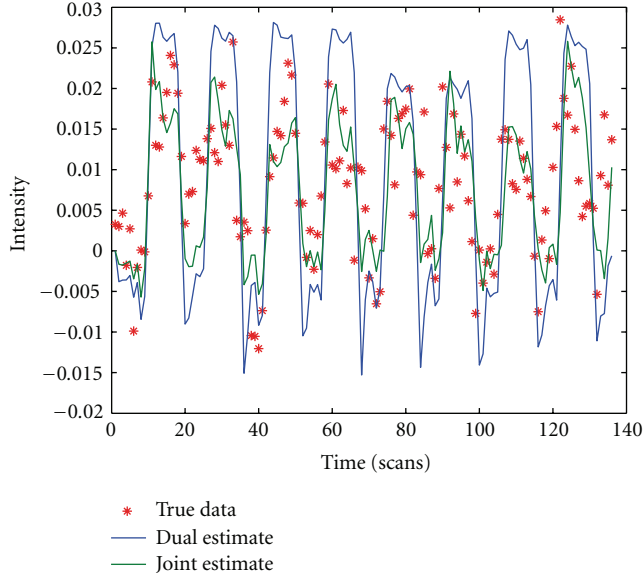
FIGURE 5: Estimated and real fMRI signal.

TABLE 3: Flops number for augmented state vector.

Algorithm	Dimension of state vector	Total flops
Joint UKF	$L = L_x + L_w = 19$	2470
Dual UKF	$L_x = 9, L_w = 11$	1582

However, in practice by slightly restructuring the state vector, the process and observation models, we may introduce the noise with the same order of accuracy as the uncertainty in the state. First, the state vector is augmented to give a $L^\alpha = L_x + L_v + L_n$ dimensional vector;

$$\mathbf{x}_k^\alpha = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{v}_k \\ \mathbf{n}_k \end{bmatrix}. \tag{11}$$

Then the process model is rewritten as a function of $\mathbf{x}_k^\alpha$, $\mathbf{x}_{k+1}^\alpha = \mathbf{F}(\mathbf{x}_k^\alpha, \mathbf{u}_k)$; the unscented transform uses sigma points that are drawn from

$$\mathbf{P}^\alpha = \begin{bmatrix} \mathbf{P}_x & 0 & 0 \\ 0 & \mathbf{R}_v & 0 \\ 0 & 0 & \mathbf{R}_n \end{bmatrix}, \tag{12}$$

where $\mathbf{R}_v$ and $\mathbf{R}_n$ are the process and observation noise covariance. In this situation, similarly we can derive Table 3.

For either case mentioned above, the flops number for joint filtering is at least 56% larger than that for dual filtering.

Comparing to flops count, overall execution time is a more tangible and practical criterion. We have tested our programs on several computers and collected their execution time data. Normally, a difference over 90% can be observed (16 s and 30 s for dual UKF and joint UKF, resp.). This result is even more impressive than that related to flops analysis, indicating that flops number is not the only factor influencing the operation time. Even if we take the rapid improvements in processing speed and memory into consideration, this variance is significant and should not be ignored.

4. Conclusions

In this paper we brought forward the dual Kalman filter as a reliable and efficient approach to estimating the states and parameters involved in balloon model. Comparing to the commonly used joint Kalman filter, its principle is in better conformity with the physiological reality, and by decoupling the dual estimation problem, it is much more calculational efficient to implement. The result of experiments showed good agreement with our conclusion.

Acknowledgments

References

[1] R. B. Buxton, E. C. Wong, and L. R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: the balloon model," *Magnetic Resonance in Medicine*, vol. 39, no. 6, pp. 855–864, 1998.

[2] R. D. Hoge, J. Atkinson, B. Gill, G. R. Crelier, S. Marrett, and G. Bruce Pike, "Investigation of bold signal dependence on cerebral blood flow and oxygen consumption: the deoxyhemoglobin dilution model," *Magnetic Resonance in Medicine*, vol. 42, pp. 849–863, 1999.

[3] Y. Zheng, J. Martindale, D. Johnston, M. Jones, J. Berwick, and J. Mayhew, "A model of the hemodynamic response and oxygen delivery to brain," *NeuroImage*, vol. 16, no. 3, pp. 617–637, 2002.

[4] T. Obata, T. T. Liu, K. L. Miller et al., "Discrepancies between BOLD and flow dynamics in primary and supplementary motor areas: application of the balloon model to the interpretation of BOLD transients," *NeuroImage*, vol. 21, no. 1, pp. 144–153, 2004.

[5] K. J. Friston, A. Mechelli, R. Turner, and C. J. Price, "Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics," *NeuroImage*, vol. 12, no. 4, pp. 466–477, 2000.

[6] K. J. Friston, "Bayesian estimation of dynamical systems: an application to fMRI," *NeuroImage*, vol. 16, no. 2, pp. 513–530, 2002.

[7] T. Deneux and O. Faugeras, "Using nonlinear models in fMRI data analysis: model selection and activation detection," *NeuroImage*, vol. 32, no. 4, pp. 1669–1689, 2006.

[8] J. J. Riera, J. Watanabe, I. Kazuki et al., "A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals," *NeuroImage*, vol. 21, no. 2, pp. 547–567, 2004.

[9] L. A. Johnston, E. Duff, and G. F. Egan, "Particle filtering for nonlinear BOLD signal analysis," *Lecture Notes in Computer Science*, vol. 4191, pp. 292–299, 2006.

[10] L. A. Johnston, E. Duff, I. Mareels, and G. F. Egan, "Nonlinear estimation of the BOLD signal," *NeuroImage*, vol. 40, no. 2, pp. 504–514, 2008.

[11] Z. H. Hu and P. C. Shi, "Nonlinear analysis of bold signal: biophysical modeling, physiological states, and functional activation," in *Proceedings of the 10th Internation Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI '07)*, pp. 734–741, Brisbane, Australia, 2007.

[12] P. Shi, Z. Hu, X. Zhao, and H. Liu, "Nonlinear analysis of the BOLD signal," *Eurasip Journal on Advances in Signal Processing*, vol. 2009, Article ID 215409, 2009.

[13] Z. Hu, C. Liu, P. Shi, and H. Liu, "Exploiting magnetic resonance angiography imaging improves model estimation of bold signal," *PloS One*, vol. 7, no. 2, Article ID e31612, 2012.

[14] K. J. Friston, N. Trujillo-Barreto, and J. Daunizeau, "DEM: a variational treatment of dynamic systems," *NeuroImage*, vol. 41, no. 3, pp. 849–885, 2008.

[15] J. Daunizeau, K. J. Friston, and S. J. Kiebel, "Variational bayesian identification and prediction of stochastic nonlinear dynamic causal models," *Physica D*, vol. 238, no. 21, pp. 2089–2118, 2009.

[16] R. van der Merwe and E. A. Wan, "The square-root unscented Kalman filter for state and parameter-estimation," in *Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 6, pp. 3461–3464, May 2001.

[17] Z. H. Hu and P. C. Shi, "Sensitivity analysis for biomedical models," *IEEE Transactions on Medical Imaging*, vol. 29, no. 11, pp. 1870–1881, 2010.

[18] K. E. Stephan, L. M. Harrison, W. D. Penny, and K. J. Friston, "Biophysical models of fMRI responses," *Current Opinion in Neurobiology*, vol. 14, no. 5, pp. 629–635, 2004.

[19] K. Uludağ, R. B. Buxton, D. J. Dubowitz, and T. T. Liu, "Modeling the hemodynamic response to brain activation," *NeuroImage*, vol. 23, supplement 1, pp. S220–S233, 2004.

[20] R. B. Buxton and L. R. Frank, "A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation," *Journal of Cerebral Blood Flow & Metabolism*, vol. 17, no. 1, pp. 64–72, 1997.

[21] K. Irikura, K. I. Maynard, and M. A. Moskowitz, "Importance of nitric oxide synthase inhibition to the attenuated vascular responses induced by topical L-nitroarginine during vibrissal stimulation," *Journal of Cerebral Blood Flow & Metabolism*, vol. 14, no. 1, pp. 45–48, 1994.

[22] D. E. Glaser, K. J. Friston, A. Mechelli, R. Turner, and C. J. Price, "Haemodynamic modelling," *Human Brain Function*, Elsevier Press, London, UK, 2003.

[23] R. L. Bellaire, E. W. Kamen, and S. M. Zabin, "A new nonlinear interated filter with application to target tracking," in *Proceedings of the AeroSense: 8th International Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, vol. 2561, pp. 240–251, 1995.

[24] Z. Hu, P. Ni, C. Liu, X. Zhao, H. Liu, and P. Shi, "Quantitative evaluation of activation state in functional brain imaging," *Brain Topography*. In press.

[25] Z. Hu, H. Liu, and P. Shi, "Concurrent bias correction in hemodynamic data assimilation," *Medical Image Analysis*. In press.

[26] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, 2 edition, 1992.

Research Article

A Meshfree Method for Simulating Myocardial Electrical Activity

Heye Zhang,[1] Huajun Ye,[2] and Wenhua Huang[3]

[1] *Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, Shenzhen 518055, China*
[2] *Department of Optical Engineering, Zhejiang University, Hangzhou 310027, China*
[3] *Institute of Clinical Anatomy, Southern Medical University, Guangzhou 510515, China*

Correspondence should be addressed to Wenhua Huang, hwh@fimmu.com

An element-free Galerkin method (EFGM) is proposed to simulate the propagation of myocardial electrical activation without explicit mesh constraints using a monodomain model. In our framework the geometry of myocardium is first defined by a meshfree particle representation that is, a sufficient number of sample nodes without explicit connectivities are placed in and inside the surface of myocardium. Fiber orientations and other material properties of myocardium are then attached to sample nodes according to their geometrical locations, and over the meshfree particle representation spatial variation of these properties is approximated using the shape function of EFGM. After the monodomain equations are converted to their Galerkin weak form and solved using EFGM, the propagation of myocardial activation can be simulated over the meshfree particle representation. The derivation of this solution technique is presented along a series of numerical experiments and a solution of monodomain model using a FitzHugh-Nagumo (FHN) membrane model in a canine ventricular model and a human-heart model which is constructed from digitized virtual Chinese dataset.

1. Introduction

Myocardial contraction is driven by a sequence of propagating electrical activations throughout the myocardium [1]. Propagation of electrical activations inside the myocardium is a highly complicated process mainly due to the fibrous structure of myocardium, as shown in many experiments [2]. There have been efforts in simulating myocardial electrical activations using computational models with known physical parameters, including the source intensities and locations, material properties, and boundary conditions, because these simulations can help to understand the measurement data, suggest new experiments, and provide insights into the basic mechanism of electrical activity in the heart. A number of computational models have been developed to simulate the macroscopic electrical propagation process [3, 4], such as cellular automata and reaction-diffusion systems. A cellular automaton is a discrete model which usually consists of a regular grid of cells, each in one of a finite number of states. Every cell has the same rule for updating, based on the states in its neighborhood. Because the simplicity of states and superior computational speed resulted from rules, cellular

automata have been applied in simulations of myocardial electrical activity in the heart [5, 6], but such simplistic and rule-based approaches cannot always properly capture the shape of transmembrane potentials. A reaction-diffusion system is a mathematical model that describes how the concentration of one or more substances distributed in space changes under the influence of two processes: local reactions in which the substances are converted into each other and diffusion which causes the substances to spread out in space. This concept in the reaction-diffusion system is borrowed and applied in the simulation of myocardial electrical activity by turning local reactions into cellular models, that is, ionic currents, and diffusion into transmission of transmembrane potentials, that is, anisotropic propagation through myofibers. Though the reaction-diffusion system can more appropriately reproduce electrical activity of excitable myocardium [3, 4, 7] than cellular automata, solving a reaction-diffusion system is computationally expensive with realistic modelling of cardiac tissue properties and cellular behaviors [3]. Recently, the Eikonal model [8, 9], which is a simplified wavefront model, has also been solved by FEMs in order to simulate anisotropic electrical activity

across myocardium [10]. The computational models have been widely applied in understanding patients' data [11–15].

In the context of modelling electrical activity in the heart, one of challenges is to establish a numerical representation of the complex geometry of the heart. This representation must not only characterize geometric complexities but also be of sufficient resolution to capture the activation wavefront and perhaps cellular behaviours. In order to properly simulate myocardial electrical activity by solving reaction-diffusion systems accurately, a large number of numerical schemes have been developed by representing the intrinsic structure of the myocardium and the inhomogeneity/anisotropy in different ways. By discretizing the diffusion tensor over the problem domain, traditional FDMs can evolve electrical activity over orthogonal and regular grids [16], but the complex geometry of heart is always a great challenge for FDMs. Thus, a few works have proposed the use of irregular grids with FDMs to deal with complex geometry by increasing the complexity of interpolation between grids [7, 17–19]. In FEMs, the integral form of the reaction-diffusion system is discretized over a finite element representation of the geometry. Typically low-order (linear Lagrange basis) elements used in FEMs [20, 21] always lead to a high number of elements in the complex geometry and a long time integration for a certain accuracy or remesh in changing geometry, such as a beating heart. Therefore, high-order elements, such as cubic Hermite basis elements [22] and quadratic Lagrange basis elements, use more nodal parameters or nodes inside one element to get better accuracy, but the size of system matrix and the computational load are also increased largely. Furthermore, meshing or remeshing for FEMs using high-order elements still remains challenge.

EFGM is developed as a meshfree method in 1990s [23] and has been successfully applied for a wide range of mechanical applications [24, 25]. A series of publications [26, 27] have explored the numerical capabilities of EFGM, including parallelization and comparison with FEMs in mechanical applications. Meshfree method has been applied into simulation of myocardial electrical activity by authors [28, 29]. However, the numerical performance of meshfree method has not been well verified in the simulation of myocardial electrical activity. Furthermore, the previous work [28, 29] only used left ventricle segmented from MR images with spurious fiber structure. Our aim of this paper is to present EFGM as a computational tool to solve reaction-diffusion systems for the simulation of myocardial electrical activity. In this paper a new representation of myocardial geometry and fiber structure by a cloud of nodes without any explicit connectivity defined between them, that is, meshfree particle representation, is first discussed. Upon this representation, the numerical performance of EFGM in solving the monodomain model [3, 4], a reaction-diffusion system, is demonstrated through experiments. The properties processed by EFGM provide quite a few advantages, such as refinement can be accomplished by adding or removing nodes in particular areas [23–25]. Moreover, fiber orientation is interpolated with nodal parameters, not inside the element any more. Hence, all the operations inside the element of FEMs, such as coordinate transform

from elemental coordinate to global coordinate and the interpolation of elemental fiber orientation, are also avoided in this approach. Furthermore, higher order approximation of a meshfree shape function can be achieved without rearranging nodal positions or adding extra degrees of freedom in nodes for example, higher consistency and continuity can be still maintained over the whole problem domain, even with a linear basis, in EFGM [23–25]. Though it has been demonstrated that EFGM can also handle material inhomogeneities and discontinuities in mechanical applications [24, 25], we cannot cover that in this paper because of limited space.

Governing equations of electrical activity over the myocardium are discussed in Section 2. A numerical scheme based on EFGM in terms of representation, shape function and the Galerkin weak form is established in Section 3. Numerical experiments are presented and compared in Section 4. And finally in Section 5, we discuss the strengths and weaknesses of the current approach and state possible future directions.

2. Governing Equations

The bidomain model [3, 4], a popular reaction-diffusion system, divides the myocardium into intracellular and extracellular space. Both spaces can be described by the same coordinate system and are separated by the membrane at each location:

$$\nabla \cdot ((\mathbf{D}_i + \mathbf{D}_e)\nabla v_e) = -\nabla \cdot (\mathbf{D}_i \nabla v_m) + I_{s1}, \qquad (1)$$

$$\nabla \cdot (\mathbf{D}_i \nabla v_m) + \nabla \cdot (\mathbf{D}_i \nabla v_e) = A_m \left(C_m \frac{\partial v_m}{\partial t} + I_{\text{ion}} \right) - I_{s2}. \qquad (2)$$

v_m is the transmembrane potential, v_e is the extracellular potential, $\mathbf{D}_i$ is the conductivity in intracellular space, $\mathbf{D}_e$ is the conductivity in extracellular space, A_m is the ration of the membrane surface area to the volume, C_m is the membrane capacitance, I_{ion} is sum of ionic currents, and I_{s1} and I_{s2} are external stimulus currents. There are a lot of cellular ionic models [3, 4] that could be used in reaction-diffusion system. If the conductivity in extracellular domain is assumed to be infinitely large, or the conductivities of extracellular and intracellular domains are assumed to be equally anisotropic, for example, $\mathbf{D}_i = k \cdot \mathbf{D}_e$, a bidomain model can be reduced to a monodomain model, which turns (1) and (2) into a single equation:

$$\nabla \cdot (\mathbf{D}\nabla v_m) = A_m \left(C_m \frac{\partial v_m}{\partial t} + I_{\text{ion}} \right) - I_s, \qquad (3)$$

with natural boundary condition $(\mathbf{D}\nabla v_m) \cdot \mathbf{n} = 0$ if heart is considered as an isolated continuum. I_s is external stimulus current, and $\mathbf{D}$ is the conductivity. The conductivity variables, $\mathbf{D}_e$, $\mathbf{D}_i$, or $\mathbf{D}$, at each point in space, are represented by a tensor containing coefficients along and across fiber
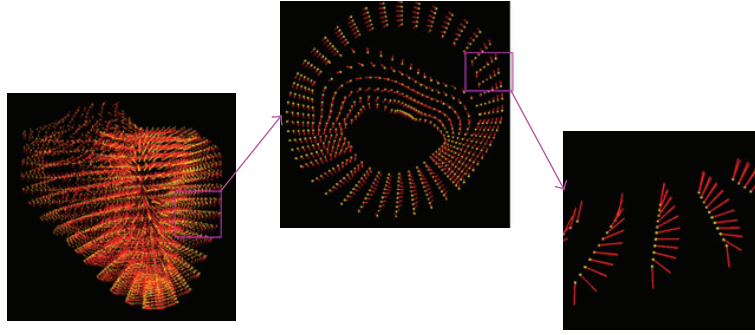
FIGURE 1: Meshfree representation of Auckland heart model.

orientation in that point. Let $\mathbf{D}_{\text{local}}$ be a diffusion tensor of a point in local fiber coordinate; then in 3D

$$\mathbf{D}_{\text{local}} = \begin{pmatrix} \sigma_f & 0 & 0 \\ 0 & \sigma_{\text{cf}} & 0 \\ 0 & 0 & \sigma_{\text{cf}} \end{pmatrix}, \tag{4}$$

where σ_f along the fiber and σ_{cf} across fiber. In some three-dimensional simulation works, directions of sheet of fiber and cross-sheet of fiber are treated differently [22, 30].

Hence the $\mathbf{D}$ of one point with α and β defining a rotation around the z- and y-axis of the global coordinate system according to the fiber orientation can be defined:

$$\mathbf{D} = \mathbf{T}\mathbf{D}_{\text{local}}\mathbf{T}^{-1}, \qquad \mathbf{T} = \mathbf{R}_{xz}\mathbf{R}_{xy}, \tag{5}$$

$$\mathbf{R}_{xy} = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \mathbf{R}_{xz} = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}. \tag{6}$$

3. Element-Free Galerkin Method

The reaction-diffusion system is a dynamic system controlled by the diffusion term and reaction term. However, there are too many cellular models, that is, reaction terms, which are beyond the scope of this paper. Therefore, we choose the monodomain model with polynomial cellular model to verify the numerical performance of EFGM in this paper: meshfree particle representation of geometry and fiber structure by unstructured nodes is established first, and then meshfree shape function is constructed from those unstructured nodes; after obtaining Galerkin weak form of the monodomain model using meshfree shape function over meshfree particle representation, a regular background mesh, served as an integration scheme, is applied to solve Galerkin weak form of the monodomain model numerically.

3.1. Meshfree Particle Representation. In FEMs the problem domain is always discretized by finite elements, such as triangular meshes in 2D and tetrahedral meshes in 3D. These elements are constructed through certain constraints, such as connectivity and size. Then field variables, such as potential or fiber direction, are interpolated by elemental

shape function. However in EFGM the problem domain is represented by a cloud of unstructured nodes without any predefined connectivity, named by meshfree particle representation, and field variables are approximated by meshfree shape function. In Figure 1, a meshfree representation of Auckland heart model and its fiber orientations is shown from whole view, one slice to one section of muscle wall. In meshfree particle representation all nodal positions can be arbitrary, so irregular boundaries or interfaces of inhomogeneity can be simply represented by nodes and nodal positions can follow the changing of boundaries or interfaces easily [23–25]. Several works also developed different adaptive meshfree representations using level set method [31], triangular meshing in 2D [25], or tetrahedral meshing in 3D [25]. Moreover, refinement of EFGM could be accomplished by adding or removing nodes into existing representation according to the requirement of accuracy in interested area, such as more nodes should be added into interested area if the error is particularly large or higher accuracy is required in this local area [24, 25].

3.2. Meshfree Shape Function. After meshfree particle representation is established, approximation of field variables can be computed using meshfree shape function and finite nodal values. Construction of shape function is the kernel of EFGM, which includes three steps: (1) determine the size of influence domain of each Gaussian point and search nodes (In this paper, the node only refers to the node of the meshfree representation, Gaussian point always refers to the quadrature point of Gaussian quadrature scheme.) which fall inside the influence domain of Gaussian point from meshfree particle representation, for example, $\mathbf{x}_I$ (In this paper, $\mathbf{x}_i$ refers to index of coordinates, and $\mathbf{x}_I$ refers to the index of nodes) and $I = 1, \ldots, n$; (2) choose proper weighting parameters and calculate weight function; (3) compute entries of meshfree shape function and its derivatives in the position of each Gaussian point using moving least square (MLS) approximation.

3.2.1. Influence Domain. The influence domain is used to determine an influence area/supporting area of one point, usually Gaussian point, inside the meshfree particle representation. The shape of influence domain can be any arbitrary
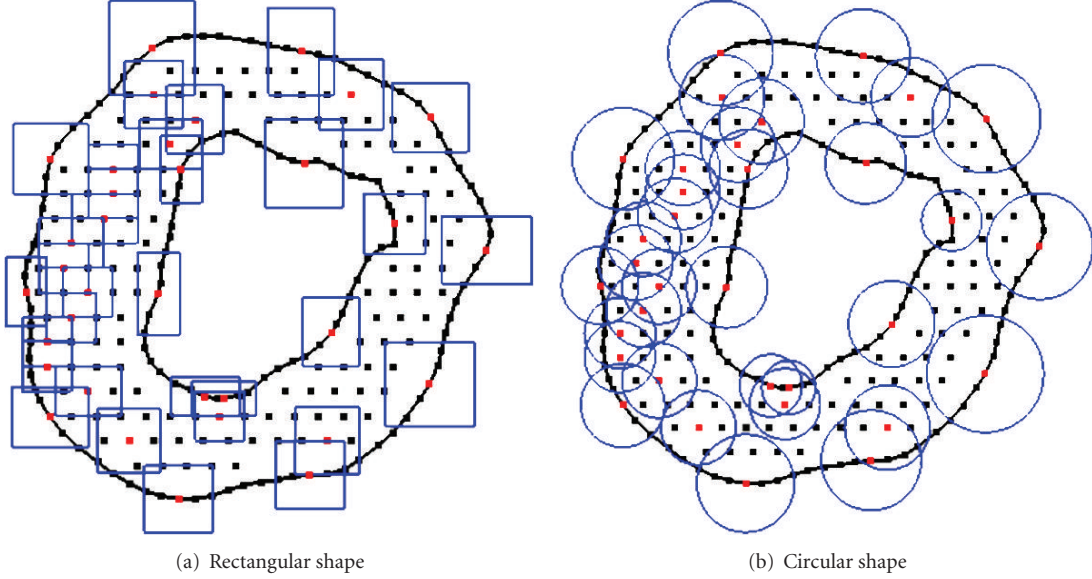
(a) Rectangular shape

(b) Circular shape

Figure 2: Examples of influence domains in 2D. (a) rectangular shape and (b) circular shape.

closed shape in space, while circle or rectangle in two dimensions and sphere or cube in three dimensions are commonly used [24, 25]. Examples of circle and rectangle influence domains are shown in Figure 2. The size of influence domain should reflect the density of nodes (i.e., the size of influence domain in coarse area should be large and the size of influence domain in dense area should be small), and besides, influence domain of one point has to be overlapped with influence domains of neighbouring points to guarantee a smooth approximation of field variables and their derivatives (C^1 continuity). The size of influence domain of node $\mathbf{x}_I$, d_{mI}, is calculated as

$$d_{mI} = d_{\max} c_I, \tag{7}$$

where $d_{\max}$ is a scaling parameter which might vary between different applications and could be determined by numerical experiments [23, 25]. The distance c_I is determined by searching enough neighbouring nodes for the matrix $\mathbf{A}$ in (22), which is discussed in the following subsection, to be invertible, which is also a good strategy to reflect the density of nodes. But influence domain of a point near to any discontinuity should be cut by the discontinuity, including boundary, if this influence domain crosses the discontinuity during the construction of meshfree shape function [24, 25, 32], because the nodes in one side of discontinuity could not affect the nodes or area in the other side of discontinuity. Though cardiac tissue is discontinuous and fiber orientations will not change smoothly at a certain scale any more [3], the heart still could be modelled as a continuum for the propagation of electrical activity, which will not damage our purpose to demonstrate the numerical performance of EFGM.

3.2.2. Weight Function. The weight function, a function of distance $\|\mathbf{x} - \mathbf{x}_I\|$, which obtains a compact support from

the influence domain, needs to be positive to guarantee all meshfree shape functions unique, smooth and continuous throughout the entire problem domain to fulfill the compatibility requirement so that the nodes further from $\mathbf{x}$ will have smaller weights [23–25]. Cubic weight function and quartic weight function are popularly used and they can be replaced by each other in EFGM without rearrangement of nodal positions. Cubic weight function is

$$w\left(\frac{\|\mathbf{x} - \mathbf{x}_I\|}{d_{mI}}\right)$$

$$\equiv w(r_I) = \begin{cases} \dfrac{2}{3} - 4r_I^2 + 4r_I^3 & \text{for } r_I \leq \dfrac{1}{2}, \\[2mm] \dfrac{4}{3} - 4r_I + 4r_I^2 - \dfrac{4}{3}r_I^3 & \text{for } \dfrac{1}{2} < r_I \leq 1, \\[2mm] 0 & \text{for } r_I > 1, \end{cases} \tag{8}$$

where $r_I = \|\mathbf{x} - \mathbf{x}_I\|/d_{mI}$. And the spatial derivative of cubic weight function in location $\mathbf{x}$ is:

$$\frac{dw(r_I)}{dx_i} = \frac{dw(r_I)}{dr_I}\frac{dr_I}{dx_i} \begin{cases} (-8r_I + 12r_I^2)\dfrac{dr_I}{dx_i} & \text{for } r_I \leq \dfrac{1}{2}, \\[2mm] (-4 + 8r_I - 4r_I^2)\dfrac{dr_I}{dx_i} & \text{for } \dfrac{1}{2} < r_I \leq 1, \\[2mm] 0 & \text{for } r_I > 1. \end{cases} \tag{9}$$

Quartic weight function is

$$w\left(\frac{\|\mathbf{x} - \mathbf{x}_I\|}{d_{mI}}\right) \equiv w(r_I) = \begin{cases} 1 - 6r_I^2 + 8r_I^3 - 3r_I^4 & \text{for } r_I \leq 1, \\[2mm] 0 & \text{for } r_I > 1, \end{cases} \tag{10}$$

where $r_I = \|\mathbf{x} - \mathbf{x}_I\|/d_{mI}$. And the spatial derivative of quartic weight function in location $\mathbf{x}$ is

$$
\frac{dw(r_I)}{dx_i}
$$

$$
= \frac{dw(r_I)}{dr_I}\frac{dr_I}{dx_i}\begin{cases} \left(-12r_I + 24r_I^2 - 12r_I^3\right)\dfrac{dr_I}{dx_i} & \text{for } r_I \le 1, \\ 0 & \text{for } r_I > 1. \end{cases}
\tag{11}
$$

3.2.3. MLS Approximation. Let $u^h(\mathbf{x})$ be the approximation of state variable $u_{\mathbf{x}}$ at point $\mathbf{x}$. In the MLS approximation:

$$
u^h(\mathbf{x}) = \sum_{i=1}^{m} p_i(x)e_i(x) \equiv \mathbf{P}^T\mathbf{e},
\tag{12}
$$

where $p_i(x)$ are the polynomial basis functions, m the number of terms in the basis functions, and $e_i(x)$ the unknown coefficients which will be determined later. The basis functions usually consist of monomials of the lowest orders to ensure minimum completeness, and common ones are linear basis:

$$
p^T = \{1, x\} \quad \text{in 1D}, \qquad p^T = \{1, x, y\} \quad \text{in 2D},
$$
$$
p^T = \{1, x, y, z\} \quad \text{in 3D},
\tag{13}
$$

and the quadratic basis:

$$
p^T = \{1, x, x^2\} \quad \text{in 1D},
$$
$$
p^T = \{1, x, y, x^2, xy, y^2\} \quad \text{in 2D},
\tag{14}
$$
$$
p^T = \{1, x, y, z, x^2, y^2, z^2, xy, xz, yz\} \quad \text{in 3D}.
$$

In EFGM, $\mathbf{P}^T$ in (12) can be replaced by any other polynomial basis $\mathbf{P}^T$ without the rearrangement of nodal positions [24, 25]. The consistency of the MLS approximation depends on the complete order of polynomial basis $\mathbf{P}^T$. If the complete order of polynomial basis, $\mathbf{P}^T$, is m, the meshfree shape function will possess C^m consistency [24, 25].

Given a set of n nodal values $u(x_1), u(x_2), \ldots, u(x_n)$ of the field variable u at a set of nodes $\{\mathbf{x}_I\} = x_1, x_2, \ldots, x_n$. The coefficients $a_i(x)$ are obtained by minimizing the difference between the local approximation $u^h(\mathbf{x})$ and the actual nodal parameter $u(\mathbf{x}_I)$ in location $\mathbf{x}$:

$$
J = \sum_{I=1}^{n} w(r_I)\left[u^h(\mathbf{x}) - u(\mathbf{x}_I)\right]^2
$$
$$
= \sum_{I=1}^{n} w(r_I)\left[\sum_{i=1}^{m} p_i(\mathbf{x}_I)e_i(\mathbf{x}) - u(\mathbf{x}_I)\right]^2,
\tag{15}
$$

where $w(r_I)$ is the weighting function with compact support within the influence domain, which is defined in Section 3.2.2 Equation (15) can be rewritten into matrix form:

$$
J = (\mathbf{Pe} - \mathbf{u})^T\mathbf{W}(\mathbf{Pe} - \mathbf{u}),
\tag{16}
$$

where

$$
\mathbf{u} = [u_1, u_2, \ldots, u_n]^T,
$$

$$
\mathbf{P} = \begin{bmatrix} p_1(\mathbf{x}_1) & p_2(\mathbf{x}_1) & \cdots & p_m(\mathbf{x}_1) \\ p_1(\mathbf{x}_2) & p_2(\mathbf{x}_2) & \cdots & p_m(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_1(\mathbf{x}_n) & p_2(\mathbf{x}_n) & \cdots & p_m(\mathbf{x}_n) \end{bmatrix},
$$

$$
\mathbf{W} = \begin{bmatrix} w(r_1) & 0 & \cdots & 0 \\ 0 & w(r_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w(r_n) \end{bmatrix}.
\tag{17}
$$

At point $\mathbf{x}$, coefficients $\mathbf{E}(\mathbf{x})$ are chosen by minimizing the weighted residual, which are realized through $\partial J/\partial e = 0$:

$$
\frac{\partial J}{\partial e} = \mathbf{AE} - \mathbf{Bu} = 0
\tag{18}
$$

therefore,

$$
\mathbf{e} = \mathbf{A}^{-1}\mathbf{Bu},
\tag{19}
$$

where

$$
\mathbf{A} = \mathbf{P}^T(\mathbf{x}_I)\mathbf{W}(r_I)\mathbf{P}(\mathbf{x}_I), \qquad \mathbf{B} = \mathbf{P}^T(\mathbf{x}_I)\mathbf{W}(r_I).
\tag{20}
$$

Substituting (19) into (12), the approximation $u^h(\mathbf{x})$ becomes

$$
u^h(\mathbf{x}) = \sum_{I=1}^{n} \phi_I(x)u_I = \phi(\mathbf{x})\mathbf{u},
\tag{21}
$$

where the meshfree shape function $\phi(\mathbf{x})$ is defined by

$$
\phi(\mathbf{x}) = \mathbf{P}(\mathbf{x})^T\mathbf{A}^{-1}\mathbf{B},
\tag{22}
$$

with m the order of the polynomial in $\mathbf{P}(\mathbf{x})$. Note that m, the number of terms of the polynomial basis, is usually set to be much smaller than n, the number of nodes used for constructing the meshfree shape function. The spatial derivative of meshfree shape function in $\mathbf{x}$ is obtained by:

$$
\phi(\mathbf{x})_{,x_i} = \left(\mathbf{P}^T(\mathbf{x})\mathbf{A}^{-1}\mathbf{B}\right)_{,x_i} = \mathbf{P}^T_{,x_i}(\mathbf{x})\mathbf{A}^T\mathbf{B} + \mathbf{P}^T(\mathbf{x})\mathbf{A}^T_{,x_i}\mathbf{B}
$$
$$
+ \mathbf{P}^T(\mathbf{x})\mathbf{A}^T\mathbf{B}_{,x_i},
\tag{23}
$$

where

$$
\mathbf{B}_{,x_i} = \frac{d\mathbf{W}(r_I)}{dx_i}\mathbf{P}(\mathbf{x}_I),
\tag{24}
$$

and $\mathbf{A}^{-1}_{,x_i}$ is computed by

$$
\mathbf{A}^{-1}_{,x_i} = -\mathbf{A}^{-1}\mathbf{A}_{,x_i}\mathbf{A}^{-1}, \qquad \mathbf{A}_{,x} = \mathbf{P}^T(\mathbf{x}_I)\frac{d\mathbf{W}(r_I)}{dx_i}\mathbf{P}(\mathbf{x}_I),
\tag{25}
$$

where $d\mathbf{W}(r_I)/dx_i$ is defined in Section 3.2.2 Then the approximation of first derivative of field variable u can be obtained in $\mathbf{x}$:

$$
u^h(\mathbf{x})_{,x_i} = \sum_{i=1}^{N} \phi(\mathbf{x})_{,x_i}\mathbf{u}_I,
\tag{26}
$$

and is continuous in the whole problem domain.

3.3. Construction of Galerkin Weak Form. In Galerkin weak form differential equations are transformed into integral form by using the weighted residual strategies so that they are satisfied over a domain in an integral sense rather than every point. Consider the integral form of (3), which also can be easily applied to bidomain (1), we have

$$\int_\Omega \left[\nabla \cdot (\mathbf{D}\nabla v_m) - A_m \left(C_m \frac{\partial v_m}{\partial t} + I_{\text{ion}} \right) \right] v d\Omega = 0, \quad (27)$$

where v is the trial function. The exact solution of (3) should always satisfy integral in (27). Substituting $f = -A_m I_{\text{ion}}$ and evaluating integral in (27) using Green's formulae

$$\int_\Omega A_m C_m \frac{\partial v_m}{\partial t} v d\Omega + \int_\Omega \nabla v_m \cdot \mathbf{D}\nabla v d\Omega - \mathbf{D}\oint_S v \frac{\partial v_m}{\partial \mathbf{n}} dS$$
$$= \int_\Omega f v d\Omega, \quad (28)$$

where S is the boundary of Ω and $\mathbf{n}$ is a vector normal to boundary. Equation (28) can automatically fulfill zero natural boundary condition, $\partial v_m / \partial \mathbf{n} = 0$, by eliminating $\mathbf{D}\oint_S v(\partial v_m/\partial \mathbf{n}) dS$ at boundary S, but an accurate numerical integral scheme should be applied to the rest parts of (28) so that zero natural boundary condition can be enforced correctly in numerical sense. In Galerkin weak form procedure, trial function could be replaced by the shape function, Φ^T, of EFGM here:

$$\int_\Omega A_m C_m \frac{\partial v_m}{\partial t} \Phi^T d\Omega + \int_\Omega \nabla v_m \mathbf{D}\nabla \Phi^T d\Omega = \int_\Omega f \Phi^T d\Omega. \quad (29)$$

To solve (29) we need to discrete them. Let $\mathbf{v}_I$ be the vector of nodal values of transmembrane potentials v_m, and let $\mathbf{f}_I$ be the vector of nodal values of $f = -A_m I_{\text{ion}}$ at node set $\mathbf{x}_I$. Then $v_m \approx \Phi \mathbf{v}_I$ and $f \approx \Phi \mathbf{f}_I$ and a continuous form of (29) can be discretized:

$$A_m C_m \frac{\partial \mathbf{v}_I}{\partial t} \int_\Omega \Phi\Phi^T d\Omega + \mathbf{D}\mathbf{v}_I \int_\Omega \nabla\Phi\nabla\Phi^T d\Omega$$
$$= \mathbf{f}_I \int_\Omega \Phi\Phi^T d\Omega. \quad (30)$$

Rewrite equations previously mentioned with matrices:

$$A_m C_m \frac{\partial \mathbf{v}_I}{\partial t} + \mathbf{M}^{-1}\mathbf{K}\mathbf{v}_I = \mathbf{f}_I,$$

$$\mathbf{M}_{i,j} = \int_\Omega \phi_i^T \phi_j d\Omega, \qquad \mathbf{K}_{i,j} = \int_\Omega \mathbf{B}_i^T \mathbf{D}\mathbf{B}_j d\Omega \quad (31)$$

$$\mathbf{B}_i = \begin{bmatrix} \phi_{i,x} \\ \phi_{i,y} \\ \phi_{i,z} \end{bmatrix},$$

with $\mathbf{D}$ the diffusion tensor transformed from fiber coordinate (5), $\phi_{i,x}$, $\phi_{i,y}$, and $\phi_{i,z}$ the derivatives of the shape function with respect to x, y, and z, ϕ_i the matrix of shape functions, and $\mathbf{B}_i$ the differential matrix at the ith node.

3.4. Integration Schemes. The shape function of EFGM does not fulfill the property of strict interpolation, that is, $\phi_i(x_j) \neq \delta_{ij}$, which implies that essential boundary condition cannot be imposed directly, so penalty method and Lagrange multiplier are proposed to enforce essential boundary condition in EFGM [32, 33]. However, zero natural boundary condition can be enforced in Galerkin weak form (Equation (28)) by placing sufficient nodes along the boundaries and then applying a correct integration scheme.

In EFGM, a regular background mesh, which consists of nonoverlapping regular cells and covers the whole problem domain, is a very popular choice to perform the integration of computing $\mathbf{M}$ and $\mathbf{K}$ matrix in (29) because of its simplicity. The regular cells of background mesh are commonly squares in two dimension, and cubes in three dimensions. The proper density of background mesh needs to be designed to approximate solutions of desired accuracy. In each cell, Gauss quadrature scheme is used. The number of quadrature points, integration points, seems to depend on the number of nodes in the cell. An empirical guideline of quadrature points suggests [25]

$$n_q = \sqrt{n_n} + 2 \quad \text{in 2D}, \qquad n_q = \sqrt[3]{n_n} + 1 \quad \text{in 3D}, \quad (32)$$

where n_n is the number of nodes in the cell and n_q is the number of quadrature points in one cell. Our experience with Gauss quadrature in EFGM suggests that a lower order quadrature (smaller n_q) with finer background mesh may be preferable to a higher order quadrature (larger n_q) with coarser background mesh. The background cells are usually independent of the arrangement of sample nodes and large enough to hold the whole problem domain, but in regular domain with regular nodes, it can be coincided with problem domain and depend on nodal positions. In the background mesh, there may exist the cell that does not entirely belong to the problem domain; that is, only a portion of this cell would contribute to (29). This contribution could be realized by counting the quadrature weights of those quadrature points in this cell, which are inside problem domain, and ignoring other quadrature points of this cell, which are outside problem domain (Figure 3). Therefore, a scheme that automatically detects the quadrature points of each cell which lie inside of the problem domain is employed. Hence the integral of (29) over irregular problem domain is solved numerically in those quadrature points inside problem domain. In [26], an irregular background mesh is proposed to achieve higher accuracy, but the improvement is not obvious and it may increase the time of assembling system matrices. Finally we can give out the flow of EFGM:

(1) set up sample nodes,

(2) set up background mesh and quadrature points in all cells,

(3) loop over all the quadrature points,

    (a) if this quadrature point is outside the problem domain, go to 3*e*,

    (b) determine nodes whose influence domains cover this quadrature point by searching enough neighboring nodes,
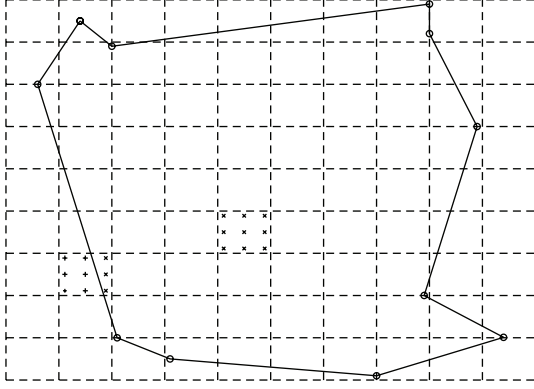
FIGURE 3: Dashed square cells consist of a background mesh, which covers the whole problem domain—the area confined by solid lines. In each cell, Gauss quadrature will be applied. Here only two cells are marked to illustrate this process. Those Gauss points, indicated by $\times$ marker, which are inside problem domain will be counted during integration, but the other Gauss points, indicated by $+$ marker, which are out of problem domain, will not be counted.

    (c) calculate quadrature weight, weight function and shape function in this quadrature point,

    (d) assemble $\mathbf{M}$ and $\mathbf{K}$ matrices in (29),

    (e) end if,

(4) End loop.

4. Experiment

Numerical experiments are implemented by Matlab, and the simulation of electrical propagation in Auckland heart model is implemented by C++ and Matlab external C++ math library in a Dell precision T3400 workstation with a quad cores 2.4 GHz CPU and 4G DDR2 memory. Let Π_i^{exact} be the analytical solution and let $\Pi_i^{\text{numerical}}$ be the numerical solution in node i, respectively. To a set of nodes, from 1 to N, root mean square (RMS) error is.

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\Pi_i^{\text{exact}} - \Pi_i^{\text{numerical}}\right)^2}. \tag{33}$$

The behaviour of reaction-diffusion equation is controlled by the diffusion term and reaction term simultaneously. The reaction term could have huge varieties in electrical propagation applications [3, 4], and it is impossible to evaluate EFGM's performance over all the forms of reaction term in this paper; however it would be valuable to compare EFGM to FEM in approximating diffusion process. So a two-dimensional heat conduction problem without reaction term is first tested by FEM and EFGM:

$$\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} = \frac{1}{\sigma}\frac{\partial C}{\partial t}, \tag{34}$$

where C temperature, σ diffusion tensor, and t time. The analytic solution of (34) in infinite media is [34]

$$C_t = \frac{C_0}{4\pi\sigma t}\exp\left(-\frac{x^2}{4\sigma t}\right), \tag{35}$$

where C_0 is initial source in $x = 0$ at $t = 0$. The numerical simulations are initialized by the analytic solution at $t = 1$, which can be calculated from (35) with $C_0 = 1$, and then numerical solutions are obtained in $t = 2$ in 20×20 area in order to approximate the effect of infinite media through a small time duration and a large enough area. Though EFGM does not always require regular nodes, it is convenient to determine the convergence rate by reducing spacing between regular nodes. The convergence behaviour of EFGM using different $d_{\max}$ and weight functions is also evaluated. Euler forward method is applied from $t = 1$ to $t = 2$ for time integration. To find a stable RMS error in each spatial discretization, more than 10^5 time steps of Euler method are used in our implementation. Because of regular problem domain, 20×20 area, the nodes of EFGM are chosen from grid points from 20×20 grids to 80×80 grids; that is, the spacing h is from 1 to 0.25. These grids are also used as background mesh for EFGM, respectively; for example, for 20×20 grids, there are 21×21 nodes for EFGM and 20×20 cells in the background mesh, and for 80×80 grids, there are 81×81 nodes for EFGM and 80×80 cells in the background mesh. In all the cells of background mesh, 4×4 Gaussian quadrature scheme is applied. The same background meshes are used as meshes of linear FEM, and the convergence curve of linear FEM is obtained using the same Gaussian quadrature scheme for fair comparison. The convergence curves of EFGM are displayed in Figure 4 along with the convergence curve of linear FEM. When $d_{\max} = 1.1$, both curves of cubic weight function and quartic weight function in EFGM show almost identical convergence behaviour as linear FEM. Without changing linear basis and nodal positions in EFGM, the convergence rates of EFGM become better in both weight functions when $d_{\max}$ increases from 1.1 to 3.0, and these curves are far below the curve of linear FEM. However, the convergence behaviours of EFGM do not become better when $d_{\max}$ has even bigger value. When $d_{\max} = 4.0$, the slopes of convergence curves (Figure 4) are larger, but RMS errors increase sharply in coarse nodes. A great value of $d_{\max}$, that is, oversized influence domain, will produce oversmoothing effect as one huge element or too coarse mesh in FEM, which is the reason that RMS errors of EFGM increase largely in coarse nodes with too bigger value of $d_{\max}$. Hence the suggested range of $d_{\max}$ is between 1 and 3 [24, 25]. As shown by all the convergence curves in Figure 4, EFGM shows much better behaviour than linear FEM. Higher-order Gaussian quadrature scheme of each cell of background mesh will help EFGM gain better accuracy, but lower-order Gaussian quadrature scheme in the cells of finer background meshes also works quite well. Another experiment, with 2×2 Gaussian quadrature scheme in each cell of background mesh and total number of cells being 4 times as large as before, is compared to previous

(a) Convergence rate of EFGM using cubic weight function

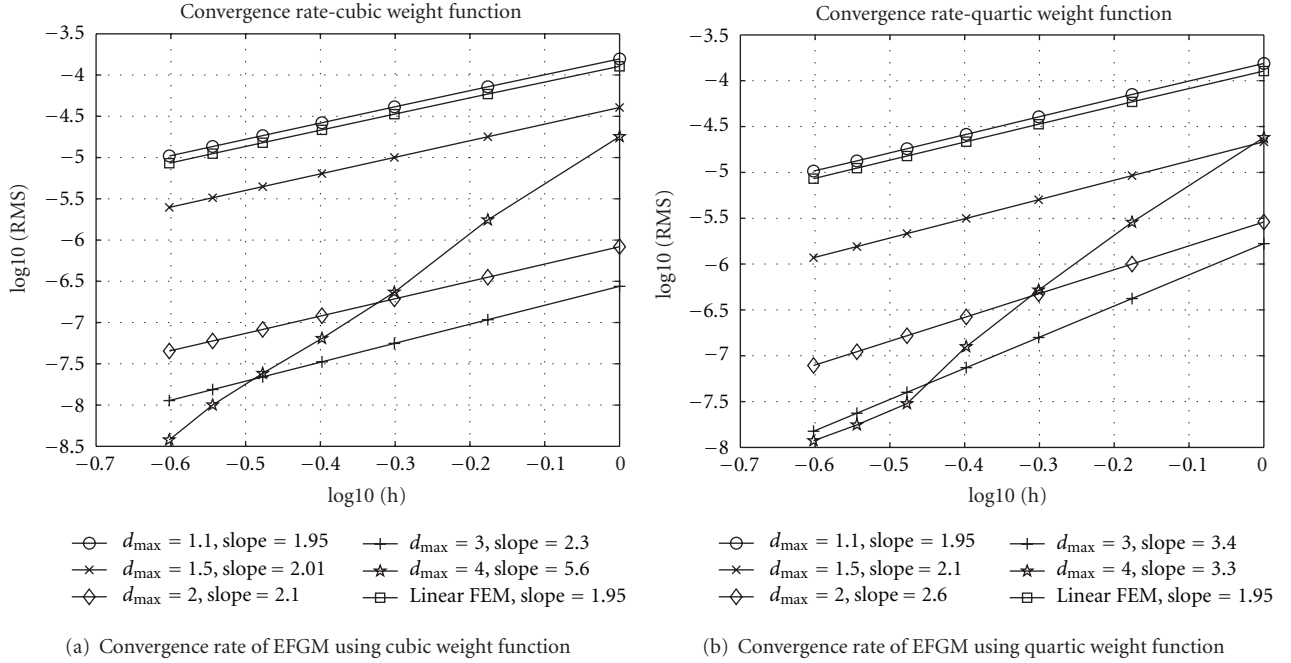(b) Convergence rate of EFGM using quartic weight function

FIGURE 4: All the experiments, including FEM, use linear polynomial basis, and their convergence rates are indicated by slope. All the nodes are regularly placed and so h is the spacing between nodes. 4×4 Gaussian quadrature scheme in each cell of background mesh is used for numerical integration.

results in Figure 4. The convergence behaviours of lower-order Gaussian quadrature scheme in finer mesh (indicated by 2×2 in one cell in Figure 5) are better than higher-order Gaussian quadrature scheme in coarse mesh (indicated by 4×4 in one cell in Figure 5).

An analytic result of reaction-diffusion system of cardiac electrical activity seldom exists that allows the performance of numerical methods to be verified. However, an analytic solution of conduction velocity in a one-dimensional fiber is available when a cubic polynomial ionic current model is used as a reaction term of the monodomain model [35]. The conduction velocity is determined by each location's activation time, which is defined by the time at which the maximum upstroke velocity occurs [3]. The cubic polynomial ionic current model is given by

$$I_{\text{ion}} = g\left[v_m\left(1 - \frac{v_m}{v_{\text{th}}}\right)\left(1 - \frac{v_m}{v_p}\right)\right], \qquad (36)$$

and the analytic conduction velocity γ is given by

$$\gamma = \sqrt{\frac{g\sigma}{A_m C_m^2}\left(\frac{S^2}{S+1}\right)},$$

$$S = \frac{v_p}{2v_{\text{th}}} - 1, \qquad (37)$$

where g is the membrane conductance. v_{th} and v_p represent the threshold potential and the plateau potential, respectively. All the potential variables in cubic polynomial ionic current model are expressed as deviations from the resting potential. The parameters used in cubic current model are

listed in Table 1. Again, the same setting of nodes is used in FDM, linear FEM, and EFGM (cubic weight function and quartic weight function), and time integration is solved by Euler forward method again. After activation times of all nodes are available, RMS error of conduction velocity can be calculated. In Figure 6, the relation between RMS errors of different numerical methods and spatial discretization is displayed. The convergence behaviours of EFGM are still better than conventional methods, FDM and FEM, after a cubic polynomial reaction term is included. In Table 2, that the computational costs to reach a similar level of error for conduction velocities of different σ values are shown. From Table 2, it can be seen that EFGM could achieve similar level of error using considerably less time. The computational costs presented in Table 2 have been split into "assemble" (the time taken to assemble the global system of equations) and "propagation" (the time taken to solve the global system of equations) times.

To explore the further ability of EFGM in simulation of cardiac electrical activity, one published monodomain model, a modified FHN model [7], is solved by EFGM. This FHN model [7] is

$$\frac{\partial v_m}{\partial t} = f(v_m, I_{\text{ion}}) + \nabla \cdot (\mathbf{D}\nabla v_m),$$

$$\frac{\partial I_{\text{ion}}}{\partial t} = b(v_m - dI_{\text{ion}}), \qquad (38)$$

$$f(v_m, I_{\text{ion}}) = c_1 v_m(v_m - a)(1 - v_m) - c_2 v_m I_{\text{ion}},$$

with natural boundary condition $(\mathbf{D}\nabla v_m) \cdot n = 0$. Values of parameters are taken from [7], which are listed in Table 3.

(a) different background meshes using cubic weight function

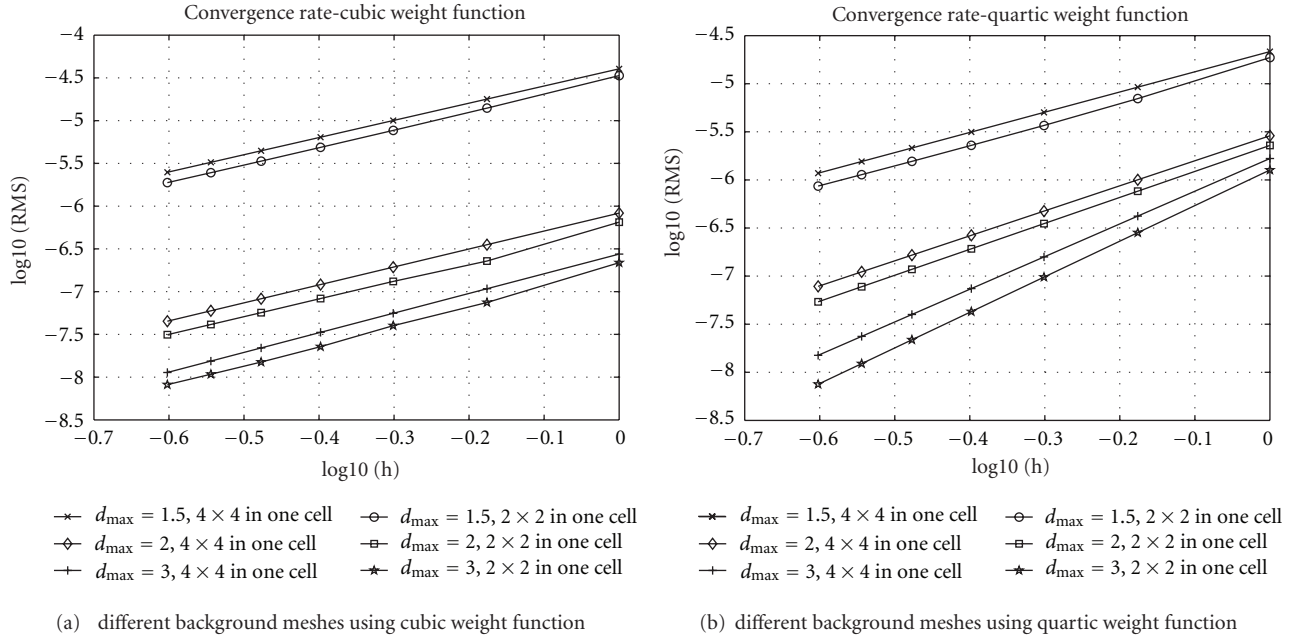(b) different background meshes using quartic weight function

FIGURE 5: Convergence behaviours of EFGM using different background mesh. h is the spacing between nodes. RMS is the error measure. It shows that 2×2 Gaussian quadrature scheme in each cell of fine background mesh works a little better than 4×4 Gaussian quadrature scheme in each cell of coarse background mesh.
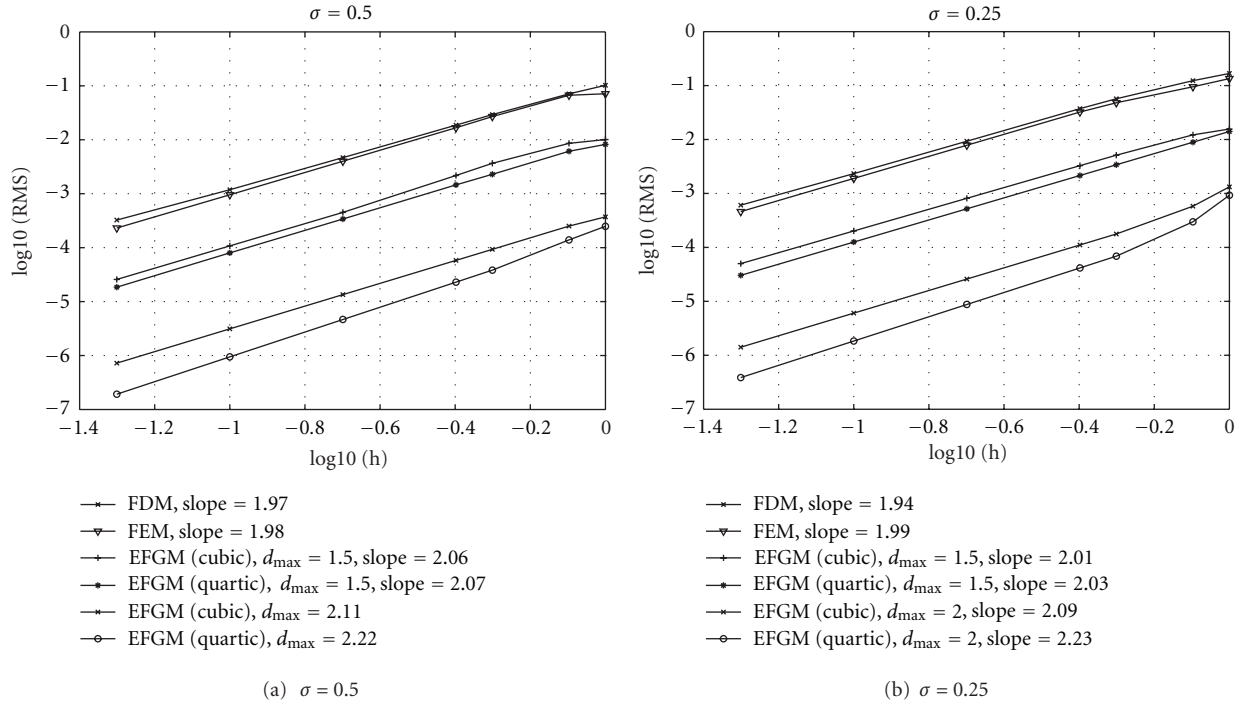

(a) $\sigma = 0.5$

(b) $\sigma = 0.25$

FIGURE 6: Results of conduction velocity using FDM, FEM, and EFGM. h is the spacing between nodes and RMS is the error measure. It shows that EFGM still has better convergence rate after the reaction term is included than FEM and FDM. 2×2 Gaussian quadrature scheme in each cell of one finer background mesh is applied.

TABLE 1: Parameters of cubic current model.

Parameter	v_{rest}	v_{th}	v_p	g	C_m	A_m
Value	$-85.0\,mV$	$-75.0\,mV$	$15.0\,mV$	$0.004\,mS\,mm^{-2}$	$0.01\,\mu F\,mm^{-2}$	$200\,\mu F\,mm^{-1}$

TABLE 2: Comparison of computational costs.

	RMS	h (mm)	Assemble (sec)	Propagation (sec)
$\sigma = 0.5$				
FDM	$3.30e-04$	0.05	n/a	258.31
FEM	$2.31e-04$	0.05	0.20	248.26
EFGM(cubic, $d_{max} = 1.5$)	$4.56e-04$	0.2	0.11	39.50
EFGM(cubic, $d_{max} = 2.0$)	$2.60e-04$	0.8	0.05	41.33
EFGM(quartic, $d_{max} = 1.5$)	$3.40e-04$	0.2	0.07	38.83
EFGM(quartic, $d_{max} = 2.0$)	$1.49e-04$	0.8	0.04	40.60
$\sigma = 0.25$				
FDM	$6.06e-04$	0.05	n/a	580.11
FEM	$4.60e-04$	0.05	0.21	591.42
EFGM(cubic, $d_{max} = 1.5$)	$8.12e-04$	0.2	0.10	48.23
EFGM(cubic, $d_{max} = 2.0$)	$5.80e-04$	0.8	0.04	50.90
EFGM(quartic, $d_{max} = 1.5$)	$5.20e-04$	0.2	0.08	47.81
EFGM(quartic, $d_{max} = 2.0$)	$3.00e-04$	0.8	0.02	50.90

(a)

(b)
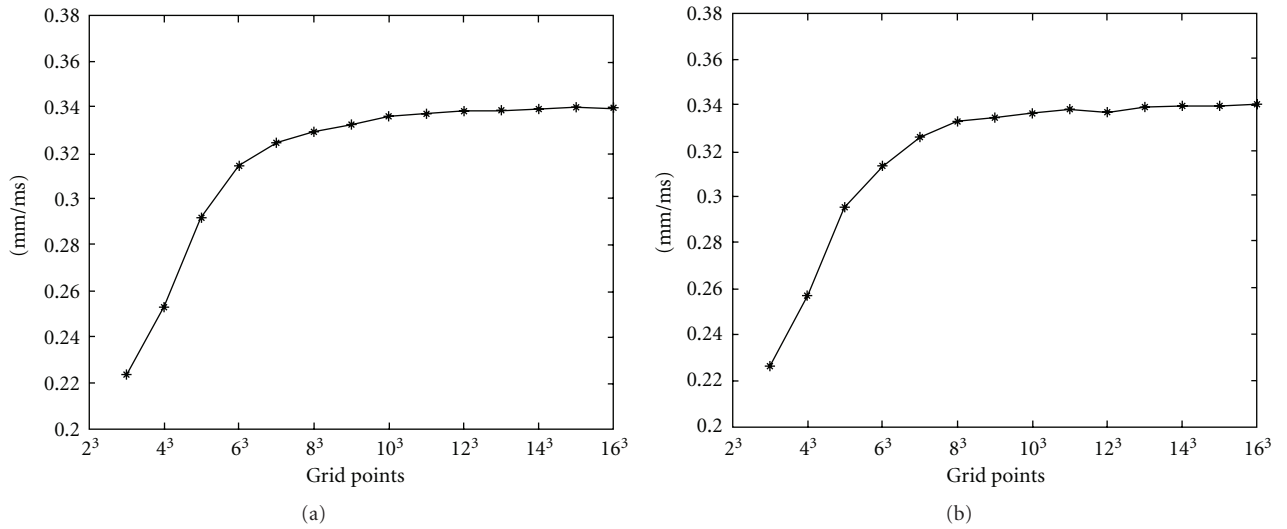
FIGURE 7: The convergence of the velocity of propagation wave with increasing density of regular sample nodes. (a) $2 \times 2 \times 2$ quadrature points in each background cell; (b) $3 \times 3 \times 3$ quadrature points in each background cell.

(a)   (b)   (c)   (d)   (e)   (f)
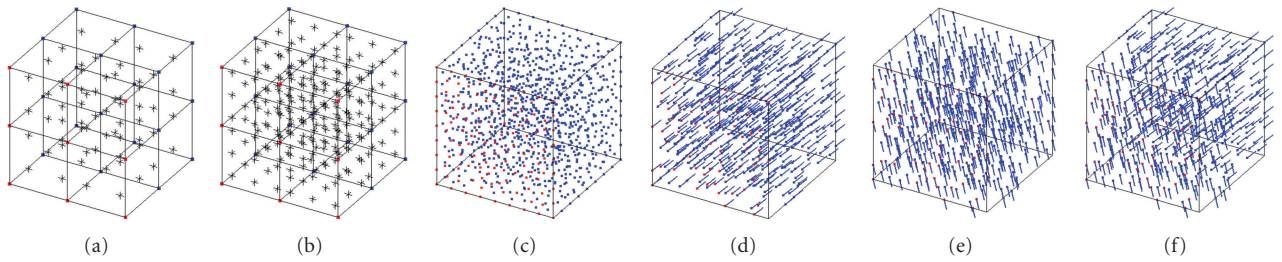
FIGURE 8: (a) $3 \times 3 \times 3$ grid points (solid) and 2^3 quadrature points (stars) in each background cell; (b) $3 \times 3 \times 3$ grid points (solid) and 3^3 quadrature points (stars) in each background cell; (c) meshfree representation of cube with irregular sample nodes (1106 nodes); (d) all the fiber directions are (0.57735, 0.57735, −0.57735); (e) all the fiber directions are (0.57735, −0.57735, 0.57735); (f) half is (0.57735, 0.57735, −0.57735) and half is (0.57735, 0.57735, −0.57735). Red points in the front side are stimulated at the beginning.
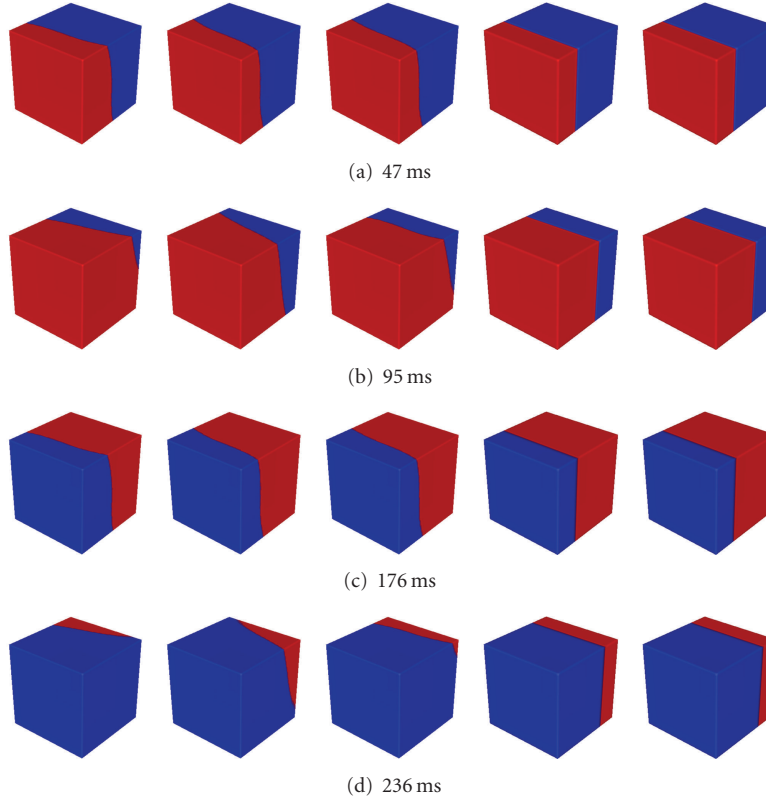
(a) 47 ms

(b) 95 ms

(c) 176 ms

(d) 236 ms

FIGURE 9: Propagation wave at 47 ms, 95 ms, 176 ms, and 236 ms with regular sample nodes ($10 \times 10 \times 10$ nodes). Fiber orientation from column 1 to column 5: $(0.57735, 0.57735, -0.57735)$; $(0.57735, -0.57735, 0.57735)$; half is $(0.57735, 0.57735, -0.57735)$ and half is $(0.57735, 0.57735, -0.57735)$; isotropic; isotropic. Except 3^3 quadrature points in each background cell are applied in column 5, 2^3 quadrature points in each background cell are applied in other columns. In each column total 10^3 background cells are used for integration of Galerkin weak form. Diffuse parameter: column 1, 2, and 3 : $d_f = 4$, $d_{cf} = 1$, column 4 and 5: $d_f = d_{cf} = 1$. Red color represents active state and blue color represents quiescent state.

State variable v_m is the excitation variable which corresponds to the transmembrane voltage, I_{ion} is the recovery current variable, n is the normal of the boundary, $f(v_m, I_{ion})$ is the excitation term, a, b, $c1$, $c2$, and d are parameters that define the shape of action potential. These parameters are constant over time but not necessary in space. The changes of state variables are determined by the excitation term $f(v_m, I_{ion})$ and diffusion term $\nabla \cdot (\mathbf{D} \nabla v_m)$, and $\mathbf{D}$ is defined in (5).

In order to find out a proper density of sample nodes in EFGM for a stable propagation wave of the FHN model in heart, two series of isotropic plane waves of electrical propagation with increasing regular sample nodes in a cube, whose size is 60 mm $\times$ 60 mm $\times$ 60 mm, are solved by setting an initial potential, 0.5, to one side of cube, and then the conduction velocity is calculated by activation time. A fourth-order Runge-Kutta method, which can select time step automatically, is applied for time integration. Two series of the isotropic electrical propagation with regular sample nodes, which change from $3 \times 3 \times 3$ grid nodes to $16 \times 16 \times 16$ grid nodes, and correspondingly the regular background mesh, whose background cells change from $2 \times 2 \times 2$ to $15 \times 15 \times 15$, are simulated, but one uses 2^3 quadrature points in each background cell and the other uses 3^3 quadrature points in each background cell. Convergence

curves of conduction velocity are plotted in Figure 7, and a stable speed of propagation wave is achieved in both curves after sample nodes are equal to or greater than $10 \times 10 \times 10$.

In Figure 9 propagations with different fiber orientations using $10 \times 10 \times 10$ regular sample nodes are displayed in different time instants. A fourth-order Runge-Kutta method, which can select time step automatically, is still used for time integration. The fiber orientations from column 1 to column 3 are illustrated from Figure 8(d) to Figure 8(f). In these first three columns d_f is set to 1, and d_{cf} is set to 4. In column 4 and column 5 isotropic propagations, but different quadrature points, are displayed. In Figure 10 propagations with 1106 irregular sample nodes are displayed in different time instants. In Figure 8(c) the positions of these irregular sample nodes are shown. In Figure 10 fiber orientations in the first three columns are the same as the fiber orientations in the first three columns in Figure 9 accordingly. Two isotropic propagations with different quadrature points are also tested in irregular sample nodes, which are displayed in column 4 and column 5 of Figure 10. From Figures 9 and 10, almost identical propagations can be seen between corresponding two columns, which demonstrate that the performance of EFGM in solving FHN model will not be damaged by using irregular nodes.

(a) 47 ms
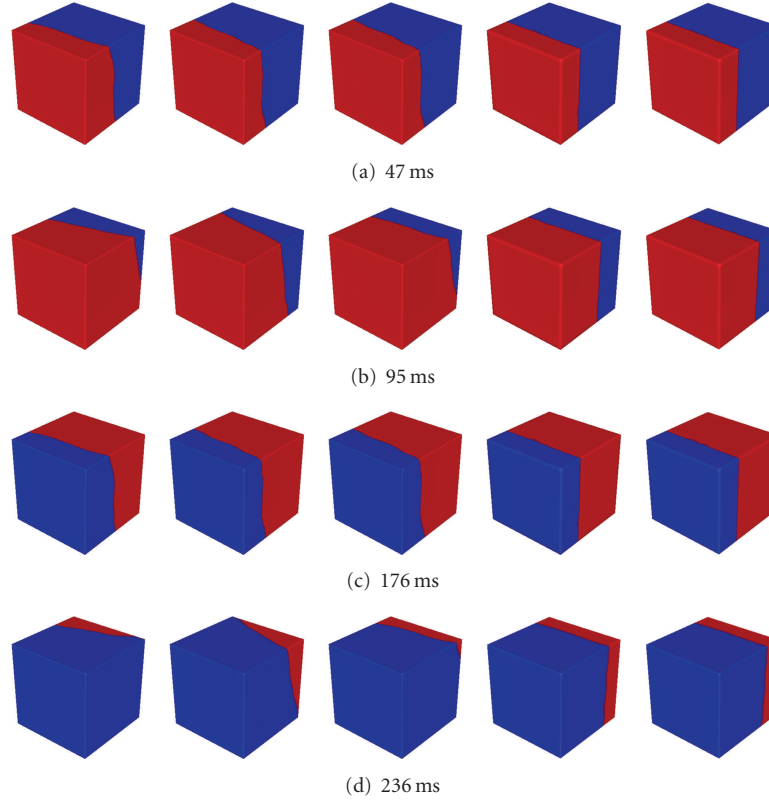

(b) 95 ms

(c) 176 ms

(d) 236 ms

FIGURE 10: Propagation wave at 47 ms, 95 ms, 176 ms, and 236 ms with irregular sample nodes (1106 nodes). Fiber orientation from column 1 to column 5: $(0.57735, 0.57735, -0.57735)$; $(0.57735, -0.57735, 0.57735)$; half is $(0.57735, 0.57735, -0.57735)$ and half is $(0.57735, 0.57735, -0.57735)$; isotropic; isotropic. Except 3^3 quadrature points in each background cell are applied in column 5, 2^3 quadrature points in each background cell are applied in other columns. In each column total 10^3 background cells are used for integration of Galerkin weak form. Diffuse parameter: column 1, 2 and 3: $d_f = 4$, $d_{cf} = 1$, column 4 and 5: $d_f = d_{cf} = 1$. Red color represents active state and blue color represents quiescent state.

TABLE 3: Parameters of FHN model.

Parameter	a	b	c_1	c_2	d	σ_f	σ_{cf}
Value	0.13	0.013	0.26	0.1	1.0	4.0	1.0

Finally we select 3164 sample nodes from Auckland heart model and use 3^3 quadrature points in each background cell as suggested by the experiment mentioned previously (Figure 7). In Auckland heart model, σ_f is set to 4 and σ_{cf} is set to 1 as we did in the cube. Purkinje network is manually chosen on endocardium because of unavailable Purkinje network locations. From Figure 11 ((with permission): http://www.bem.fi/book/) which is generated from Durrer's [36] measurements from isolated human hearts, it can be seen that propagation of electrical activity starts from several locations on the endocardium, that is, Purkinje network extremities. Hence, a small set of nodes (around 6 nodes) around corresponding locations on the endocardium of Auckland heart model are initialized with a starting potential, 0.5 in our simulation, and the result solved by EFGM is displayed in Figure 11 in different time instants. It is reported that isolation of the heart would lead to an increase in conduction velocity [36]. Actually durations of QRS waveforms in healthy individuals vary from 80 ms to 100 ms since durations of QRS waveforms are determined by depolarization processes in the healthy hearts. That is the reason that the duration of propagation in Durrer's data is shorter than the duration of propagation in our simulation. The activation process in our simulation is qualitatively close to the published measurements as we can see in Figure 11. Once cycle of simulate of electrical propagation in Auckland heart model includes generating sample nodes, assembling of matrices and time integration. The time integration is done by the Runge-Kutta method using automatic time step. It takes 21 minutes to simulate the whole cycle of electrical propagation in Auckland hear model.

In the end, we also simulate the propagation in the human left ventricle extracted from digitized virtual Chinese dataset [37]. In this simulation, we only demonstrate the ability of EFMG in simulating in different cardiac geometry because of the lack of ground truth. The results are displayed in Figure 12.
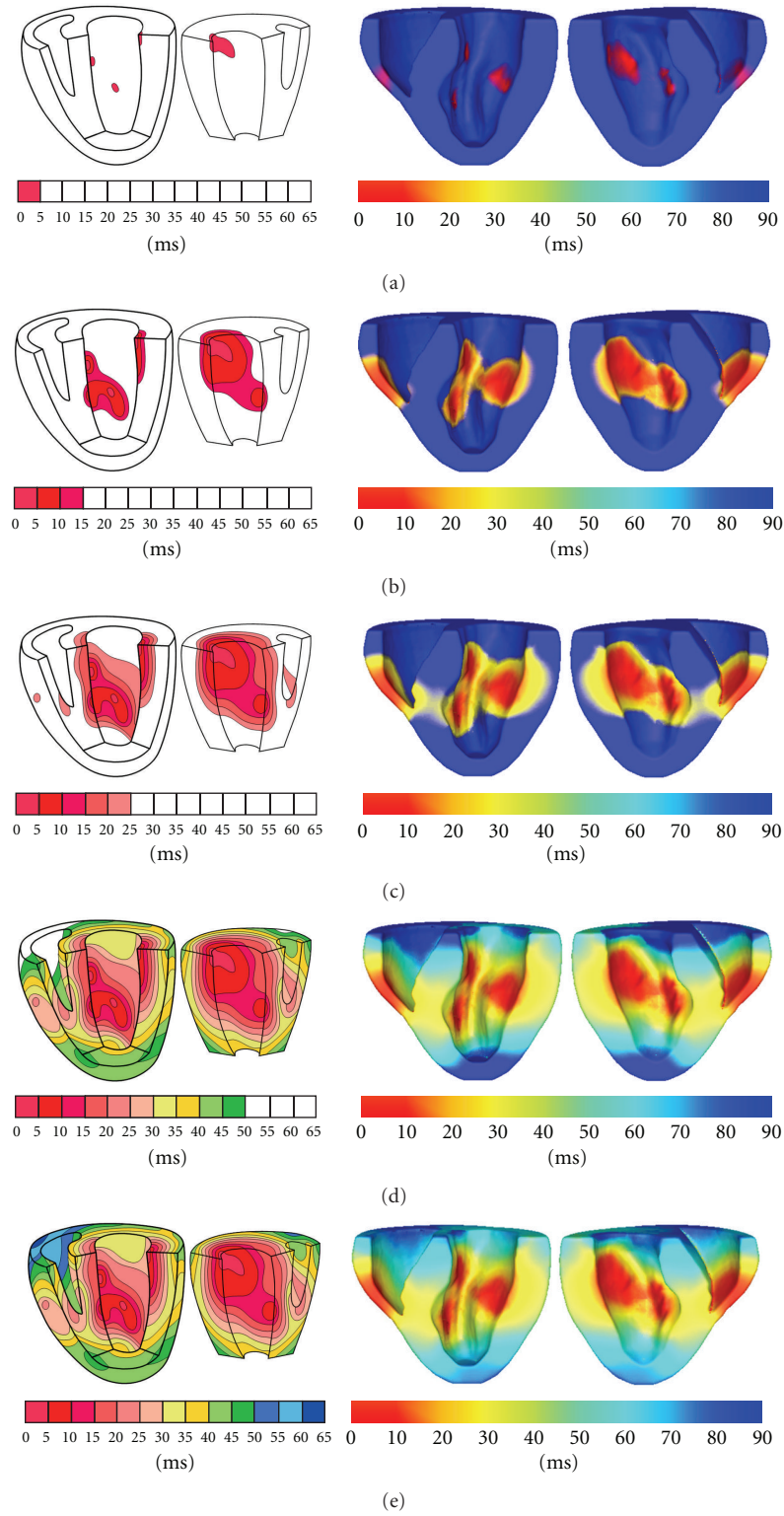
FIGURE 11: Comparison between Durrer's measurements and our simulation results. (a) 5 ms (left) and 10 ms (right), (b) 15 ms (left) and 30 ms (right), (c) 25 ms (left) and 40 ms (right), (d) 50 ms (left) and 70 ms (right), (e) 65 ms (left) and 90 ms (right).
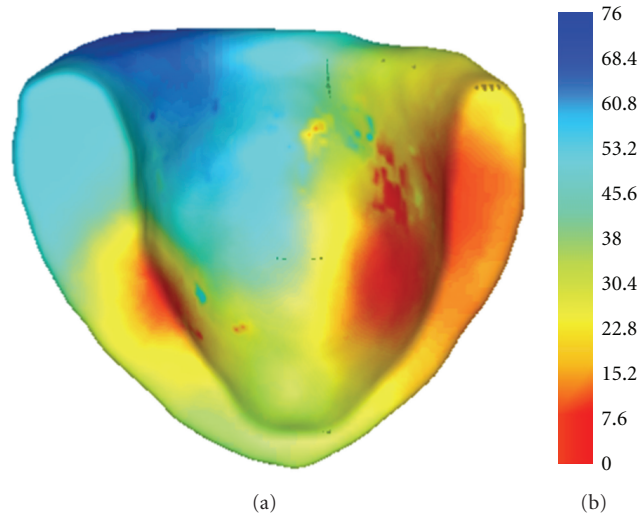
FIGURE 12: Simulation of electrical propagation in a human left ventricle (a) and its color map (b).

5. Discussion

In this paper, a numerical method without mesh constraint, EFGM, is adopted to solve reaction-diffusion equation for simulating cardiac electrical activity. This work was motivated by the successes of EFGM in mechanical modellings [24], but in our implementation, more aspects of EFGM, including the effects of influence domain (d_{max}), weight function, and integration scheme, in solving reaction-diffusion equations are evaluated.

One of main attractions of EFGM is the meshfree particle presentation, which provides not only a convenient representation, particles without predefined connectivity, of cardiac geometry and fiber orientation, but also high interpolation accuracy for dynamical process. Our tests show that convergence behaviour of EFGM will be mainly affected by the size of influence domain. In a certain range, that is, from 1 to 3 for d_{max}, the slope of convergence curve will increase along with the value of d_{max}. However, too small size of influence domain will cause singularity in system matrices, and too large size of influence domain will also introduce large error in coarse nodes and increase the assembling cost hugely in dense nodes. We also found that EFGM performance could be minimally affected by nodal positions in simulation of propagation if the nodal density does not change largely. Different weight functions also affect the accuracy of EFGM, but the performance of cubic weight function and quadratic weight function is closed, which could be selected upon user's opinion. The numerical integration of EFGM is only evaluated on a popular regular background mesh in this paper though some works proposed irregular background meshes [24, 26], because the performance of EFGM on regular background mesh is already good enough. Especially in 3D, a lower-order Gaussian quadrature scheme in one cell of regular and fine background mesh not only saves time in assembling system matrices but also achieves rational accuracy in the irregular problem domain. Hence we would recommend regular background mesh

because of easy implementation and acceptable accuracy. There has been the discussion about the construction of FEMs shape function is faster than the construction of EFGM shape function in the same spatial discretization [24], but we found that EFGM can reach a certain level of error with less computational cost than FEMs and FDMs because of higher order accuracy of EFGM shape function. Moreover, EFGM does not need to rearrange nodal positions if weight functions or polynomial bases are changed.

To fully utilize the ability of EFGM is not an easy process because wrong parameters will affect the performance of EFGM a lot, especially in 3D simulation. However, the computational cost of EFGM could be appropriately depressed by proper adjustments. First a finer background mesh with lower order quadrature, such as $2 \times 2 \times 2$ Gauss points or even one Gauss point in one background cell, is preferable to a coarser background mesh with higher order quadrature because of cheaper computation and acceptable accuracy. Second the size of influence domain should be selected as small as possible according to local nodal density, since the time to compute shape functions and their derivatives is proportional to the number of sample nodes inside the influence domain of each Gaussian integration point. The time to assemble mass matrix and stiff matrix will also increase and the spareness of those matrix will be destroyed as result of large size of influence domain. From the point of view of accuracy, there is a minimum size of influence domain to compute the shape functions and their derivatives. In our implementation we choose a big size of influence domain first and adjust the background mesh. Then we fix the background mesh and adjust the size of influence domain. After several rounds of such adjustment, we can find suitable size of influence domain and corresponding background mesh to obtain reasonable accuracy with acceptable computational cost.

EFGM offers great potentials in simulation of cardiac behaviour, especially electrical activity because of its meshless property. This kind of numerical discretization is defined

simply by placing unstructured nodes in interested area, which not only offers great convenience in implementation of adaptivity but also possibly decreases the complexity to customize the patient-specific model a lot as reasonable propagation of electrical activity in Auckland heart model could be computed in a standard desktop computer. However, further experiments with more physiological meanings, such as sustained reentry or sophisticated cellular models, in EFGM will be demanded in the future. Furthermore, a heart model with realistic geometry and components, such as with atria, ventricles, Purkinje systems, and authentic fiber structure, should be considered for better understanding of electrical activity of the whole heart.

Acknowledgment

References

[1] W. J. Germann and C. L. Stanfield, *Principles of Human Physiology*, Pearson Benjamin Cummings, 2002.

[2] D. E. Roberts, L. T. Hersh, and A. M. Scher, "Influence of cardiac fiber orientation on wavefront voltage, conduction velocity, and tissue resistivity in the dog," *Circulation Research*, vol. 44, no. 5, pp. 701–712, 1979.

[3] A. J. Pullan, M. L. Buist, and L. K. Cheng, *Mathematically Modelling the Electrical Activity of the Heart: From Cell to Body Surface and Back Again*, World Science, Singapore, 2005.

[4] F. B. Sachse, *Computational Cardiology: Modeling of Anatomy, Electrophysiology, and Mechanics*, Springer, Berlin, Germany, 2004.

[5] P. Siregar, J. P. Sinteff, N. Julen, and P. Le Beux, "An interactive 3D anisotropic cellular automata model of the heart," *Computers and Biomedical Research*, vol. 31, no. 5, pp. 323–347, 1998.

[6] C. D. Werner, F. B. Sachse, and O. Dossel, "Electrical excitation propagation in the human heart," *International Journal of Bioelectromagnetism*, vol. 2, no. 2, 2000.

[7] J. M. Rogers and A. D. McCulloch, "A collocation-Galerkin finite element model of cardiac action potential propagation," *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 8, pp. 743–757, 1994.

[8] P. C. Franzone and L. Guerri, "Spreading of excitation in 3-D models of the anisotropic cardiac tissue. I. Validation of the eikonal model," *Mathematical Biosciences*, vol. 113, no. 2, pp. 145–209, 1993.

[9] J. P. Keener, "An eikonal-curvature equation for action potential propagation in myocardium," *Journal of Mathematical Biology*, vol. 29, no. 7, pp. 629–651, 1991.

[10] K. A. Tomlinson, P. J. Hunter, and A. J. Pullan, "A finite element method for an eikonal equation model of myocardial excitation wavefront propagation," *SIAM Journal on Applied Mathematics*, vol. 63, no. 1, pp. 324–350, 2002.

[11] H. Liu, H. Hu, A. J. Sinusas, and P. Shi, "An H_∞ approach for elasticity properties reconstruction," *Medical Physics*, vol. 39, no. 1, pp. 475–481, 2012.

[12] H. Liu and P. Shi, "State-space analysis of cardiac motion with biomechanical constraints," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 901–917, 2007.

[13] H. Liu and P. Shi, "Maximum a posteriori strategy for the simultaneous motion and material property estimation of the heart," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 378–389, 2009.

[14] K. C. L. Wong, L. Wang, H. Zhang, H. Liu, and P. Shi, "Physiological fusion of functional and structural images for cardiac deformation recovery," *IEEE Transactions on Medical Imaging*, vol. 30, no. 4, pp. 990–1000, 2011.

[15] K. C. L. Wong, H. Zhang, H. Liu, and P. Shi, "Physiome-model-based state-space framework for cardiac deformation recovery," *Academic Radiology*, vol. 14, no. 11, pp. 1341–1349, 2007.

[16] R. L. Winslow, D. F. Scollan, J. L. Greenstein et al., "Mapping, modeling, and visual exploration of structure-function relationships in the heart," *IBM Systems Journal*, vol. 40, no. 2, pp. 342–359, 2001.

[17] E. M. Cherry, H. S. Greenside, and C. S. Henriquez, "A space-time adaptive method for simulating complex cardiac dynamics," *Physical Review Letters*, vol. 84, no. 6, pp. 1343–1346, 2000.

[18] P. Shi and H. Liu, "Stochastic finite element framework for simultaneous estimation of cardiac kinematic functions and material parameters," *Medical Image Analysis*, vol. 7, no. 4, pp. 445–464, 2003.

[19] M. L. Trew, B. H. Smaill, D. P. Bullivant, P. J. Hunter, and A. J. Pullan, "A generalized finite difference method for modeling cardiac electrical activation on arbitrary, irregular computational meshes," *Mathematical Biosciences*, vol. 198, no. 2, pp. 169–189, 2005.

[20] C. R. Johnson, R. S. MacLeod, and P. R. Ershler, "A computer model for the study of electrical current flow in the human thorax," *Computers in Biology and Medicine*, vol. 22, no. 5, pp. 305–323, 1992.

[21] W. J. Karlon, S. R. Eisenberg, and J. L. Lehr, "Effects of paddle placement and size on defibrillation current distribution: a three-dimensional finite element model," *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 3, pp. 246–255, 1993.

[22] A. J. Pullan and C. P. Bradley, "A coupled cubic hermite finite element/boundary element procedure for electrocardiographic problems," *Computational Mechanics*, vol. 18, no. 5, pp. 356–368, 1996.

[23] T. Belystchko, Y. Y. Lu, and L. Gu. . Int, "Element-free galerkin methods," *International Journal for Numerical Methods in Engineering*, vol. 37, pp. 229–256, 1994.

[24] T. Belytschko, Y. Krongauz, D. Organ, M. Fleming, and P. Krysl, "Meshless methods: an overview and recent developments," *Computer Methods in Applied Mechanics and Engineering*, vol. 139, no. 1–4, pp. 3–47, 1996.

[25] G. R. Liu, *Mesh Free Methods: Moving Beyond the Finite Element Method*, CRC Press, 2003.

[26] J. Dolbow and T. Belytschko, "Numerical integration of the Galerkin weak form in meshfree methods," *Computational Mechanics*, vol. 23, no. 3, pp. 219–230, 1999.

[27] I. V. Singh, "Parallel implementation of the EFG method for heat transfer and fluid flow problems," *Computational Mechanics*, vol. 34, no. 6, pp. 453–463, 2004.

[28] L. Wang, K. C. L. Wong, H. Zhang, H. Liu, and P. Shi, "Noninvasive computational imaging of cardiac electrophysiology for 3-D infarct," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 4, pp. 1033–1043, 2011.

[29] H. Zhang and P. Shi, "A meshfree method for solving cardiac electrical propagation," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS '05)*, pp. 349–352, 2005.

[30] P. J. Mulquiney, N. P. Smith, K. Clarke, and P. J. Hunter, "Mathematical modelling of the ischaemic heart," *Nonlinear Analysis, Theory, Methods and Applications*, vol. 47, no. 1, pp. 235–244, 2001.

[31] T. Rabczuk and T. Belytschko, "Adaptivity for structured meshfree particle methods in 2D and 3D," *International Journal for Numerical Methods in Engineering*, vol. 63, no. 11, pp. 1559–1582, 2005.

[32] M. Fleming, Y. A. Chu, B. Moran, and T. Belytschko, "Enriched element-free galerkin methods for crack tip fields," *International Journal for Numerical Methods in Engineering*, vol. 40, no. 8, pp. 1483–1504, 1997.

[33] L. Gavete, J. J. Benito, S. Falcón, and A. Ruiz, "Penalty functions in constrained variational principles for element free Galerkin method," *European Journal of Mechanics, A/Solids*, vol. 19, no. 4, pp. 699–720, 2000.

[34] J. Crank, *The Mathematics of Diffusion*, Oxford University Press, Oxford, UK, 2nd edition, 1975.

[35] P. J. Hunter, P. A. McNaughton, and D. Noble, "Analytical models of propagation in excitable cells," *Progress in Biophysics and Molecular Biology*, vol. 30, pp. 99–144, 1976.

[36] D. Durrer, R. T. van Dam, G. E. Freud, M. J. Janse, F. L. Meijler, and R. C. Arzbaecher, "Total excitation of the isolated human heart," *Circulation*, vol. 41, no. 6, pp. 899–912, 1970.

[37] S. Z. Zhong, L. Yuan, L. Tang et al., "Research report of experimental database establishment of digitized virtual Chinese No.1 female," *Academic Journal of the First Medical College of PLA*, vol. 23, no. 3, pp. 196–209, 2003.

Research Article

Parametric Mapping of Brain Tissues from Diffusion Kurtosis Tensor

Yuanyuan Chen,[1] Xin Zhao,[1] Hongyan Ni,[2] Jie Feng,[2] Hao Ding,[1] Hongzhi Qi,[1] Baikun Wan,[1] and Dong Ming[1]

[1] *Department of Biomedical Engineering, Tianjin University, Tianjin 300072, China*
[2] *Department of Radiology, Tianjin First Center Hospital, Tianjin 300192, China*

Correspondence should be addressed to Xin Zhao, zhaoxin@tju.edu.cn and Dong Ming, richardming@tju.edu.cn

Diffusion kurtosis imaging (DKI) is a new diffusion magnetic resonance imaging (MRI) technique to go beyond the shortages of conventional diffusion tensor imaging (DTI) from the assumption that water diffuse in biological tissue is Gaussian. Kurtosis is used to measure the deviation of water diffusion from Gaussian model, which is called non-Gaussian, in DKI. However, the high-order kurtosis tensor in the model brings great difficulties in feature extraction. In this study, parameters like fractional anisotropy of kurtosis eigenvalues (FAek) and mean values of kurtosis eigenvalues (Mek) were proposed, and regional analysis was performed for 4 different tissues: corpus callosum, crossing fibers, thalamus, and cerebral cortex, compared with other parameters. Scatterplot analysis and Gaussian mixture decomposition of different parametric maps are used for tissues identification. Diffusion kurtosis information extracted from kurtosis tensor presented a more detailed classification of tissues actually as well as clinical significance, and the FAek of D-eigenvalues showed good sensitivity of tissues complexity which is important for further study of DKI.

1. Introduction

Diffusion magnetic resonance imaging can detect the water molecule diffusion in human tissues noninvasively, which indicates the microstructure of biotissue such as one of the most popular methods: diffusion tensor imaging (DTI), in which the three-dimensional water diffusion probability distribution in an anisotropic medium has been quantified by a 2-ranked tensor. The three eigenvectors of it are corresponded to the axes of a triaxial diffusivity ellipsoid [1]. The commonly employed rotationally invariant parameters are derived from the diffusion tensor (DT) including the mean diffusivity (MD) and fractional anisotropy (FA).

However, in the conventional diffusion tensor imaging (DTI), water diffusion is assumed that diffusion appears to be a free and nonrestricted environment within a Gaussian distribution of diffusion displacement. Actually in biological tissue, complex cellular microstructure makes water diffusion a highly hindered or restricted and also non-Gaussian

process [2]. The diffusion tensor loses many details of the tissues microstructure. There are many studies about the non-Gaussian of diffusion in real tissues. DKI uses the kurtosis to estimate this non-Gaussian distribution providing insights into the microstructure of biological tissues. Recent studies have demonstrated that DK (diffusion kurtosis) measures offer an improved sensitivity in detecting developmental and pathological changes in neuronal tissues, compared to conventional DTI [3, 4]. In addition, directional kurtosis analyses have been formulated to reveal directionally specific information, such as the water diffusional kurtosis along the direction parallel or perpendicular to the principle water diffusion direction [5–8]. Because kurtosis is a measure of the deviation of the diffusion displacement profile from a Gaussian distribution, DKI analyses quantify the degree of diffusion restriction or tissue complexity.

However, what is difficult is that the high-order (3-dimensional 4-order) kurtosis tensor in DKI is complex to analyses and it is an important feature like the diffusion

tensor in DTI (diffusion tensor imaging). This paper here is to promote some parameters mapping of the kurtosis tensor.

2. Materials and Method

2.1. Theory

2.1.1. Diffusion Kurtosis Imaging. As DTI assumes Gaussian diffusion, the apparent diffusivity (D_{app}) is derived by linearly fitting the DW signals acquired with one or more nonzero b values to the following linear equation:

$$\ln\left[\frac{S(b)}{S(0)}\right] = -bD_{app}. \tag{1}$$

In DKI, logarithmic expansion of DW signal can estimate both apparent diffusivity (D_{app}) and diffusion kurtosis (K_{app}), which keep an extra b-square term compared with DTI. Thus, it forms a nonlinear equation [8]:

$$\ln\left[\frac{S(b)}{S(0)}\right] = -bD_{app} + \left(\frac{1}{6}\right)b^2 D_{app}{}^2 K_{app},$$

$$D_{app} = Dx^2, \tag{2}$$

$$K_{app} = \left(\frac{MD}{D_{app}}\right)^2 Wx^4,$$

where $S(b)$ is the DW signal intensity at a b value and $S(0)$ is the signal without diffusion gradient; x is the gradient magnetic encoding directions; MD is the mean diffusivity; here appear the diffusion tensor (D) and kurtosis tensor (W), which characterize the different diffusion motion. The theory also indicates that the apparent diffusional kurtosis approaches the true diffusional kurtosis in the limit of short gradient pulse durations, which is analogous to the relationship between D_{app} and the true water diffusion coefficient D.

2.1.2. D-Eigenvalues of Kurtosis Tensor. DKI model provides a high-order tensor except a two-order diffusion tensor, so that the problem becomes more knotty. Just like the eigenvalues and eigenvectors of diffusion tensor from which we can easily get a visualized structure model, and also more insights into the tissues microstructure, which is viable and potential via this 4-order 3-dimensional fully symmetric tensor, diffusion kurtosis tensor (DK) .

D-eigenvalues of the DK were proposed mathematically with an assumption that D tensor is always positive definite [9]. Then, a conversion was used:

$$D^{-1}(Wx^3) = \lambda x,$$

$$Dx^2 = 1. \tag{3}$$

A number of D-eigenvectors are obtained using the Sylvester formula of the resultant of a two variable system. The D-eigenvalues values are

$$\lambda_i = Wx_i^4. \tag{4}$$

2.2. Data Processing

2.2.1. Data Acquisition. The whole study and all the human experiments have got the medical ethics authentication and each subject or volunteer has knew it clearly. All human experiments were conducted on a Siemens 3. 0 T Scanner System with the physicist and 20 volunteers were normal adults in the age between 20 and 30. The DW data were acquired with SE-EPI (single shot echo-echo planar imaging) sequence, following 30 gradient magnetic encoding directions and three b values (0, 1000, 2000 ms/μm^2). Additional image parameters were that image orientation is transverse, TR = 10500 ms, TE = 103 ms, average = 1, TA = 11′14″, noise level = 30, acquisition matrix = 128 × 128, FOV = 230 × 230 mm^2, slice thickness = 1.8 mm, number of slices = 73, no gap.

For each subject a 128 × 128 × 73 × 61 metric data was acquired and prepared to fitting the tensors. Imaging processing including eddy current correction and 3D motion correction was conducted with FSL software. And for accessibility the normalization to Standard brain was used under SPM. Then the D_{app} and K_{app} were fitted using LMS method and LS method to estimate the optimal components of tensors.

2.2.2. Parameters Mapping. Although there is a way to analyse the high kurtosis tensor, the number of D-eigenvalues is random within the range between 3 and 13, even the practical meaning of these vectors and values is not clear. Anyhow, some parameters can be acquired for a further research.

In DKI, both DT and KT are obtained, and some kurtosis parameters can be formed following DT's method. Then the DTI- & DKI-derived parametric maps were analysed on the contrast between different tissues. Here are vD-eigenvalues (λ_i) or D-eigenvectors (x_i) from kurtosis tensor, FAek (FA about eigenvalues of kurtosis tensor), Mek (mean eigenvalues of kurtosis tensor), AKC (apparent kurtosis coefficient), AKCd (apparent diffusion kurtosis coefficient) are defined as following:

$$\mathrm{FAek} = \sqrt{\left(\frac{v}{(v-1)}\right)} \cdot \sqrt{\left(\frac{\left(\sum(\lambda_i - \mathrm{Mek})^2\right)}{\left(\sum \lambda_i^2\right)}\right)}$$

$$\mathrm{Mek} = \mathrm{mean}(\lambda) \tag{5}$$

$$\mathrm{MK} = \mathrm{mean}\left(K_{app}\right)$$

$$\mathrm{AKCd} = \mathrm{AKC} \cdot \mathrm{MD}^2.$$

In the mapping of MK, these negative K_{app} values are revised as zeros as no medical significance. And for comparison, we also get AKC (mean of K_{app}) without revising. Except these parameters above, this paper also considers FA (fractional anisotropy) and MD (mean diffusivity) from DTI. And that the MK (mean kurtosis) [6, 7] is considered as its popular use. Figure 1 shows some parametric maps.

2.2.3. Data Analysis. For each subject, regions of interest (ROIs) were manually defined in several transverse slices by
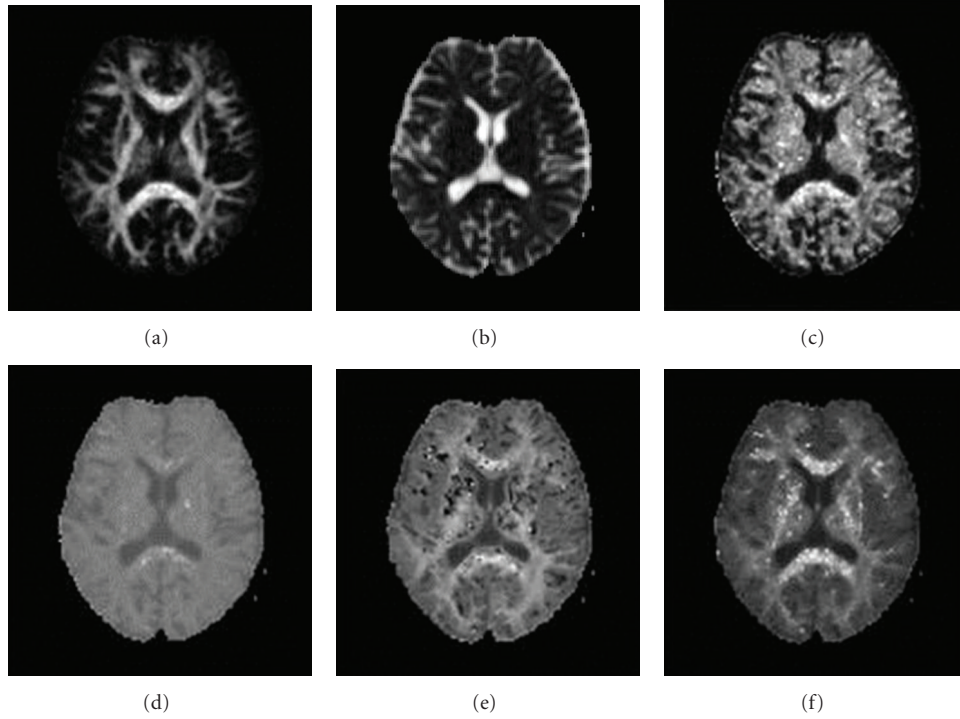
(a) (b) (c)

(d) (e) (f)

FIGURE 1: Parametric mapping from a same anatomical slice, ((a)–(f)) FA, MD, FAek, Mek, MK, and AKCd.

referencing to the anatomical structure. Anatomical landmarks were identified from both FA and MD. 6 WM structures were chosen, including the knee and the splenium of corpus callosum (CC); 4 crossing fiber areas, 4 of which are the extend areas along the knee and the splenium of CC and where are full of crossing fibers; 2 areas of thalamus; 6 GM structures, namely, 4 cerebral cortex CSF, more details from Figure 2. Various properties such as diffusivity or kurtosis can help to recognize the tissues [10, 11].

The mean and standard-variance were computed by volume averaging within the multislice ROIs for each structure. For each parameter, analysis of variance (ANOVA) was performed to compare the measurements among different tissues, followed by independent-samples t-test to detect intergroup differences.

Also in order to combining the whole image and the relativity between each other, we scattered two of the parameters with their gray histogram curves. As we already know that same tissues' voxel gray values' distribution is similarly Gaussian, and that we can assume that the histogram curves a first-order Gaussian mixed signal. The decomposition may be some single independent parameter histogram curve and also a scattered figure.

3. Results

3.1. Results of ROIs Analysis. The RIOs' statistics are figured in Figure 3. Different ROIs show pronounced average values ($P < 0.05$), and mainly five kinds of tissues in the ROIs. In a general view, the corpus callosum, cerebral cortex and CSF can be recognized obviously, while ((c)–(f)) cannot
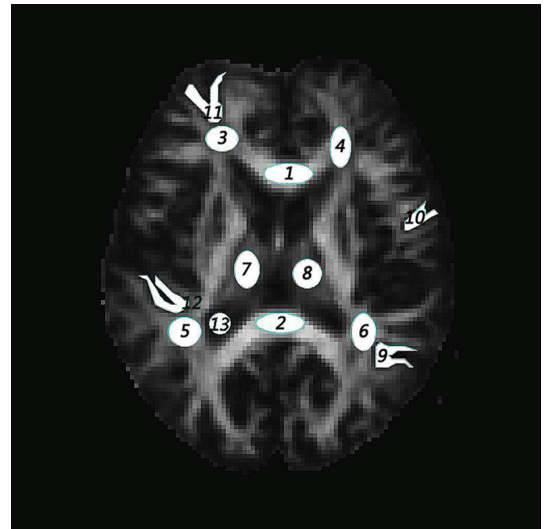

FIGURE 2: The ROIs selection according FA map. (1-2) the knee and splenial of callosum; (3–6) the crossing fibers; (7-8) the thalamus; (9–12) the cerebral cortexes; 13 is CSF.

distinguish the crossing fiber tissues and thalamus. In detail, the first two ROIs, which are two parts of corpus callosum: the knee and the splenium, result in pronounced different values ($P < 0.05$). As the fibers in the splenium are mostly more slender than the knee, and its diffusion environment is more restricted or non-Gaussian. The MK shows similar values of the two parts (1.94 ± 0.15, 1.96 ± 0.14), but it is MK that can only differs from the crossing fiber (1.67 ± 0.16)
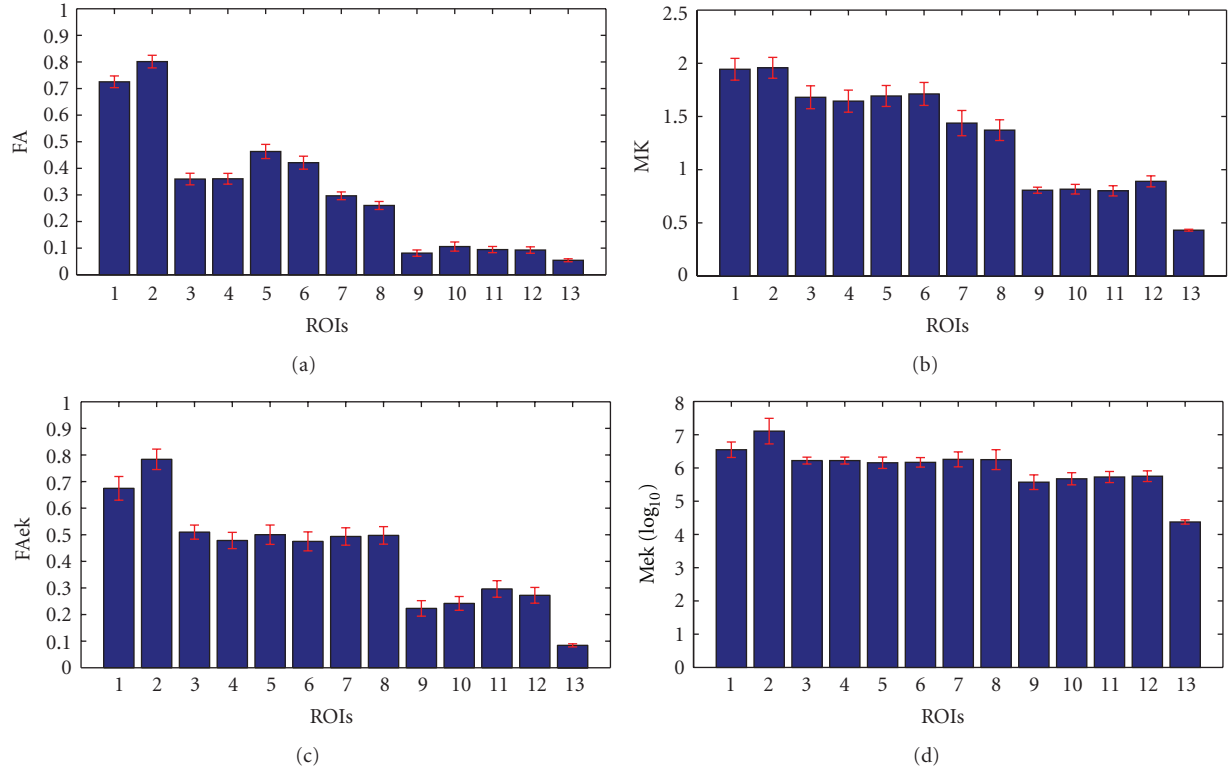
(a)

(b)

(c)

(d)

Figure 3: Mean variance of every ROIs, (1-2) the knee and splenial of callosum; (3–6) the crossing fibers; (7-8) the thalamus; (9–12) the cerebral cortexes; 13 is CSF.

and thalamus (1.42 ± 0.17) which is full of both cytons and fibers. The cerebral cortex, which mainly consists of cytons or cell membranes is which all parameters can obviously differ from white matter but the CSF in FA. What is special is that Mek is just showing the index number of the mean of D-eigenvalues, as the values is very large and resulting in a lower gray contrast between different tissues.

Having a whole picture of these ROIs tissues' structure and considering the diffusion environment, we can select the freest and the most restricted: CSF and corpus callosum especially the splenium. Then the following crossing fiber area, basal ganglia (thalamus here) and cerebral cortex are less free successively. Freer the environment is, more Gaussian the Diffusion displacement distribution is. So, MK gives a good distinguish, but not very precise; FAek distinguishes different tissues more in details. Compared with AKCd, MK does not show stably to specific structure while the AKCd performances better.

Figure 4 gives a visualized comparison of different tissues about the same property (anisotropy and kurtosis) with specific method. From the figure, FAek also performances similarly with FA, but gains better contrast in cerebral cortex (0.28 ± 0.03). FAek shows high sensibility to gray matter as well as that like thalamus. Considering Figure 4(b), the kurtosis, MK shows low gray contrast, and AKCd and MK are better recognized, but MK shows a significant difference between crossing fibers and thalamus.

3.2. Results of Histogram Analysis. With the principle that the gray values or parameters of the same characteristic tissues will be under a displacement of Gaussian function and independent from different tissues, the parametric map's histogram is decomposed using first-order Gaussian mixed signals. The mask was used in order to ignore the zeros background. MD map can gives a practical view of the tissues, so FA, FAek, Mek, and AKCd are compared with it in Figure 5, and also the kurtosis parameters' relativity are shown in Figures 5(e) and 5(f).

In Figure 5(a), FA has a wide range when MD is low which represents white matter mainly, and MD shows also a wide range when FA is low which represent gray matter and CSF mainly. But there is no relativity between them and most information is distributed where both FA and MD are low. Following Figure 5(b), FAek has more balanced distribution of histogram, an obvious subpeak, so the most information distributes where higher FAek and low MD, which indicates more sensitive to white matter. And there has some negative correlation. In Figures 5(c) and 5(d), MK and AKCd have the similar distribution with MD, while MK shows more balanced with its MK value range is wider than AKCd. But in Figures 5(e) and 5(f), the decompositions of AKCd's histogram are much independent and signi orientation is transverseficant, while the MK's are much overlapped with others. As known that Mek has low gray contrast, and here it has an obvious negative correlation with MD in Figure 5(d),
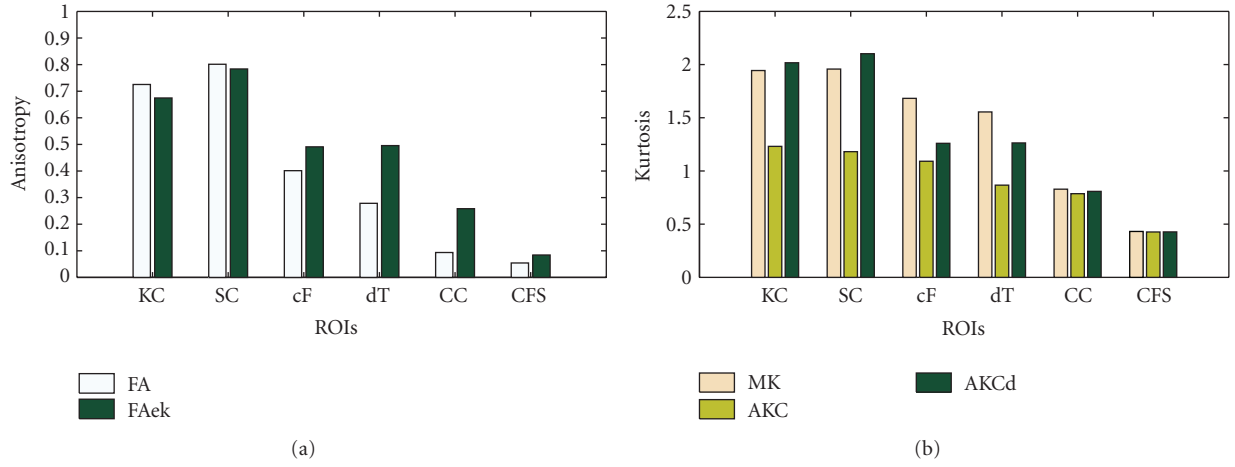
(a)

(b)

FIGURE 4: Anisotropy and kurtosis. In (a) several anisotropy values (FA, FAek) were detected averaging the volume value in different ROIs (knee of callosum, splenium callosum, crossing fiber, dorsal thalamus, cerebral cortex, CFS) as well as in (b) about kurtosis values; the statistical significance is all $P < 0.05$.
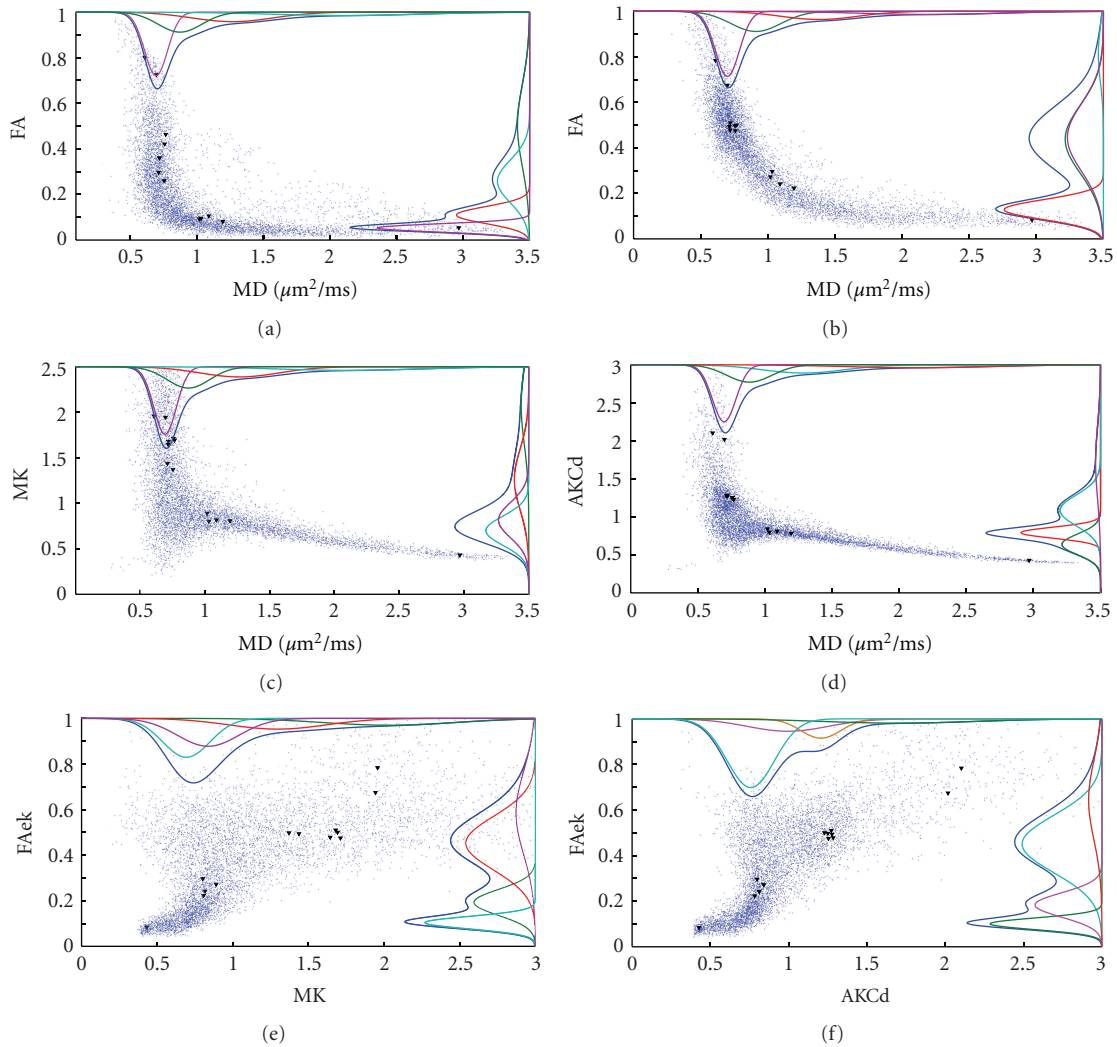

(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 5: Scatterplot analysis. ((a)–(f)) give a visualized relationship between different parameter maps, and the Gaussian decomposition of the histogram is drew along the axis. Previous ROIs for different tissues' data were also marked on it.

because the points are distributed along a similar line with MD increasing.

4. Discussion

As known theoretically, the DKI specifically the kurtosis tensor characterizes more detailed information and can show us a real insight into the microstructure of brain tissues. Practically, many work about kurtosis using DKI were carried out verifying that kurtosis provides more information and is more sensitive than diffusivity, but the kurtosis absolutely because of its complexity.

It is obvious that MK has great potential in biotissues mapping, such as tumor diagnosis and other Nerve damage disease, and kurtosis tensor can also provide an insight into the brain tissues' microstructure especially for white matter. These D-eigenvalues of kurtosis indicate more information about tissues microstructure where diffusivity cannot. However, FAek performs much better than MEK, that is to say that these D-eigenvalues can really indicate the complexity of the microenvironment but the average level of kurtosis is less sensitive. Here the cause maybe the number of the D-eigenvalues or the algorithm. What is common and important is that all kurtosis information gives more information or sensitivity about white matter microstructures [3–8].

From the black triangles in Figures 5(a) and 5(b), we can see FA recognizing cortex and thalamus as the same, but FAek recognizes crossing fibers and thalamus as the same, and FAek differs from the cortex and CSF obviously. In Figures 5(e) and 5(f), the difference is that MK recognizes the thalamus and cortex which is better than AKCd.

In the Figure 5(c), MK is the directly kurtosis calculation while the Mek is from D-eigenvalues of kurtosis tensor, which the "D-" means diffusion in original paper [9]. That is to say MK is the real kurtosis information but not accurate, but Mek is the "diffusivity" information of kurtosis tensor, because there is a stronger relativity between Mek and MD. Here "diffusivity" means the level of 2-order diffusion coefficient. But it shows actually different evaluation of the tissues, maybe just the level of complexity. In the results other parameters' classification between the crossing fibers and thalamus show that they are likely the same complexity while not exactly the diffusivity in some degree, except MK.

Kurtosis information can be used to get a more delicate classify of the tissues [12, 13]. The scatter figure and the Gaussian distribution classifying can give us ideas about image segmentation of different tissues. Within Figure 5 different two parameters put the tissues in different locations. In this part, FAek associated with AKCd and MD classified the typical tissues well. According to this, Figure 6 shows a different cluster of the scatter points.

In Figure 6, with previous knowledge, these 4 Gaussian distribution represent 4 kinds of tissues: CSF, Gray matter, crossing fibers; single diffusion orientation areas (corpus callosum). However, the third Gaussian class has a wide range which represents a collection of all crossing fiber areas as their complicated various fiber structures.

For achieving better jobs, there are several aspects need be considered further.
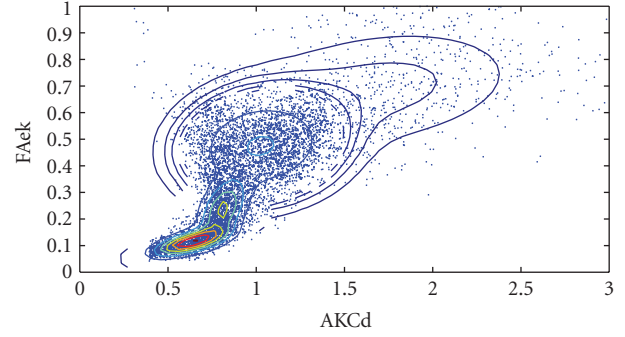

FIGURE 6: Scatter between AKCd and FAek, showing a 4 independent Gaussian distribution and probability gradients.

Firstly, the calculation of kurtosis anisotropy is just similar to DT. However, the D-eigenvalues and vectors have no clear meaning, and even more important is that DT's three eigenvectors are always of orthogonal to each other while the D-eigenvectors' orientation are random here. For example, same eigenvalues with different eigenvectors give different anisotropy properties.

Secondly, these parameters in this paper are just aimed at the contrast or discrimination between different anatomical structures at the 2-D imaging level. Anyhow this work is a start, and what's need more is the white matter microstructure and 3D space reconstruction. Several studies about the partition of diffusion or ODF have carried out aiming at the 3D space diffusion property distribution [13, 14].

Thirdly, clearer relationship between DT and KT should be clearer and what to do with them should be clear. For example, how the 4-order statistic kurtosis is related practically to the 2-order variance which suggests that diffusion coefficient here should be clear. About this, popular speaking think the kurtosis can revise or make up the deviation from Gaussian, but not in a specific way. Another idea about this is that DKI gives the Gaussian diffusion and non-Gaussian diffusion.

Finally, not least, the imaging processing to reducing noises or advancing SNR before fitting (2) is most important. Because high-order kurtosis is more sensitive to errors than diffusivity such as what appears as error spots in Figures 1(e) and 1(f) though using appropriate pre- and postprocessing. What is worse is that it will reduce the fitting results. So better signal processing is needed, or just increase the repeat times of signal acquisition (AVERAGE or NEX) but losing acquisition time (TA) for better signals from the view of experiment setting [15].

5. Conclusion

DKI is a straightforward extension of DW signals that provides a sensitive measurement of tissue structure by quantifying the non-Gaussian degree of water diffusion. DKI has been demonstrated to be highly sensitive and directionally specific in detecting brain maturation processes, and the parametric analysis of kurtosis tensor was carried out in this paper. The results indicated that more detailed insights of the

microstructure can be detected and differed from diffusivity of diffusion tensor by the kurtosis tensor.

The D-eigenvalues of kurtosis tensor give the diffusivity information about tissues complexity which is different from diffusion coefficient, but the FAek (fractional anisotropy of these eigenvalues) shows more different properties than FA of diffusion tensor, which means the level of tissues complexity. Multiparameters analysis can give more detailed tissues of human brain. Diffusion kurtosis tensor can show a more comprehensive and sensitive detection of subtle difference, but more energy should be paied for this kurtosis tensor.

Acknowledgments

References

[1] J.-B. Heidi and E. J. B. Timothy, "Diffusion MRI," in *Quantitative Measurement To in Vivo neuroanaTomy*, Academic Press, Boston, Mass, USA, 2009.

[2] M. Susumu, *Introduction To Diffusion Tensor Imaging*, Elsevier Science, Amsterdam, The Netherlands, 2007.

[3] B. Ines, G. De Groof, and M. Verhoye, "Microstructural changes observed with DKI in a transgenic Huntington rat model: evidence for abnormal neuro development," *NeuroImage*, vol. 59, pp. 957–967, 2012.

[4] E. Fieremans, J. H. Jensen, and J. A. Helpern, "White matter characterization with diffusional kurtosis imaging," *NeuroImage*, vol. 58, no. 1, pp. 177–188, 2011.

[5] E. X. Wu and M. M. Cheung, "MR diffusion kurtosis imaging for neural tissue characterization," *NMR in Biomedicine*, vol. 23, no. 7, pp. 836–848, 2010.

[6] M. M. Cheung, E. S. Hui, K. C. Chan, J. A. Helpern, L. Qi, and E. X. Wu, "Does diffusion kurtosis imaging lead to better neural tissue characterization? A rodent brain maturation study," *NeuroImage*, vol. 45, no. 2, pp. 386–392, 2009.

[7] E. S. Hui, M. M. Cheung, L. Qi, and E. X. Wu, "Towards better MR characterization of neural tissues using directional diffusion kurtosis analysis," *NeuroImage*, vol. 42, no. 1, pp. 122–134, 2008.

[8] J. H. Jensen, J. A. Helpern, A. Ramani, H. Lu, and K. Kaczynski, "Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 53, no. 6, pp. 1432–1440, 2005.

[9] L. Qi, Y. Wang, and E. X. Wu, "D-eigenvalues of diffusion kurtosis tensors," *Journal of Computational and Applied Mathematics*, vol. 221, no. 1, pp. 150–157, 2008.

[10] R. Delgado Y Palacios, A. Campo, K. Henningsen et al., "Magnetic resonance imaging and spectroscopy reveal differential hippocampal changes in anhedonic and resilient subtypes of the chronic mild stress rat model," *Biological Psychiatry*, vol. 70, no. 5, pp. 449–457, 2011.

[11] J. C. Klein, M. F. S. Rushworth, T. E. J. Behrens et al., "Topography of connections between human prefrontal cortex and mediodorsal thalamus studied with diffusion tractography," *NeuroImage*, vol. 51, no. 2, pp. 555–564, 2010.

[12] A. Iraji, E. Davoodi-Bojd, H. Soltanian-Zadeh, G. A. Hossein-Zadeh, and Q. Jiang, "Diffusion kurtosis imaging discriminates patients with white matter lesions from healthy subjects," in *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2796–2799, Boston, Mass, USA, August 2011.

[13] M. Lazar, J. H. Jensen, L. Xuan, and J. A. Helpern, "Estimation of the orientation distribution function from diffusional kurtosis imaging," *Magnetic Resonance in Medicine*, vol. 60, no. 4, pp. 774–781, 2008.

[14] C. Meier, W. Dreher, and D. Leibfritz, "Diffusion in compartmental systems. II. Diffusion-weighted measurements of rat brain tissue in vivo and postmortem at very large b-values," *Magnetic Resonance in Medicine*, vol. 50, no. 3, pp. 510–514, 2003.

[15] D. H. J. Poot, A. J. Den Dekker, E. Achten, M. Verhoye, and J. Sijbers, "Optimal experimental design for diffusion kurtosis imaging," *IEEE Transactions on Medical Imaging*, vol. 29, no. 3, pp. 819–829, 2010.

Research Article

Computerized Segmentation and Characterization of Breast Lesions in Dynamic Contrast-Enhanced MR Images Using Fuzzy c-Means Clustering and Snake Algorithm

Yachun Pang,[1] Li Li,[2] Wenyong Hu,[1] Yanxia Peng,[2] Lizhi Liu,[2] and Yuanzhi Shao[1]

[1] *School of Physics and Engineering, Sun Yat-sen University, Guangzhou 510275, China*
[2] *Imaging Diagnosis and Interventional Center, Cancer Center, Sun Yat-sen University, Guangzhou 510060, China*

Correspondence should be addressed to Yuanzhi Shao, stssyz@mail.sysu.edu.cn

This paper presents a novel two-step approach that incorporates fuzzy c-means (FCMs) clustering and gradient vector flow (GVF) snake algorithm for lesions contour segmentation on breast magnetic resonance imaging (BMRI). Manual delineation of the lesions by expert MR radiologists was taken as a reference standard in evaluating the computerized segmentation approach. The proposed algorithm was also compared with the FCMs clustering based method. With a database of 60 mass-like lesions (22 benign and 38 malignant cases), the proposed method demonstrated sufficiently good segmentation performance. The morphological and texture features were extracted and used to classify the benign and malignant lesions based on the proposed computerized segmentation contour and radiologists' delineation, respectively. Features extracted by the computerized characterization method were employed to differentiate the lesions with an area under the receiver-operating characteristic curve (AUC) of 0.968, in comparison with an AUC of 0.914 based on the features extracted from radiologists' delineation. The proposed method in current study can assist radiologists to delineate and characterize BMRI lesion, such as quantifying morphological and texture features and improving the objectivity and efficiency of BMRI interpretation with a certain clinical value.

1. Introduction

Breast cancer is the most common cancer and a leading cause of deaths in cancer for women worldwide [1]. In the United States, the chance of developing invasive breast cancer in a woman's life is nearly 1 in 8 [2]. Medical imaging, specifically, magnetic resonance imaging (MRI) plays a crucial role in detecting and diagnosing breast lesions and tumors. While mammography, as recently reported, might fail to spot up to 20 percent of tumors, MRI can detect breast cancer missed by mammography [3, 4]. Because of MRI's effectiveness in detecting breast cancer, American Cancer Society has published the guidelines for recommending women with high risk of breast cancer to receive MRI screening [5].

With its high sensitivity and variable specificity, MRI has been increasingly used for a breast cancer detection and characterization [6–8]. As a result, there is an urgent need

to develop a computer-aided diagnosis system to release radiologists from the heavy works of medical image analysis. Unfortunately, compared with mammography, relatively fewer automated CADs have been developed specifically for breast MRI. Chen et al. [9] applied the region-growing method to segment lesions and later they [10] proposed a semiautomated algorithm based on the fuzzy c-means (FCMs) clustering with the shortcoming oversensitivity to noise; Liney et al. [11] presented a user-interaction-threshold method to extract the region of interest (ROI), requiring manual intervention; Stoutjesdijk et al. [12] designed an automated computer program to select the ROIs on the basis of a mean-shift-clustering method, the method is an accurate method to automatically determine a contiguous region of interest. Shi et al. [13] used the FCMs clustering algorithm followed by a 3D level set (LS) method for segmentation refinement, and a recent paper by Meinel et al. [14] reported

a computerized segmentation method for mass-like breast MRI lesion involving robust seed-point selection, which is more reproducible than manual method in measuring the size and shape of a lesion. Because MR images are a sequence of two-dimensional images, the segmentation in 2D is still important and the basis of 3D segmentation.

Texture analysis is extensively utilized to quantify image characteristics (i.e., homogeneity and regularity with diagnosis potential in MR images). Gray level cooccurrence matrix (GLCM) method, proposed by Haralick et al. [15], is instrumental in medical image analysis. Various studies on texture analysis have been reported, including brain disease [16], bone [17], and abdominal tumor [18]. The GLCM method is also applied to the analysis of breast cancer. Chan et al. [19] put forward a method based on the texture features for discriminating mammography lesions by using linear discriminant analysis. Gibbs and Turnbull [20] manually delineated the breast MRI lesions, and then employed the GLCM method to differentiate benign and malignant lesions.

Computer-extracted morphological features have demonstrated to be of certain usefulness for characterizing breast lesions [11, 21, 22]. Breast Imaging Reporting and Data System (BI-RADS) lexicon has been introduced to categorize lesion appearance. However such lexicons are subject to radiologists' assessment. The objective computer-extracted features may benefit a radiologist to improve the interpretation and acceptability of a distinguishing feature.

In this study, we investigate systematically the segmentation and characterization of both benign and malignant breast lesions inside breast MR images using a computerized segmentation and characterization package we developed specifically for Breast MRI. The computational results of both segmentation and characterization of breast lesions are also compared with the manual delineation and the pathological results given by experienced radiologists.

2. Materials and Methods

Figure 1 shows the flowchart for our computerized breast lesion segmentation and characterization method. Our computer program performs an automated segmentation and an image analysis consequently after the manual lesion identification of a breast MRI (2D) is input. In the computerized segmentation section, FCMs clustering based method is used to produce an initial segmentation of the input image, while the gradient vector flow (GVF) snake model is applied to the initial segmentation to obtain the final segmentation. The initial segmentation method is referred to as the FCMs-based and the final segmentation method is referred to as the GVF-FCMs for short. The segmentation performance of both methods is evaluated with manual segmentation by experienced radiologists on dynamic contrast-enhanced (DCE) MRI. In the computerized characterization section, we extract morphological and texture features from both the GVF-FCMs method and radiologists' delineation. Finally, Fisher stepwise discriminant analysis (FSDA) is applied to select the features extracted by the GVF-FCMs and the
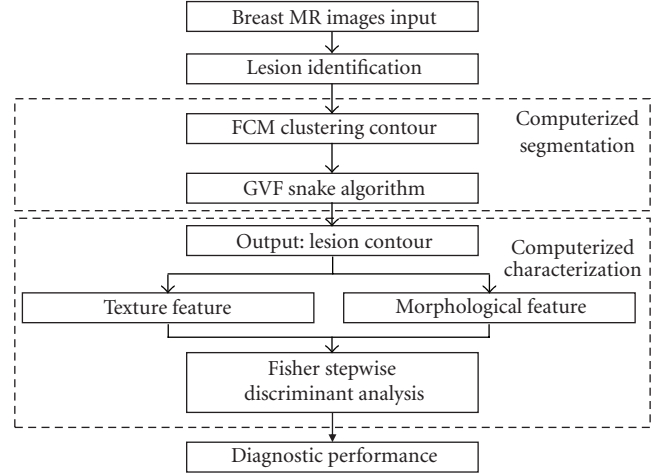

Figure 1: Flowchart of computerized lesion segmentation and characterization on breast MRI.

radiologists' manual delineation in differentiating between benign and malignant lesions.

2.1. Breast Lesions Database. This study consists of 22 benign and 38 malignant breast lesions which have been examined with a final histopathology confirmation (age range = 27–65 years old, mean age ± standard deviation = 42 ± 9 years). Only mass-like lesions that showed strong contrast enhancements were selected for this study. The database of the images for each case includes one sagittal postcontrast image slice that shows an obvious contrast enhancement and demonstrates the maximum dimension of a mass lesion. The size of the image is 512×512 pixels with a 256-gray level.

MR imaging was performed on a 1.5 T superconductive magnetic system (GE, Signa, HDx). A breast-specific 4-channel phased-array surface coil was used. Contrast medium was injected through a hand venipuncture technique. Patients were scanned in the prone position with bilateral breast naturally hanging into the two holes of the coil and their feet were first placed into the machine. A cross-sectional FSE T1WI was first employed using the following parameters: TR = 650 ms, TE: Min Full, ETL = 2, and BW = 20.83. The cross-sectional and sagittal FSE T2WI were then employed using the following parameters: TR = 4650 ms, TE = 85 ms, ETL = 16 and BW = 20.83, THK = 6 mm, spacing = 1 mm, and FOV was adjusted based on the breast size, ranging from 18 cm to 28 cm, matrix = 320×224, Frep DIR = A/P and NEX = 2. Except cross-sectional T1WI, all other sequences were fat suppression sequence. DCE-MRI was conducted after plain scan as following: (1) dynamic scanning was initiated after satisfied image quality was obtained in prescanning by simultaneously push the high-pressure syringe button and the dynamic scan button; (2) using MRI-specific high-pressure syringe (Medrad injector system, Pittsburgh) to inject 0.1 mmol/kg body weight contrast medium gadolinium diethylenetriamine penta-acetic acid (Gd-DTPA) using hand venipuncture technique at rate of 3 mL/s and then inject 10 mL saline at 3 mL/s to wash the

tube; (3) all patients accepted sagittal vibrant multitemporal DCE-MRI using 3D Fast FSPGR pulse sequence and the following parameters: FA = 12, BW = 83.33, matrix = 288 × 288, FOV = 38 mm, phase FOV = 0.90, Frep DIR = A/P, multiphase = 8~10, Zip = 2, THK = 3.4 mm and locs per slab = 50 mm; (4) the initial section of the dynamic study was obtained in the sagittal plane at 20 second intervals for 11 minutes. After that, cross-sectional and sagittal MRI was employed using fat-suppressed enhanced T1WI sequence.

2.2. Initial Segmentation.

Segmentation accuracy has a considerable influence on the subsequent characterization used to differentiate between benign and malignant breast lesions. Because of this reason, an experienced radiologist is included to identify the suspicious areas of breast lesions by firstly locating and defining a rectangle region of interest, as shown in Figure 2(a). The regions of interest serve as an input to the following sections. Then a two-step segmentation method is used to find out the accurate contour of a lesion. FCMs clustering based method is used to produce an initial segmentation of the ROI before the GVF snake for refinement is carried out.

The FCMs is an unsupervised machine learner in the pattern-recognition field and it has been widely used in image processing as well [23]. MR images always present overlapping intensities for different tissues because of the noise and blur in acquisition. The borders between different tissues are intrinsically fuzzy. The conventional (hard) clustering methods forces pixels to belong exclusively to one class. Therefore, fuzzy c-means clustering (FCMs) method allows uncertain belonging by a varying membership map and turns out to be particularly suitable for the segmentations of MR images.

In this study, the FCMs method is applied to the ROI for building the likelihood membership map (cluster number, 2; weighting exponent, 2; stop criteria, 0.0005, max iteration, 100). To binarize the membership map, we have referred to some articles [10, 13] and experimentally determined a likelihood threshold $T = 0.5$. Within the binary membership map, the processes including hole-filling, morphological opening, and two-dimensional connected-component labeling (8-connected objects) are carried out to remove the disconnections from the main lesion part. Finally, an initial segmentation is obtained with a slightly reduced size. Figure 2(b) shows an initial lesion segmentation using the FCMs-based method.

2.3. GVF Segmentation.

In this study, the gradient vector flow (GVF) snake model is applied to further refine the initial segmentation. Here the word "snake" refers to a curve that can deform under the influence by both "internal" and "external" forces [24]. GVF snake model are commonly applied to medical images because they can capture the irregular shapes and shape deformations found in anatomical structures. Its main contributions are to overcome leakage at weak boundaries in progressing snakes into concave boundary regions. As for the GVF snake model, the external force field is defined as a diffusion of the gradient vectors of

a gray-level edge map derived from the image [25]. The edge map $f(x, y)$ derived from an image $I(x, y)$ is defined as

$$f(x, y) = |\nabla I(x, y)|^2. \tag{1}$$

The GVF external field is the vector field $\mathbf{v}(x, y) = (u(x, y), v(x, y))$ that minimizes the energy functional

$$\varepsilon = \iint \mu \left(u_x^2 + u_y^2 + v_x^2 + v_y^2 \right) + |\nabla f|^2 |\mathbf{v} - \nabla f|^2 \, dxdy, \tag{2}$$

where μ is a regularization parameter governing the tradeoff between the first term and the second term in (2). According to [25], we chose $\mu = 0.1$ experimentally for the segmentation task in our study.

A GVF snake is a parameter curve defined as

$$\mathbf{x}(s) = (x(s), y(s)), \tag{3}$$

where s denotes an arc length parameter. The curve deforms iteratively until reaching a balance between the internal force $\mathbf{F}_{\text{int}}$ and the external force $\mathbf{F}_{\text{ext}}$. The internal and external forces are

$$\mathbf{F}_{\text{int}} = \alpha \mathbf{x}''(s) - \beta \mathbf{x}''''(s),$$

$$\mathbf{F}_{\text{ext}} = \mathbf{v}(x, y), \tag{4}$$

where α and β are weighting parameters that control the snake's tension and rigidity and experimentally set as 0.01 and 0 according to [24, 25]. Double and quadruple primes represent the second- and fourth-order derivatives of $\mathbf{x}(s)$, respectively. The GVF snake model is solved numerically by discretization and iteration in similar fashion to the traditional snake [24]. In the iterative procedure, the internal force prevents the snake contour from stretching and bending excessive [25], while the external force pulls the snake toward the real contour. We will set the max iterations when the snake is iterating to reach a balance. It is hard to reach a balance when the image is quite blurred and complex. Figure 2(c) shows the deformation of the GVF snakes initialized by an FCMs-based method.

2.4. Feature Extraction

2.4.1. Texture Features.

Texture is one of the intrinsic characteristics of an object, and it is important for medical image analysis [26]. Various textural algorithms have been proposed by researchers, such as fractal-based description, texture spectrum, and Markov random field model [27–29]. The GLCM texture method is widely used in medical image processing through utilizing the relative positions of pixels [15]. The matrix element $p_{\theta,d}(i, j)$ of the GLCM is the joint probability density of the occurrence for a pixel pair in an ROI with a defined distance d, direction θ, and gray levels i and j. We calculated thirteen textural measures for the nearest pixels (distance: 1 pixel) in four limited directions, 0°, 45°, 90° and 135°, respectively. Thirteen features derived from the GLCM are angular second moment, contrast, correlation, inverse difference moment, sum average, sum variance, sum
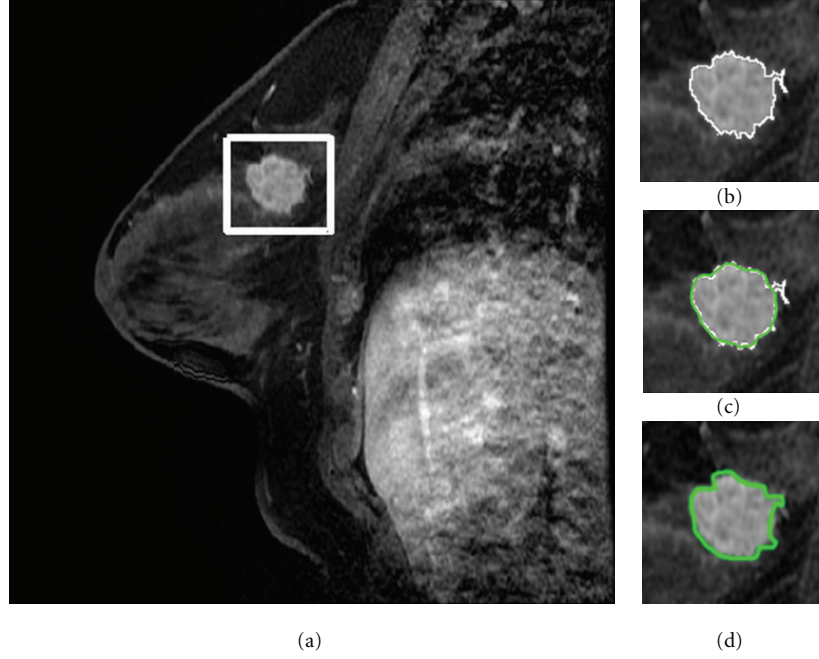
FIGURE 2: Lesion segmentation on a breast MRI scan: (a) locate a rectangle ROI box that contained a postcontrast breast MRI lesion; (b) initial segmentation by the FCMs-based method; (c) deformation of GVF snake using FCMs-based contour for initialization; (d) radiologists' manual delineation. The average time cost and dynamic memory cost of the method we proposed are 2.4180 seconds and 1256.75 KB.

entropy, entropy, difference average, difference variance, difference entropy, information measure of correlation 1, and information measure of correlation 2, respectively. Owing to the isotropic texture of the images investigated, the features we evaluated in the current study are the averages over the four directions. These texture features contain some important information on homogeneity, contrast, and other organized structures of images.

2.4.2. Morphological Features. Eight morphological features, including compactness, spiculation, extent, elongation, solidity, circularity, and entropy of radial length distribution, are selected and computed to describe the morphological properties f as defined in the Breast Imaging Reporting and Data System lexicon. Listed below are the definitions of these features.

p_1: Compactness

$$p_1 = \frac{P^2}{4\pi S},\tag{5}$$

where P and S are the perimeter length and area for a given breast MRI lesion contour, respectively.

p_2: Spiculation

$$p_2 = \frac{1}{N}\sum_{i=1}^{N}|r_i - r_{i+1}|, \ r_{N+1} = r_1,\tag{6}$$

where N is the number of pixels on the lesion contour and r_i is the individual radial length. The individual radial length is defined as the Euclidean distance from the object's center to each of contour pixels.

p_3: Extent

$$p_3 = \frac{S}{S_{\text{box}}},\tag{7}$$

where S_{box} is the area of the smallest rectangle containing the given lesion contour.

p_4: Elongation

$$p_4 = \frac{\min(H, L)}{\max(H, L)},\tag{8}$$

where H and L are the vertical and horizontal length of the smallest rectangle containing the given lesion contour.

p_5: Solidity

$$p_5 = \frac{S}{S_{\text{convex}}},\tag{9}$$

where S_{convex} is the area of the smallest convex polygon that can contain the given lesion contour.

p_6: Circularity

$$p_6 = \frac{1}{N}\sum_{i=1}^{N} r_i - \mu_r,\tag{10}$$

where μ_r is the average of r_i.

p_7: Entropy of radial length distribution

$$p_7 = -\sum p(r_i)\log(p(r_i)),\tag{11}$$

where $p(r_i)$ is the probability density of a given r_i.

p_8: Eccentricity

Eccentricity is a scalar that specifies the eccentricity of the ellipse that has the same second-moments as the lesion region. It is the ratio of the distance between the foci of the ellipse and its major axis length.

2.5. Segmentation Performance Measure. It is somewhat difficult to appraise the segmentation performance of a computerized segmentation method, because there is no golden truth in delineating accurate contour. In this paper, we take the manual delineation by two experienced radiologists in interpreting BMRI as a reference standard. All images were manually delineated by the two radiologists who were blinded to the histological results, and the disagreements were resolved by consensus. Figure 2(d) demonstrates the delineation of the radiologists.

The lesion areas extracted by the FCMs-based initial segmentation and the GVF-FCMs are compared with their counterparts segmented manually by the radiologists. Pearson's correlation coefficient (Pearson's r) and Paired Student's t-test are used to evaluate the consistency between computerized and manual segmentation. In the following discussion, A_C and A_R denote the lesion area calculated by computer and radiologists for a given lesion, respectively. $A_C \cap A_R$ means an intersection set of the lesion areas returned from both methods, while $A_C \cup A_R$ means a union set. AOR$_1$ and AOR$_2$ are defined as two overlapping measures to compare the computerized segmentation with the radiologists' delineation [10, 13] as follows:

$$\mathrm{AOR}_1 = \frac{A_C \cap A_R}{A_R},$$
$$\mathrm{AOR}_2 = \frac{A_C \cap A_R}{A_C \cup A_R}. \tag{12}$$

We calculate the AOR$_1$ and AOR$_2$ to evaluate the segmentation performance of the FCMs-based initial segmentation and the GVF-FCMs methods, respectively. Generally, a better segmentation attains when the AOR value approaches one.

2.6. Fisher Stepwise Discriminant Analysis Model. Discriminant analysis involves deriving a variate, which is a linear combination of the independent variables that would discriminate the best from a priori defined groups [30]. The method transforms the coordinates of the initial data to realize the least overlapping of the projections of data points in different groups for maximizing the diagnostic accuracy.

2.7. Statistical Analysis. The FSDA involves entering and removing features to get a statistically significant subset that predicts malignancy well, according to the discriminatory power of the subset adding to the group membership prediction [30]. Referring to [31], we set the value of the entering critical probability and the removal critical probability as $P = 0.10$ and $P = 0.15$, respectively. The FSDA is used to do the selection and classification of the features. In this study, a single database has been used for both training and testing, with the use of a "leave-one-out

cross validation" method to avoid overfit. All the diagnostic performance details were calculated by the "leave-one-out cross validation" method.

The accuracy of a model in making predictions is evaluated regularly using a ROC analysis. An ROC curve is generated by combining the true positive fraction (sensitivity) and false positive fraction (1-specificity) with different setting decision thresholds. The area under an ROC curve (AUC) is taken to estimate the classification accuracy. Generally, a larger AUC stands for a better predictive performance.

3. Results and Discussion

While an accurate delineation of lesions on breast MRI is crucial for diagnosis and associated image-guided biopsy, a slice-by-slice manual delineation by radiologists is both time-consuming and subject to interobserver and intraobserver variations [32]. Our current study involves both computerized segmentation and characterization. This study is aimed at overcoming these problems.

3.1. Segmentation Performance. Table 1 summarizes the mean values and standard deviations of the areas from the lesion contours which were segmented by the FCMs-based, GVF-FCMs and the radiologists' manual delineation, respectively. The differences between the computerized method and radiologists' manual delineation are analyzed using the Pearson's correlation coefficient (Pearson's r) and Paired Student's t-test (Table 1). The original hypothesis is that there is no significant difference between the two groups of lesion areas segmented by different methods.

Pearson's r between the lesion areas segmented by the FCMs-based method and the radiologists' manual delineation was 0.891 while the paired t-test between the areas extracted by the two methods achieves a P value of 0.105. The result indicates that the areas worked out by the two methods are highly correlated without a significant difference at the averages. After refined by the GVF method, the r and P values were both increasing, which still showed highly correlation between the areas without a significant difference at averages ($P > 0.05$). These results indicate that both the two computerized methods have certain potentials to help radiologists in an accurate delineation, and the GVF-FCMs method showed the better performance among the two methods.

Figure 3 shows the log-log scatter plot of the areas measured using the computerized method versus radiologists' manual segmentation. The lesion area is the pixels numbers in the lesion region. We drew the log-log scatter plot because the range of lesion area is wide. Judged by the distribution of the data points in Figure 3, the computerized methods have somewhat underestimated the lesion area when compared with the radiologists' reference area, since the most of the data points are distributed below the reference diagonal line. The GVF-FCMs method has the smaller underestimated. One drawback of the FCMs implementation is that the method depends simply on the intensity information and does not include the pixels' spatial relationships. For

TABLE 1: Areas, statistical comparisons and area overlap measures of computerized delineation and radiologists' manual delineation.

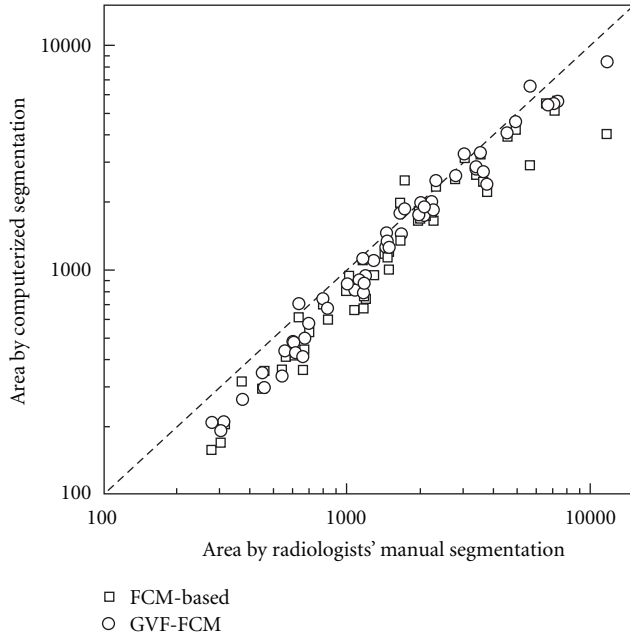| Segmentation method	Area (mean ± SD pixels)	Pearson's correlation	t-test P value	AOR_1 (mean ± SD)	AOR_2 (mean ± SD)
FCM-based	1599.5 ± 1355.4	0.891	0.105	0.75 ± 0.13	0.72 ± 0.12
GVF-FCM	1815.3 ± 1722.2	0.976	0.437	0.81 ± 0.10	0.78 ± 0.08
Radiologists' manual	2114.9 ± 2093.8	—	—	—	—

FIGURE 3: Scatter plot of the lesion areas segmented by computerized and radiologists' manual delineation. The diagonal line is represented the most perfect segmentation performance. *Square* is for areas segmented by FCMs-based initial method, *circle* is for areas extracted from GVF-FCMs method.

a more complicated lesion enhancement, it is difficult for the FCMs-based method to locate the contour that approaches near to the realistic lesion contour. The GVF-FCMs method improves the initial segmentation when deforming to a balance of internal and external forces.

Figure 4 exhibits the histograms of the overlap measures on the computerized methods: the FCMs-based and the GVF-FCMs. It turns out that all lesions segmented by the GVF-FCMs method have the values of AOR_1 and AOR_2 over 0.6. The GVF-FCMs method has the better performance in overlap measures, too. From [10], 3D segmentation over the threshold value 0.4 indicates that this method has a successful segmentation of the lesion. The threshold should be stricter in 2D segmentation and is set to 0.6. At the overlap threshold, mass lesions were all segmented correctly after the refinements by GVF method. Two sets of overlap value were compared by using the Paired Student's t-test, and the P value between AOR_1 was 0.064, while AOR_2 was 0.005. AOR_2 values were found to be statistically significant in average between the two computerized segmentation methods ($P < 0.05$).

3.2. Feature Selection and Performance of the Fisher Stepwise Discriminant Analysis Model.

For the computerized characterization part, morphological and texture features are assessed to find out whether they can be used for classifying breast lesions, and whether the features from computerized segmentation method can have a better diagnostic performance in discriminating between benign and malignant lesions. Within the two training sets, features extracted by the two methods both had no statistically significant correlations between pairs of features.

3.2.1. Features Extracted by GVF-FCMs Method.

Among two computerized segmentation methods, the GVF-FCMs method achieves the better segmentation performance. Thus GVF-FCMs method is therefore adopted in the following analysis as a preferred method for the computerized characterization. When morphological features are taken into account alone, the classifier involves three features: spiculation, eccentricity, and solidity, with an AUC of 0.883. When using GLCM texture features, however, the classifier contains four features: entropy, difference average, difference variance and information measure of correlation 1, and the classifier could attain an AUC of 0.921. When combining all the morphological and texture features, five features were selected by the classifier with the improved AUC of 0.968. They were entropy, correlation, sum average, difference average and solidity. The diagnostic measure details are shown in Table 2.

3.2.2. Features Extracted by Radiologists' Manual Delineation.

The classifier selects only one morphological feature: spiculation with an AUC of 0.836. In view of the GLCM texture features, the classifier selects three features: entropy, difference average, and information measure of correlation 1 for ROC analysis with an AUC of 0.914. When combining the morphological and texture features, only the three aforementioned texture features were selected without any morphological feature. So the AUC was the same as only using texture features. The details of diagnostic performance are given in Table 2.

3.2.3. Comparison of the Diagnostic Performance Based on Computerized and Manual Segmentation Methods.

Different morphological features are selected when using different segmentation methods. Spiculation, eccentricity, and solidity are selected when GVF-FCMs segmentation method is applied, whereas only the spiculation is selected by means of radiologists' delineations. These features are both weighting
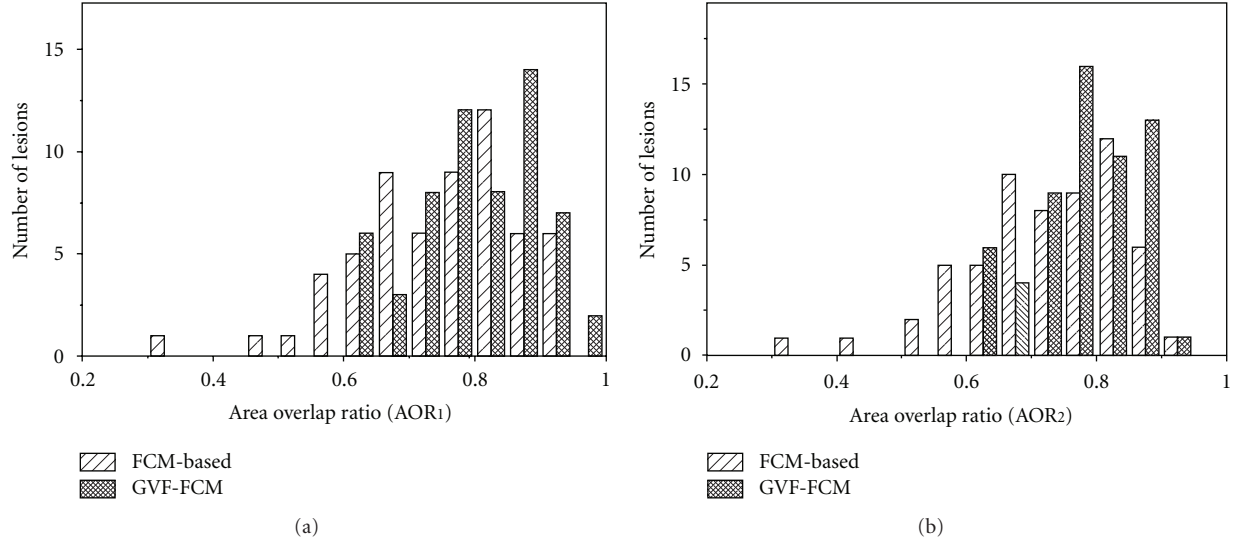
(a)

(b)

FIGURE 4: Histograms of the overlap measures on computerized methods: (a) AOR_1; (b) AOR_2. The closer the AOR value approximates to one, the better the segmentation performs. The GVF-FCMs method has the better performance among the two methods.

TABLE 2: Diagnostic performance details of the segmentation by computerized and manual delineation methods.

Segmentation Method	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)
	Morphology (three selected)	83.3	84.2	81.2
GVF-FCM	GLCM (four selected)	86.7	86.8	86.3
	Combing all features (five selected)	88.3	86.8	90.9
	Morphology (one selected)	75.0	73.7	77.3
Radiologists' manual	GLCM (three selected)	81.7	84.2	77.3
	Combing all features (three selected)	81.7	84.2	77.3

the irregularity of the contour. Generally, a spiculated contour and irregular shape are attributed to a malignant lesion while smooth contour and circle-like shape are attributed to a benign one. The computerized segmentation method can improve the discriminatory power of morphological features, comparing with the results from radiologists' delineations.

When considering texture features, the features selected by the two segmentation methods are nearly the same. entropy, difference average, and information measure of correlation 1 are all selected by the two methods, but difference variance only selected by the computerized method. Entropy is related with the heterogeneity and complexity of lesion texture. The texture feature is presumably associated with a smooth margin, homogeneous, and lower enhancements of a benign lesion in comparison with an irregular margin, heterogeneous, and higher enhancements of a malignant lesion. The diagnostic performance is similar between the texture features from different segmentation methods.

By combining the morphological and texture features, none of morphological features is selected based on the radiologists' delineation while solidity is selected by the computerized segmentation method. This possibly could be due to the coarse polygon-like contour delineated by radiologists, and the morphological features only have moderate discriminatory powers. Since the GVF-FCMs method involves stretching and bending contour until the force balance, it can fit in with the real lesion contour well, and therefore the features from the GVF-FCMs method are more eligible for the classification of a breast lesion.

3.3. Comparison of the Areas under the ROC Curve. Figure 5 displays the ROC curves of the two discriminant functions. Applying the method by Delong et al. [33], no significant difference on the two AUCs is observed between the two classifiers ($P = 0.231$). The result yields two implications: firstly, the features extracted by the computerized segmentation method have the similar discriminant power with the situation when the contour is given by radiologists; secondly, the computerized characterization of a lesion probably provides a more efficient and objective method to quantify both the appearance (texture) and shape (morphology) features.
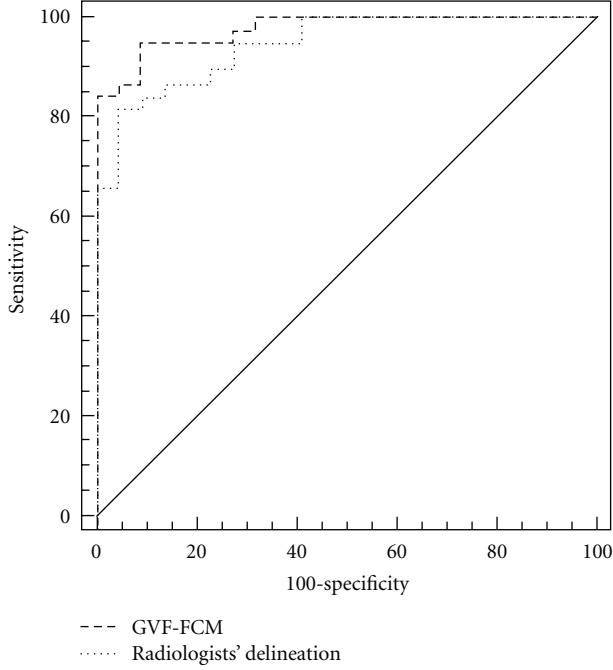
--- GVF-FCM
······ Radiologists' delineation

FIGURE 5: The ROC curves of classifer based on FSDA method by different features extracted by GVF-FCMs and radiologists' manual segmentation methods, respectively. The dotted line represented the ROC curve from radiologists' manual segmentation method with AUC of 0.914. The dash line denoted the ROC curve from computerized method (GVF-FCMs) with AUC of 0.968.

4. Conclusion

In this study, we have developed an approach based on FCMs clustering and the GVF snake model for mass-like lesion contour segmentation and computerized characterization on breast MRI. The segmentation performance measures show that the two step computerized segmentation method is an accurate method to automatically determine a suspicious lesion region and can help radiologists in their detection and delineation of breast MRI. At the computerized characterization part, Fisher stepwise discriminant analysis is used to select morphological and texture features and make classifications with the use of a "leave-one-out cross validation" method. The predictive performance based on the GVF-FCMs segmentation is better than the radiologists' manual method, but the difference is insignificant with the use of ROC curve analysis. The application of the breast MRI computerized segmentation and characterization package we developed may help radiologists to quantify the morphological and texture features and improve the objectivity and efficiency in interpreting breast MRI. In future, we intend to do further verification and assessment on a larger independent database.

Appendices

A. Fuzzy c-Means Clustering Algorithm

The fuzzy c-means clustering is an unsupervised learning method in pattern recognition [23]. The algorithm is used to minimize the objective function J as follows:

$$J = \sum_{k=1}^{c} \sum_{i=1}^{N} u_{ki}^b \|\mathbf{x}_i - \mathbf{c}_k\|^2, \qquad \text{(A.1)}$$

with constraints

$$\sum_{k=1}^{c} u_{ki} = 1, \quad 1 \le i \le N,$$

$$u_{ki} \in [0,1], \quad 1 \le i \le N, \ 1 \le k \le c, \qquad \text{(A.2)}$$

$$0 < \sum_{i=1}^{N} u_{ki} < N, \quad 1 \le k \le c,$$

where N is the number of pixels in the region of interest (ROI); c is the number of clusters (the value is set to 2); b is a weighting exponent (A.2); $\mathbf{x}_i$ is the gray-level of a pixel in the ROI; $\mathbf{c}_k$ is the cluster center (start with random assignment); u_{ik} is the likelihood that the pixel $\mathbf{x}_i$ belongs to cluster j; the $\|\cdot\|$ denotes the Euclidean distance.

The objective function may minimize only if

$$u_{ki} = \frac{1}{\sum_{j=1}^{c} \left(\|\mathbf{x}_i - \mathbf{c}_k\| / \|\mathbf{x}_i - \mathbf{c}_j\| \right)^{2/(b-1)}}, \qquad \text{(A.3)}$$

$$1 \le i \le N, \ 1 \le k \le c,$$

$$\mathbf{c}_k = \frac{\sum_{i=1}^{N} u_{ki}^b \mathbf{x}_i}{\sum_{i=1}^{N} u_{ki}^b}. \qquad \text{(A.4)}$$

The values of u_{ki} and $\mathbf{c}_k$ will iteratively update with the (A.3) and (A.4). The iteration will stop when the stop criterion is reached ($|J_{n+1} - J_n| \le 0.0005$) or the max iteration is reached (100).

B. Gray-Level Cooccurrence Matrix (GLCM) and Features Extracted from GLCM

Spatial gray-level cooccurrence matrix estimates image properties related to the second-order statistics [15]. Each element (i, j) in GLCM specifies the number of times that the pixel with gray-level value i occurred adjacent to a pixel with value j at a given offset $(\Delta x, \Delta y)$. Mathematically, GLCM element over an image S is given as

$$p(i,j) = \frac{\#\{[(x,y),(x+\Delta x, y+\Delta y)] \in S \mid f(x,y) = i, f(x+\Delta x, y+\Delta y) = j\}}{\#S}, \qquad \text{(B.1)}$$

where # represented the number of specific pixel-pair. If N is the number of distinct gray levels of an image, we denote that

$$p_x(i) = \sum_{j=1}^{N} p(i,j), \quad i = 1, 2, \ldots, N,$$

$$p_y(j) = \sum_{i=1}^{N} p(i,j), \quad j = 1, 2, \ldots, N,$$

$$p_{x+y}(k) = \sum_{i=1}^{N} \sum_{j=1}^{N} p(i,j), \quad k = i+j = 2, 3, \ldots, 2N,$$

$$p_{x-y}(k) = \sum_{i=1}^{N} \sum_{j=1}^{N} p(i,j), \quad k = |i-j| = 0, 1, \ldots, N-1.$$

(B.2)

Then the thirteen texture features are calculated as follows.

f_1: Angular second moment

$$f_1 = \sum_i \sum_j p(i,j)^2.$$

(B.3)

f_2: Contrast

$$f_2 = \sum_{k=0}^{N-1} k^2 p_{x-y}(k).$$

(B.4)

f_3: Correlation

$$f_3 = \frac{1}{\sigma_x \sigma_y} \sum_{i=1}^{N} \sum_{j=1}^{N} (ij) p(i,j) - \mu_x \mu_y,$$

(B.5)

where μ_x and σ_x are the mean and standard deviations of p_x, respectively; μ_y and σ_y are the mean and standard deviations of p_y, respectively.

f_4: Inverse difference moment

$$f_4 = \sum_i \sum_j \frac{1}{1 + (i-j)^2} p(i,j).$$

(B.6)

f_5: Sum average

$$f_5 = \sum_{k=2}^{2N} k p_{x+y}(k).$$

(B.7)

f_6: Sum variance

$$f_6 = \sum_{k=2}^{2N} (k - f_5)^2 p_{x+y}(k).$$

(B.8)

f_7: Sum entropy

$$f_7 = -\sum_{k=2}^{2N} p_{x+y}(k) \log(p_{x+y}(k)).$$

(B.9)

f_8: Entropy

$$f_8 = -\sum_i \sum_j p(i,j) \log(p(i,j)).$$

(B.10)

f_9: Difference average

$$f_9 = \sum_{i=1}^{N} (i - \mu)^2 p_x(i),$$

(B.11)

where μ is the average of $p(i,j)$.

f_{10}: Difference variance

$$f_{10} = \sum_{k=0}^{N-1} (k - d) p_{x-y}(k),$$

(B.12)

where d is the mean of p_{x-y}.

f_{11}: Difference entropy

$$f_{11} = -\sum_{k=0}^{N-1} p_{x-y}(k) \log(p_{x-y}(k)).$$

(B.13)

f_{12}: Information measure of correlation 1

$$f_{12} = \frac{f_9 + \sum_i \sum_j p(i,j) \log(p_x(i) p_y(j))}{\max(-\sum_j p_x(i) \log(p_x(i)), -\sum_j p_y(i) \log(p_y(i)))}.$$

(B.14)

f_{13}: Information measure of correlation 2

$$f_{13} = \sqrt{1 - \exp(-2(E - f_8))},$$

(B.15)

where $E = -\sum_i \sum_j p_x(i) p_y(j) \log(p_x(i) p_y(j))$.

Authors' Contribution

The first and second author contributed equally to this study.

Acknowledgment

References

[1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[2] A. C. Society. Learn about breast cancer, http://www.cancer.org/.

[3] C. D. Lehman, C. Gatsonis, C. K. Kuhl et al., "MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer," *The New England Journal of Medicine*, vol. 356, no. 13, pp. 1295–1303, 2007.

[4] N. Shute, "Beyond Mammograms: research to improve breast cancer screening focuses on sound, light, breath and tissue elasticity," *Scientific American*, vol. 304, pp. 32–34, 2011.

[5] D. Saslow, C. Boetes, W. Burke et al., "American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography," *CA: A Cancer Journal For Clinicians*, vol. 57, pp. 75–89, 2007.

[6] C. Kuhl, "The current status of breast MR imaging. Part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice," *Radiology*, vol. 244, no. 2, pp. 356–378, 2007.

[7] C. K. Kuhl, "Current status of breast MR imaging: part 2. Clinical applications," *Radiology*, vol. 244, no. 3, pp. 672–691, 2007.

[8] J. C. Weinreb and G. Newstead, "MR imaging of the breast," *Radiology*, vol. 196, no. 3, pp. 593–610, 1995.

[9] W. Chen, M. L. Giger, L. Lan, and U. Bick, "Computerized interpretation of breast MRI: investigation of enhancement-variance dynamics," *Medical Physics*, vol. 31, no. 5, pp. 1076–1082, 2004.

[10] W. Chen, M. L. Giger, and U. Bick, "A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images," *Academic Radiology*, vol. 13, no. 1, pp. 63–72, 2006.

[11] G. P. Liney, M. Sreenivas, P. Gibbs, R. Garcia-Alvarez, and L. W. Turnbull, "Breast lesion analysis of shape technique: semi-automated vs. manual morphological description," *Journal of Magnetic Resonance Imaging*, vol. 23, no. 4, pp. 493–498, 2006.

[12] M. J. Stoutjesdijk, J. Veltman, H. Huisman et al., "Automated analysis of contrast enhancement in breast MRI lesions using mean shift clustering for ROI selection," *Journal of Magnetic Resonance Imaging*, vol. 26, no. 3, pp. 606–614, 2007.

[13] J. Shi, B. Sahiner, H. P. Chan et al., "Treatment response assessment of breast masses on dynamic contrast-enhanced magnetic resonance scans using fuzzy c-means clustering and level set segmentation," *Medical Physics*, vol. 36, no. 11, pp. 5052–5063, 2009.

[14] L. A. Meinel, T. Buelow, D. Huo et al., "Robust segmentation of mass-lesions in contrast-enhanced dynamic breast MR images," *Journal of Magnetic Resonance Imaging*, vol. 32, no. 1, pp. 110–119, 2010.

[15] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[16] V. A. Kovalev, F. Kruggel, H. J. Gertz, and D. Y. Von Cramon, "Three-dimensional texture analysis of MRI brain datasets," *IEEE Transactions on Medical Imaging*, vol. 20, no. 5, pp. 424–433, 2001.

[17] R. Tuli, S. Tuli, S. Nandi et al., "Characterization of multi-potential mesenchymal progenitor cells derived from human trabecular bone," *Stem Cells*, vol. 21, no. 6, pp. 681–693, 2003.

[18] D. Mitrea, P. Mitrea, S. Nedevschi et al., "Abdominal tumor characterization and recognition using superior-order cooccurrence matrices, based on ultrasound images," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 348135, 17 pages, 2012.

[19] H. P. Chan, D. Wei, M. A. Helvie et al., "Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space," *Physics in Medicine and Biology*, vol. 40, no. 5, pp. 857–876, 1995.

[20] P. Gibbs and L. W. Turnbull, "Textural analysis of contrast-enhanced MR images of the breast," *Magnetic Resonance in Medicine*, vol. 50, no. 1, pp. 92–98, 2003.

[21] L. A. Meinel, A. H. Stolpen, K. S. Berbaum, L. L. Fajardo, and J. M. Reinhardt, "Breast MRI lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system," *Journal of Magnetic Resonance Imaging*, vol. 25, no. 1, pp. 89–95, 2007.

[22] K. Nie, J. H. Chen, H. J. Yu, Y. Chu, O. Nalcioglu, and M. Y. Su, "Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI," *Academic Radiology*, vol. 15, no. 12, pp. 1513–1525, 2008.

[23] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, vol. 4, Kluwer Academic, 1999.

[24] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[25] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, 1998.

[26] G. Castellano, L. Bonilha, L. M. Li, and F. Cendes, "Texture analysis of medical images," *Clinical Radiology*, vol. 59, no. 12, pp. 1061–1069, 2004.

[27] C. Jiang, R. E. Pitt, J. E. A. Bertram, and D. J. Aneshansley, "Fractal-based image texture analysis of trabecular bone architecture," *Medical and Biological Engineering and Computing*, vol. 37, no. 4, pp. 413–418, 1999.

[28] L. Wang and D. C. He, "Texture classification using texture spectrum," *Pattern Recognition*, vol. 23, no. 8, pp. 905–910, 1990.

[29] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 1, pp. 25–39, 1983.

[30] J. Hair, R. L. Tatham, R. E. Anderson, and W. Black, *Multivariate Data Analysised*, Prentice-Hall, 5th edition, 1998.

[31] T. P. Beresford, F. C. Blow, E. Hill, K. Singer, and M. R. Lucey, "Comparison of CAGE questionnaire and computer-assisted laboratory profiles in screening for covert alcoholism," *The Lancet*, vol. 336, no. 8713, pp. 482–485, 1990.

[32] S. Mussurakis, D. L. Buckley, A. M. Coady, L. W. Turnbull, and A. Horsman, "Observer variability in the interpretation of contrast enhanced MRI of the breast," *British Journal of Radiology*, vol. 69, no. 827, pp. 1009–1016, 1996.

[33] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, 1988.

Research Article

Quantitative Imaging of Young's Modulus of Soft Tissues from Ultrasound Water Jet Indentation: A Finite Element Study

Min-Hua Lu,[1, 2, 3] Rui Mao,[4] Yin Lu,[3] Zheng Liu,[3] Tian-Fu Wang,[1, 2, 3] and Si-Ping Chen[1, 2, 3]

[1] *National-Regional Engineering Laboratory for Key Technology of Medical Ultrasound, Shenzhen University, Shenzhen 518060, China*

[2] *Guangdong Key Laboratory of Biomedical Information Detection and Ultrasound Imaging, Shenzhen University, Shenzhen 518060, China*

[3] *Department of Biomedical Engineering, School of Medicine, Shenzhen University, Shenzhen 518060, China*

[4] *Shenzhen Key Labratory of Service Computing and Applications, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China*

Correspondence should be addressed to Min-Hua Lu, minhua_lu@hotmail.com

Indentation testing is a widely used approach to evaluate mechanical characteristics of soft tissues quantitatively. Young's modulus of soft tissue can be calculated from the force-deformation data with known tissue thickness and Poisson's ratio using Hayes' equation. Our group previously developed a noncontact indentation system using a water jet as a soft indenter as well as the coupling medium for the propagation of high-frequency ultrasound. The novel system has shown its ability to detect the early degeneration of articular cartilage. However, there is still lack of a quantitative method to extract the intrinsic mechanical properties of soft tissue from water jet indentation. The purpose of this study is to investigate the relationship between the loading-unloading curves and the mechanical properties of soft tissues to provide an imaging technique of tissue mechanical properties. A 3D finite element model of water jet indentation was developed with consideration of finite deformation effect. An improved Hayes' equation has been derived by introducing a new scaling factor which is dependent on Poisson's ratios v, aspect ratio a/h (the radius of the indenter/the thickness of the test tissue), and deformation ratio d/h. With this model, the Young's modulus of soft tissue can be quantitatively evaluated and imaged with the error no more than 2%.

1. Introduction

Physiologic processes may change tissue properties significantly. Tissue elasticity is one of the typical mechanical properties which generally correlate with pathological changes [1], such as cancer [2] and osteoarthritis [3]. During the past a few decades, there have been many efforts to develop various techniques to measure or image the elasticity of soft tissue, such as elastography [4–6], indentation [7–9], and atomic force elasticity microscopy [10–12].

Indentation is one of the most commonly used method to measure mechanical properties of soft tissues *in situ* or *in vivo*, because it does not require special preparation of regular shaped tissues and can be employed to perform tests on small specimens [13]. A typical application of indentation is to assess the degeneration of articular cartilage (AC). AC is

often characterized as a single phase, isotropic, homogeneous linear elastic model when no interstitial fluid flows during instantaneous, and equilibrium responses [7, 8]. Rigid cylindrical flat-ended or spherical indenters have been employed in early models of indentation, and the Young's modulus of soft tissues [14–16] can be calculated using the following equation given by Hayes et al. [7]:

$$E = \frac{(1 - v^2)}{2ak(v, a/h)} \cdot \frac{F}{d}, \tag{1}$$

where F is the indentation force, v is the Poisson's ratio of soft tissues, a is the radius of indenter, h is the tissue thickness, d is the indentation depth, and k is a scaling factor, which depends on aspect ratio a/h and Poisson's ratio v. It is emphasized that the deformation was assumed to be small,

which is difficult to be controlled when indentation test is performed manually [9]. Zhang et al. developed a finite element model using (1) by including the effects of large deformation up to15% strain, and a new set of k values were calculated [14].

Young's modulus can be calculated from (1) by using experimentally obtained data of F and d. The indentation force F is normally recorded by a force sensor. The deformation d can be measured by optical [17, 18], needle probe [19–21] or ultrasound [9, 22–25] methods. Poisson's ratio v is conventionally assumed to be an assigned value [7, 15] or separately measured using other methods [13, 16, 26]. Among them, ultrasound measurement provides a non-invasive and accurate tool for obtaining both tissue thickness and deformation simultaneously. Nevertheless, traditional ultrasound indentation typically operates at the frequency range between 2 and 10 MHz and is normally used to measure the mechanical properties of entire tissue layers. Its resolution is not sufficient to map the mechanical properties of soft tissues with fine structures. Moreover, most of current indentation instruments use a contact way, so that tissue damage caused by the measurement instrumentation cannot be avoided. High frequency ultrasound can improve the resolution to a microscopic level. However, for some technical reason, high frequency ultrasound transducers are not suitable for traditional contact indentation [27]. Considering that ultrasound can propagate through water with very small attenuation, Lu et al. [28] developed a water jet system to achieve noncontact high frequency (20–50 MHz) ultrasound indentation. Water jet not only serves as a soft indenter but also as the coupling medium for high-frequency ultrasound.

The ultrasound water jet indentation system has been employed to obtain modulus image of soft tissues [27], to assess articular cartilage degeneration [29] and to evaluate the bone-tendon junction healing progress [28], which has shown great potential to image the modulus distribution of soft tissues for clinical assessment and diagnosis, and/or to perform indentation tests on small specimens at microscopic levels for biological tissues and other materials. However, all the studies mentioned above measured the stiffness ratio of soft tissue as an indicator of tissue pathological state, which is not the intrinsic property of soft tissue. There's still large discrepancy between the values of stiffness ratio and Young's modulus.

In this study, we simulated the water jet indentation using finite element (FE) analysis to investigate the interaction between fluid (water jet) and solid (soft tissue). With this FE model, the Young's modulus of soft tissue can be calculated using an improved indentation solution based on Hayes' equation by introducing the geometry-, material- and deformation-dependent factor k. A new set of scaling factor k is presented by considering the finite deformation effect of indentation.

2. Methods

The water jet indentation system consists of a 3D translating device (Arthroscopic supporting arm, Medtronic Inc., MN,
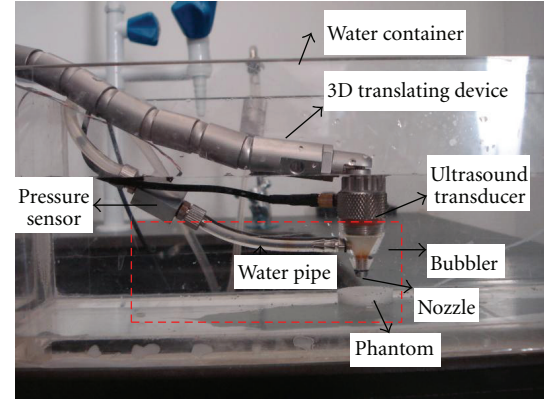

FIGURE 1: Diagram of the noncontact ultrasound indentation system. The 3D translating device facilitated the system to move easily and monitor the deformation of the phantom. Water flowed from the water pipe and water jet was used as an indenter. Focused high-frequency ultrasound was transmitted through water. The central square region, including a long water pipe, a bubbler, a nozzle, and the soft phantom, was modeled in ANSYS WORKBENCH (see Figure 2). Pressure sensor was located 60 mm from the middle of the bubbler. The dimensions of the important components are: nozzle diameter 1.7 mm, water supply pipe diameter 2 mm, the height of nozzle is 8.5 mm and the distance from the nozzle to tissue approximately 0.95 mm.

USA), a water container, a pressure sensor (EPB-C12, Entran Devices, Inc., Fairfield, NJ, USA), a water pipe, an ultrasound transducer (SEUT-506, Acoustic Sensor Co., Ltd., Taiwan), a bubbler, and a nozzle (Figure 1). The 3D translating device facilitates the system to move easily to adjust the distance from the transducer to the tissue sample, and to apply C-scan to obtain modulus image of the region of interest. Focused high-frequency ultrasound is transmitted through the bubbler when it is filled up with water. The central frequency of the ultrasound transducer is 50 MHz, with focal length at 12 mm, and aperture size at 6 mm. The dimensions of the important components are nozzle diameter 1.7 mm, water supply pipe diameter 2 mm, the height of nozzle is 8.5 mm, and the distance from the nozzle to tissue approximately 0.95 mm which is determined by adjusting the ultrasound beam focused at tissue surface. Pressure sensor was located 60 mm from the middle of the bubbler.

A three-dimensional finite element (FE) model whose geometry was as same as the experimental system was established using ANSYS (version 11.0, Canonsburg, PA, USA) to simulate indentation (Figure 2). The interaction between the water jet and soft tissue involves fluid and structural solid coupling, therefore, the simulation was performed by computational fluid dynamics (CFD) analysis that was performed in ANSYS CFX 11 and ANSYS 11 structural codes, coupled through the ANSYS MFX solver. One-way fluid-structure interaction (FSI) theory was applied in this problem.

2.1. Model Geometries. The fluid domain model is shown as Figure 2. The dimensions of the important components are
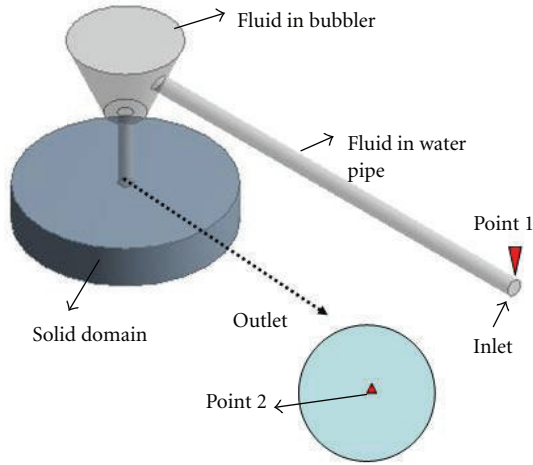
FIGURE 2: Model of the water jet indentation in ANSYS. The upper part represents fluid domain, which included a long water pipe with 60 mm in length, a bubbler, and a nozzle. To make the outlet clearer, an enlarged part was displayed at the right bottom. Radius of inlet was 1 mm and outlet 0.85 mm. Point 1, the upper point of inlet, and Point 2, the central point of outlet, were set as two monitoring points. The lower part represents solid domain. Diameters of the tissue phantom was 25 mm and its thickness was 5 mm.

totally same as the experimental system described above. The CFX-Mesh method was applied to mesh the fluid domain. With consideration of velocity gradients in near-wall regions, inflation theory was used. In our study, when the flow rate at the inlet was increased from 1 m/s to 10 m/s, different meshes influenced the results significantly. After a proper grid sensitivity analysis, the number of inflated layers was set as 11 and the maximum thickness of inflated boundary was set as 0.1 mm. Triangle elements were generated outside of the boundary layer regions. Refinement of the mesh is implemented at the boundaries. Totally a mesh of 665242 nodes and 2877024 elements were employed for all simulations.

The model of solid part is built up as a cylinder soft tissue with its thickness at 5 mm and diameter at 25 mm. The patch conforming mesher under tetrahedrons is used. Regular meshes consisting of about 101911 nodes and 65034 elements were adopted, after a proper grid sensitivity analysis, for all the models developed.

2.2. Boundary Conditions. The simulated indentation test considers a specimen supported by a rigid impermeable plate and indented by water jet from the nozzle. One-way FSI was adopted in our study, therefore, the fluid part and solid part were modeled separately. After fluid models were solved, pressure calculated from fluid outlet was mapped to the contact region of soft tissue.

The following boundary conditions were imposed on all the fluid models. (1) Inlet boundary was adopted at the beginning of the water pipe, with speed varied from 1 m/s to 10 m/s and intensity of 5% of turbulence model was used. (2) Outlet boundary was defined at the interface of fluid and relative pressure was 0 Pa. (3) Wall influence on flow is

defined as no slip and wall roughness as smooth wall. As to the solid model, the nodes were constrained in the vertical direction at the bottom of the specimen and pressure was mapped from the fluid outlet to specimen surface.

2.3. Material Properties. Water at 22°C was used for the fluid models. Soft tissue was assumed to be a linear-elastic, homogeneous, and isotropic thin layer adhere to a rigid foundation [15, 16]. The mechanical properties were described by Young's modulus E and Poisson's ratio v. Different values of Young's modulus obtained from literature review and our previous studies, 10 kPa, 52 kPa, 146 kPa, 270 kPa, 740 kPa, and 1000 kPa were assigned to the tissue model in FE analysis to mimic human normal liver, human diseased liver, breast benign lesion, breast malignant lesion, cancerous skin, and articular cartilage, respectively, [30–32]. Density of the soft tissue was 1060 kg/m^3.

2.4. Extract of Young's Modulus of Soft Tissue. Different from the traditional indentation, our water-jet indenter can be regarded as a "soft indenter", therefore, Hayes' equation cannot be used to derive Young's modulus from the force-displacement curve obtained with our experimental system anymore. An improved indentation solution was proposed by taking into account the finite deformation effect as

$$E = \frac{(1 - v^2)}{2ak(v, a/h, d/h)} \cdot \frac{F}{d}, \tag{2}$$

therefore,

$$k\left(v, \frac{a}{h}, \frac{d}{h}\right) = \frac{(1 - v^2)}{2aE} \cdot \frac{F}{d}, \tag{3}$$

where E is Young's modulus, F is the indentation force which was calculated by the FE simulation, v is the Poisson's ratio of soft tissues which was generally assumed as the values reported by literature review, a is the radius of indenter and h is the tissue thickness which can be obtained in advance before an indentation test, d is the indentation depth which was assigned to the nodes on the upper surface of the indenter during the simulation of indentation test, and k is a scaling factor, which depends on aspect ratio a/h, deformation ratio d/h, and Poisson's ratio v.

The simulated force-deformation data were used to calculate the Young's modulus and then derive the k value. These new k values were calculated using (3). Then the Young's modulus of the soft tissues under water jet indentation can be calculated by substituting these new k-values to (2). Different fixed Poisson's ratio ($v = 0.1$–0.5), indentation depth (0.1%–10%), and aspect ratio a/h (0.17, 0.4, 0.6, 0.8, and 1) were also assigned to the tissue model to investigate the effects of these factors on the estimation of Young's modulus. Therefore, we have a total of 1650 simulation processes in this study.

3. Results

First, the relationship between the pressure (or flow rate) measured at the inlet and the total force applied to the soft
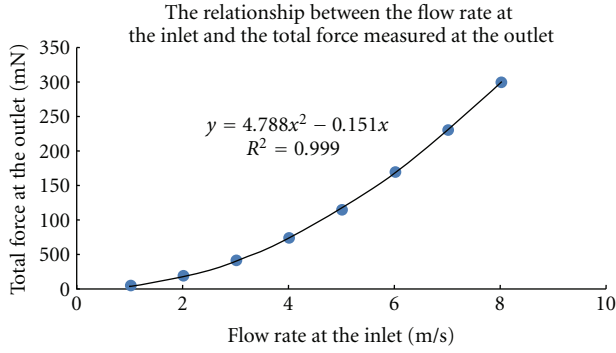
FIGURE 3: The relationship between the flow rate measured from Point 1 at inlet and the force calculated from Point 2 at outlet.

FIGURE 4: The correlation between the pressure and the flow rate measured both at Point 1 from inlet.

tissue was analyzed. This relationship is crucial because it is difficult for us to directly measure the force applied to the soft tissue by water jet during experiment. Two measurement points were set at the inlet and outlet (Points 1 and 2 as indicated in Figure 2), respectively. Figure 3 shows a quadratic function that fit the flow rate measured from Point 1 at inlet and the force calculated from Point 2 at outlet. Figure 4 shows the relationship between the flow rate and the pressure measured both at Point 1 at inlet. The nonlinear but monotonic relationship indicates that we may measure only pressure or flow rate at the inlet to calculate the force applied to the soft tissue. It will highly facilitate the water jet indentation when it is used in clinical applications.

Table 1 shows the k values with different aspect ratios (0.17–1.0), Poisson's ratios (0.1–0.5), and deformation ratios (0.01–0.1). The k values increased with the increase of aspect ratios and Poisson's ratios. However, the nonlinear and nonmonotonic relationships between the k values and deformation ratio were found.

In comparison with the ground truth, the percentage errors of the calculated Young's moduli are between -0.07% and -1.47% when the aspect ratio varied from 0.17 and 1, which suggested a quite good agreement between the actual and calculated values of Young's modulus.

4. Discussion

In this study, a 3D FE model was constructed to simulate the water jet indentation on soft tissues. One-way fluid-solid coupling analysis was conducted to find the relationship among the parameters of the indentation force, tissue deformation, tissue Young's modulus and Poisson's ratio and the aspect parameters, including the indenter size and tissue sample size. An improved Hayes' equation was developed to calculate Young's modulus of soft tissue from the force-deformation curve by introducing a new scaling factor k. A new set of k values for different Poisson's ratio v, aspect ratio a/h, and strain level d/h were calculated. We also investigated the relationship of the water parameters, such as pressure and flow rate, in the water jet indentation system and the influence of these parameters on the indentation force. The percentage errors of the estimated Young's modulus ranged
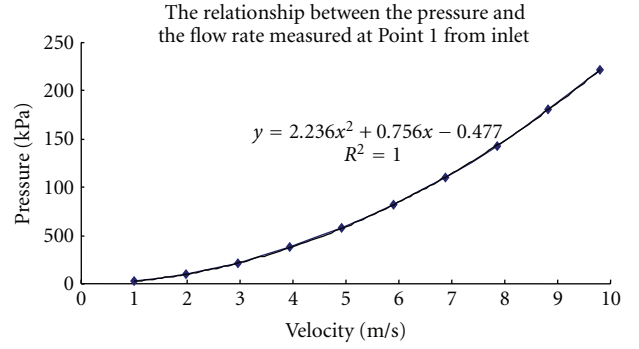
from -0.07% to -1.47% in comparison with the assigned values in FE simulation when the aspect ratio a/h ranged from 0.17 to 1. This robustness of the estimation could be further improved by averaging the parameters calculated using different deformation levels.

It has been demonstrated that our system can measure the Young's modulus of the soft tissue under water jet indentation at one single site. As shown in Figure 1, a flexible arm is used as the supporting arm to control the water jet indenter in this system. This arm, actually utilized the system to conduct C-scan over the tissue sample easily. C-scan imaging provides a useful view of an object, showing a plane perpendicular to the ultrasound beam. It has been widely used in the aerospace industry to detect the surface corrosion, delaminations, voids, cracks and other faults in aging aircraft [33, 34]. In our study, it's easy to apply C-scan with different water pressures for C-scan sequences. After analyzing both the ultrasound signal and flow rate collected from different C-scans, the modulus image could be obtained.

The water jet ultrasound indentation system has been proven that it is useful for the assessment of tissue pathology. Tissue elasticity is one important parameter which generally correlates with tissue pathological changes. This study, for the first time, extracts Young's modulus of soft tissue from water jet indentation using FE analysis. In our FE model, soft tissue is assumed as a linear elastic, homogeneous and isotropic material. However, biphasic theory suggests that most tissues are composed of solid and fluid materials, and they possess very complicated structure-function behaviors and exhibit time-dependent behavior, that is, nonlinear or visco-elastic, heterogeneous and anisotropic behavior. Normally, solid matrix represents the elastic properties while fluid materials represent the viscous properties. In order to interpret these complicated behaviors of soft tissues, the acquired force-deformation data should be interpreted carefully by taking the biphasic theory of soft tissue into accounts.

This study only focused on the nonlinear effect caused by the finite deformation of indentation. The nonlinear and visco-elastic properties of soft tissues have not been addressed yet. Studies on nonlinear tissues' models [35] and indentation of anisotropic biomaterials [36] have been reported. In the future study, the effects of nonlinear

TABLE 1: The scaling factor (k) used in this study for different aspect ratios (a/h), Poisson's ratios (v), and deformation ratios (d/h).

	Poisson's ratio					
d/h	0.10	0.20	0.30	0.40	0.45	0.50
$a/h = 0.17$						
0.001	0.8009	0.8050	0.8112	0.8215	0.8306	0.8379
0.01	0.7924	0.7967	0.8043	0.8171	0.8277	0.8385
0.05	0.7532	0.7565	0.7662	0.7858	0.8015	0.8248
0.07	0.7343	0.7354	0.7471	0.766	0.7825	0.8102
$a/h = 0.4$						
0.001	0.9307	0.9422	0.9617	0.9963	1.0254	1.0909
0.01	0.9266	0.9384	0.9589	0.9957	1.0261	1.0961
0.05	0.9074	0.92	0.9437	0.9886	1.0261	1.1129
0.1	0.8827	0.8945	0.9199	0.9712	1.0149	1.1186
$a/h = 0.6$						
0.001	1.0721	1.0918	1.1278	1.1948	1.2526	1.3683
0.01	1.0695	1.0896	1.1267	1.1965	1.2564	1.3782
0.05	1.0572	1.079	1.1204	1.2011	1.2715	1.4181
0.1	1.0407	1.0632	1.1087	1.2002	1.2822	1.4643
$a/h = 0.8$						
0.001	1.2303	1.2628	1.3208	1.4338	1.5346	1.7026
0.01	1.2289	1.2621	1.3217	1.4368	1.542	1.7191
0.05	1.2225	1.2581	1.3240	1.4566	1.5771	1.7923
0.1	1.2135	1.2513	1.3237	1.4744	1.6152	1.8761
$a/h = 1.0$						
0.001	1.4239	1.4715	1.5601	1.7385	1.9	2.197
0.01	1.4238	1.4722	1.5628	1.7465	1.9128	2.2257
0.05	1.4227	1.4746	1.5738	1.7803	1.973	2.3573
0.1	1.4204	1.4759	1.5847	1.8364	—	—

visco-elaticity, inhomogeneity and anisotropy should be considered. Another issue is about the structure of the tissue model. In this study, tissue model with a single layer was developed. However, in most cases, soft tissues are multi-layers and behave complex. Our high frequency ultrasound can differentiate tissue layers with high resolution. Hence, for a further study, soft tissues with multi-layers should be considered in the model, so that, both the depth-dependent and site-dependent Young's modulus distribution can be imaged with our system. What's more, experiments about indentation with the water jet indentation system will be conducted, and results will be compared with those obtained from this FE study.

Acknowledgments

References

[1] Y. C. Fung, *Biomechanics: Mechanical Properties of Living Tissues*, Spring, 1993.

[2] T. R. Tilleman, M. M. Tilleman, and M. H. A. Neumann, "The elastic properties of cancerous skin: Poisson's ratio and Young's modulus," *Israel Medical Association Journal*, vol. 6, no. 12, pp. 753–755, 2004.

[3] M. Stolz, R. Gottardi, R. Raiteri et al., "Early detection of aging cartilage and osteoarthritis in mice and patient samples using atomic force microscopy," *Nature Nanotechnology*, vol. 4, no. 3, pp. 186–192, 2009.

[4] O. Rouvière, M. Yin, M. A. Dresner et al., "MR elastography of the liver: preliminary results," *Radiology*, vol. 240, no. 2, pp. 440–448, 2006.

[5] A. S. Khalil, R. C. Chan, A. H. Chau, B. E. Bouma, and M. R. K. Mofrad, "Tissue elasticity estimation with optical coherence elastography: toward mechanical characterization of in vivo soft tissue," *Annals of Biomedical Engineering*, vol. 33, no. 11, pp. 1631–1639, 2005.

[6] J. F. Greenleaf, M. Fatemi, and M. Insana, "Selected methods for imaging elastic properties of biological tissues," *Annual Review of Biomedical Engineering*, vol. 5, pp. 57–78, 2003.

[7] W. C. Hayes, G. Herrmann, L. F. Mockros, and L. M. Keer, "A mathematical analysis for indentation tests of articular cartilage," *Journal of Biomechanics*, vol. 5, no. 5, pp. 541–551, 1972.

[8] A. F. Mak, W. M. Lai, and V. C. Mow, "Biphasic indentation of articular cartilage-I. Theoretical analysis," *Journal of Biomechanics*, vol. 20, no. 7, pp. 703–714, 1987.

[9] Y. P. Zheng and A. F. T. Mak, "An ultrasound indentation system for biomechanical properties assessment of soft tissues in-vivo," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 9, pp. 912–918, 1996.

[10] A. Ebert, B. R. Tittmann, J. Du, and W. Scheuchenzuber, "Technique for rapid in vitro single-cell elastography," *Ultrasound in Medicine and Biology*, vol. 32, no. 11, pp. 1687–1702, 2006.

[11] N. Gavara and R. S. Chadwick, "Noncontact microrheology at acoustic frequencies using frequency-modulated atomic force microscopy," *Nature Methods*, vol. 7, no. 8, pp. 650–654, 2010.

[12] D. C. Hurley, M. Kopycinska-Müller, A. B. Kos, and R. H. Geiss, "Nanoscale elastic-property measurements and mapping using atomic force acoustic microscopy methods," *Measurement Science and Technology*, vol. 16, no. 11, pp. 2167–2172, 2005.

[13] H. Jin and J. L. Lewis, "Determination of Poison's ratio of articular cartilage by indentation using different-sized indenters," *Journal of Biomechanical Engineering*, vol. 22, no. 7, pp. 843–848, 2004.

[14] M. Zhang, Y. P. Zheng, and A. F. T. Mak, "Estimating the effective Young's modulus of soft tissues from indentation tests—nonlinear finite element analysis of effects of friction and large deformation," *Medical Engineering and Physics*, vol. 19, no. 6, pp. 512–517, 1997.

[15] M. H. Lu and Y. P. Zheng, "Indentation test of soft tissues with curved substrates: a finite element study," *Medical and Biological Engineering and Computing*, vol. 42, no. 4, pp. 535–540, 2004.

[16] A. P. C. Choi and Y. P. Zheng, "Estimation of Young's modulus and Poisson's ratio of soft tissue from indentation using two different-sized indentors: finite element analysis of the finite deformation effect," *Medical and Biological Engineering and Computing*, vol. 43, no. 2, pp. 258–264, 2005.

[17] A. S. Khalil, R. C. Chan, A. H. Chau, B. E. Bouma, and M. R. K. Mofrad, "Tissue elasticity estimation with optical coherence elastography: toward mechanical characterization of in vivo soft tissue," *Annals of Biomedical Engineering*, vol. 33, no. 11, pp. 1631–1639, 2005.

[18] Y. P. Huang, Y. P. Zheng, S. Z. Wang, Z. P. Chen, Q. H. Huang, and Y. H. He, "An optical coherence tomography (OCT)-based air jet indentation system for measuring the mechanical properties of soft tissues," *Measurement Science and Technology*, vol. 20, no. 1, Article ID 015805, 11 pages, 2009.

[19] D. M. Ebenstein and L. A. Pruitt, "Nanoindentation of biological materials," *Nano Today*, vol. 1, no. 3, pp. 26–33, 2006.

[20] C. Li, L. A. Pruitt, and K. B. King, "Nanoindentation differentiates tissue-scale functional properties of native articular cartilage," *Journal of Biomedical Materials Research A*, vol. 78, no. 4, pp. 729–738, 2006.

[21] L. Ng, H. H. Hung, A. Sprunt, S. Chubinskaya, C. Ortiz, and A. Grodzinsky, "Nanomechanical properties of individual chondrocytes and their developing growth factor-stimulated pericellular matrix," *Journal of Biomechanics*, vol. 40, no. 5, pp. 1011–1023, 2007.

[22] L. P. Han, J. A. Noble, and M. Burcher, "A novel ultrasound indentation system for measuring biomechanical properties of in vivo soft tissue," *Ultrasound in Medicine and Biology*, vol. 29, no. 6, pp. 813–823, 2003.

[23] K. Mori, Y. Nakagawa, H. Kuroki et al., "Non-contact evaluation for articular cartilage using ultrasound," *JSME International Journal A*, vol. 49, no. 2, pp. 242–249, 2006.

[24] P. Kiviranta, E. Lammentausta, J. Töyräs, I. Kiviranta, and J. S. Jurvelin, "Indentation diagnostics of cartilage degeneration," *Osteoarthritis and Cartilage*, vol. 16, no. 7, pp. 796–804, 2008.

[25] K. Hattori, K. Ikeuchi, Y. Morita, and Y. Takakura, "Quantitative ultrasonic assessment for detecting microscopic cartilage damage in osteoarthritis," *Arthritis Research & Therapy*, vol. 7, no. 1, pp. 38–46, 2005.

[26] Y. P. Zheng, A. P. C. Choi, H. Y. Ling, and Y. P. Huang, "Simultaneous estimation of Poisson's ratio and Young's modulus using a single indentation: a finite element study," *Measurement Science and Technology*, vol. 20, no. 4, Article ID 045706, 9 pages, 2009.

[27] M. H. Lu and Y. P. Zheng, "A novel method to obtain modulus image of soft tissues using ultrasound water jet indentation: a phantom study," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 1, pp. 114–121, 2007.

[28] M. H. Lu, Y. P. Zheng, Q. H. Huang et al., "Noncontact evaluation of articular cartilage degeneration using a novel ultrasound water jet indentation system," *Annals of Biomedical Engineering*, vol. 37, no. 1, pp. 164–175, 2009.

[29] M. H. Lu, Y. P. Zheng, H. B. Lu, Q. H. Huang, and L. Qin, "Evaluation of bone-tendon junction healing using water jet ultrasound indentation method," *Ultrasound in Medicine and Biology*, vol. 35, no. 11, pp. 1783–1793, 2009.

[30] F. J. Carter, T. G. Frank, P. J. Davies, D. McLean, and A. Cuschieri, "Measurements and modelling of the compliance of human and porcine organs," *Medical Image Analysis*, vol. 5, no. 4, pp. 231–236, 2001.

[31] A. Athanasiou, A. Tardivon, M. Tanter et al., "Breast lesions: quantitative elastography with supersonic shear imaging—preliminary results," *Radiology*, vol. 256, no. 1, pp. 297–303, 2010.

[32] H. Abe, K. Hayashi, and M. Sato, *Data Book on Mechanical Properties of Living Cells, Tissue and Organs*, Spring, Tokyo, Japan, 1996.

[33] S. Birks and R. E. Green, *Nondestructive Testing Handbook: Ultrasonic Testing*, American Society for Nondestructive Testing, Columbus, Ohio, USA, 1991.

[34] M. Lasser, B. Lasser, J. Kula, and G. Rohrer, "On-line, large area ultrasonic imaging for composite manufacturing," in *Proceedings of the The American Society for Nondestructive Testing*, Phoenix, Arizona, October 1999.

[35] Y. Liu, A. E. Kerdok, and R. D. Howe, "A nonlinear finite element model of soft tissue indentation," *Lecture Notes in Computer Science*, vol. 3078, pp. 67–76, 2004.

[36] J. E. Bischoff, "Static indentation of anisotropic biomaterials using axially asymmetric indenters—a computational study," *Journal of Biomechanical Engineering*, vol. 126, no. 4, pp. 498–505, 2004.

Research Article

Sparse Reconstruction for Bioluminescence Tomography Based on the Semigreedy Method

Wei Guo,[1] Kebin Jia,[1] Qian Zhang,[2] Xueyan Liu,[3] Jinchao Feng,[1] Chenghu Qin,[2] Xibo Ma,[2] Xin Yang,[2] and Jie Tian[2]

[1] *The College of Electronic Information & Control Engineering, Beijing University of Technology, Beijing 100124, China*
[2] *Intelligent Medical Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[3] *Sino-Dutch Biomedical and Information Engineering School of Northeastern University, Liaoning, Shenyang 110004, China*

Correspondence should be addressed to Kebin Jia, kebinj@bjut.edu.cn and Jie Tian, tian@ieee.org

Bioluminescence tomography (BLT) is a molecular imaging modality which can three-dimensionally resolve the molecular processes in small animals *in vivo*. The ill-posedness nature of BLT problem makes its reconstruction bears nonunique solution and is sensitive to noise. In this paper, we proposed a sparse BLT reconstruction algorithm based on semigreedy method. To reduce the ill-posedness and computational cost, the optimal permissible source region was automatically chosen by using an iterative search tree. The proposed method obtained fast and stable source reconstruction from the whole body and imposed constraint without using a regularization penalty term. Numerical simulations on a mouse atlas, and *in vivo* mouse experiments were conducted to validate the effectiveness and potential of the method.

1. Introduction

Due to its ability of monitoring physiological and pathological activities at the molecular level, small animal optical molecular imaging has become an important method for biomedical research. Bioluminescence imaging, as one of optical molecular imaging modalities, has attracted attention for its advantages in sensitivity, specificity, and cost effectiveness in cancer research and drug development [1–3]. Compared with planar bioluminescent imaging, BLT can three-dimensionally reconstruct the bioluminescent probes in small animals [4].

The generic BLT model is ill-posed. However, it has been theoretically proven that the solution uniqueness can be established under practical constraints using a *priori* knowledge [5]. In most existing reconstructions, multi spectral measurement [6–10], permissible source region (PSR) [9–12], and sparse reconstruction [13–15] are three common strategies to reduce ill posedness of BLT. Although multi spectral techniques improve reconstruction

qualities to a certain degree by increasing the measurable data, they in turn impose some limitations in practical applications such as increased signal acquisition time and a high computational cost [16]. Besides, the PSR strategy can significantly improve the location accuracy of reconstructed source and reduce the computational cost by limiting the reconstruction region into a small area. However, in practical applications, both the size and position of the permissible region have significant impact on imaging results [5]. Additionally, since the bioluminescent source distribution is usually sparse in practical applications and only insufficient boundary measurements are available, the compressed sensing can bring benefits in spatial resolution and algorithm stability to BLT reconstruction. Recently, many sparse reconstruction methods have emerged in BLT [17–19]. The majority of them reformulate the BLT inverse problem into minimizing an objective functions that integrate a sparse regularization term with a quadratic error term and solve it via computationally tractable convex optimization methods, such as linear programming and

Table 1: Optical parameters for each organ in the mouse atlas [34].

	Muscle	Heart	Lungs	Liver	Spleen	Bone
μ_a (mm^{-1})	0.032	0.022	0.071	0.128	0.075	0.002
μ'_s (mm^{-1})	0.586	1.129	2.305	0.646	2.178	0.935

gradient methods, However, the efficacy of the sparse regularization methods strongly depends on the choice of regularization parameter in practical applications [19].

The existing works have demonstrated that PSR can improve reconstruction qualities by reducing the number of unknown variables. Although the meaningful results can be obtained by using the PSR that is based on the bioluminescent signals and *a priori* knowledge available from a specific biomedical application [4, 11, 12], in most cases, it is rather difficultly to manually select such a small and appropriate region. Recently, some optimal permissible source region methods have emerged in BLT. Feng et al. presented a reconstruction algorithm for a spectrally resolved BLT based on an adaptive rough estimate of an optimal permissible source region and multilevel finite element method approach (FEM), where Tikhonov regularization was used to solve the constrained BLT inverse problem [9]. Naser and Patterson proposed a two-step reconstruction algorithm of bioluminescence, in which the permissible regions were shrunk by using an iterative minimization solution based on the L1 norm [10]. These works demonstrated the feasibility and potential of the optimal PSR techniques with numerical simulation. However, both of two previous reconstructions also needed the regularization methods to reconstruct the sources in the allowed region, which made the reconstruction results also depend on the choice of regularization parameter. Furthermore, they were demonstrated with only regular phantoms simulations and presented no *in vivo* experiment validation.

In this work, a novel BLT reconstruction algorithm based on the semi greedy method was proposed. The optimal PSR problem was cast into a search for the correct support of source distribution among a number of dynamically evolving candidate subsets, and the optimal PSR was chosen automatically by using an iterative search tree. Therefore, the columns of the system matrix were treated as the nodes for building up the search tree where each path from the root to a leaf node denoted a candidate. The search tree was initialized with some unspecific nodes. At each iteration, new nodes were appended to the most promising path, which were selected to minimize the cost function based on the residue. The permissible source region was expended by adding nodes with high a probability to contribute to the source. Among the system matrix, the columns that were corresponding to the nodes contained in the most promising path were selected to obtain the source distribution. By automatically choosing an optimal PSR, the method reduced the ill posedness of the problem and imposed constraint without using a regularization penalty term.

This paper is organized as follows. In Section 2, the forward photon propagation model, the inverse problem for BLT with FEM, and the proposed algorithm are introduced.

In Section 3, the numerical simulations in a mouse atlas demonstrate the performance of the proposed method. In Section 4, an *in vivo* mouse experiment is conducted to further evaluate its reliability. Finally, we discuss the results and conclude this paper.

2. Method

Light propagation in biological media is essentially the transport of radiant energy. The radiative transfer equation (RTE) can rigorously describe light transport in turbid media [20]. Compared with the other approximations of RTE such as simple spherical harmonics, spherical harmonics and discrete ordinates, the following steady-state diffusion approximate equation (DA) in (1) is the most popular one as a result of its moderate computational efficiency and explicit physical meaning [4, 5, 11, 15, 21]. Assuming that the bioluminescence imaging experiment is performed in a totally dark environment and no photon travels into Ω through the boundary $\partial\Omega$, the equation is subject to the Robin boundary condition in (2) as follows:

$$-\nabla \cdot D(r)\nabla\Phi(r) + \mu_a(r)\Phi(r) = S(r) \quad (r \in \Omega), \quad (1)$$

$$\Phi(r) + 2A(r; n, n')D(r)(\nu(r) \cdot \nabla\Phi(r)) = 0 \quad (r \in \partial\Omega), \quad (2)$$

where Ω is the domain of the problem, $S(r)$ donates the source energy distribution, $\Phi(r)$ represents the photo fluence rate, $\mu_a(r)$ is the absorption coefficient, $\mu'_s(r)$ is the reduced scattering coefficient, $D(r) = 1/3(\mu_a(r) + \mu'_s(r))$ indicates the optical diffusion coefficient, $\partial\Omega$ donates the boundary of the problem, and $A(r, n, n')$ represents the mismatch coefficient between Ω and its surrounding medium. The measured quantity on the boundary $\partial\Omega$ is given by the outgoing radiation as follows:

$$Q(r) = -D(r)(\nu(r) \cdot \nabla\Phi(r)) = \frac{\Phi(r)}{(2A(r; n, n))} \quad (r \in \partial\Omega). \quad (3)$$

FEM is a powerful tool for solving the DA equation [4, 6–10]. By using FEM to discretely approach the solving domain and making a series of transformations and rearrangements, the linear relationship links the source distribution inside the heterogeneous medium, and the photon fluence rate on the surface is established as follows:

$$M\Phi = FX, \quad (4)$$

where X is the source distribution of the interior nodes, Φ is the measurable photon flux photon on the boundary nodes, M is the positive definite matrix, and F is the source weight matrix. The nonmeasurable entries in Φ and corresponding rows in $M^{-1}F$ can be removed. Then a new linear relationship can be obtained as follows [4, 22]:

$$AX = \Phi^m. \quad (5)$$

For BLT, the domains of the bioluminescent sources are usually very small and sparse compared with the entire
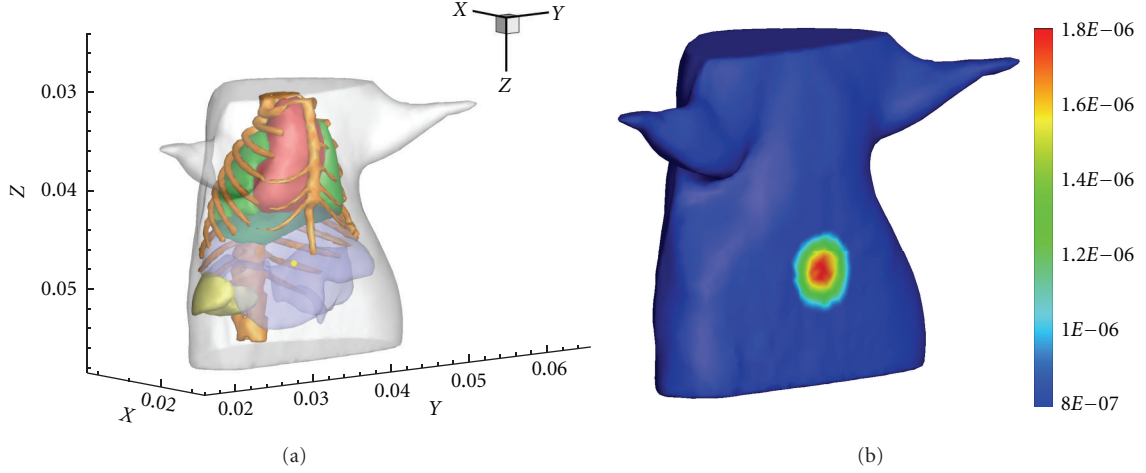
FIGURE 1: Reconstruction model with a single source. (a) The torso of the mouse atlas model with one source in the liver. (b) The simulated photon distribution on the surface.

reconstruction domain. That means that there are only fewer nonzero components in X. Therefore, the system matrix A can be seen as the dictionary, and Φ^m has a less-term representation over the dictionary. As a consequence, the aim of the proposed method is imposing the constraint on the source space by choosing only the part that contributes to the source distribution. In the language of sparse approximation, greed pursuit algorithms are useful methods for solving this problem [23, 24]. For instance, Orthogonal Matching Pursuit (OMP) is to pick columns in a greedy fashion [25]. According to the introduction of OMP, the reconstruction starts with an empty index set. At each iteration, we choose the single column of A that is most strongly correlated with the remaining part of Φ^m. Then, we subtract its contribution to Φ^m and iterate on the residual. The reconstruction is stopped after a number of iterations. Unfortunately, from experimental results shown in Figure 2 and Table 2, we found that OMP, as the single-path algorithm, could not achieve the desired expectations for reconstructing the bioluminescent source. In the experiments, when the computation of a single path selects a wrong column, the correct one is still in the set of candidate representations. Therefore, incorporation of a multipath search strategy is motivated to improve reconstruction. In this section, the semigreedy method was used to search for the correct support of Φ^m among a number of dynamically evolving candidate subsets.

A general best first (GBF) is a search algorithm which constructs a tree T by expanding the most promising node chosen according to a specified rule. Search algorithm A^*, as one of the most studied versions of GBF, can find path in combinatorial search. It selects an optimal path by minimizing an additive evaluation function $f(n) = g(n) + h(n)$, where $g(n)$ is the cost of the currently evaluated path from start node s to n, and $h(n)$ is a heuristic estimate of the cost of the path remaining between n and some goal nodes [26–29]. In our problem, the A^* search tree was iteratively built up by nodes which represent the dictionary atoms. Each path from the root to a leaf node denoted a subset of dictionary atoms which was a candidate support

for Φ^m. Let us define the notation. S_i and C_i denote the atoms contained in path i and the vector of corresponding coefficients obtained after orthogonal projection of the residue onto the set of selected columns [25]. Similarly, s_i^l and c_i^l represent the selected atom at the lth node on path S_i and corresponding coefficient.

The search tree starts with less unspecified nodes. A simple way is selecting the $I = N|500$ nodes that have the highest absolute inner product with Φ^m. In order to find the fewest possible nodes, the search must constantly make an evaluation to decide which available paths should be expended next. Therefore, the evaluation function $g(S^l)$ is defined as follows:

$$g\left(S^l\right) = \left\|r^l\right\|_2 = \left\|\Phi^m - \sum_{j=1}^{l} c^j s^j\right\|_2. \tag{6}$$

Beside the evaluation function, the auxiliary function is also needed to assess the cost brought by adding a preferred goal node to the path. Generally, according to the expectation on average equal contribution of unopened nodes, the auxiliary function can be built as follows:

$$d\left(S_i^l\right) \geq \left(\left\|r_i^{l-1}\right\|_2 - \left\|r_i^l\right\|_2\right)t, \tag{7}$$

where coefficient t is defined as $t = \alpha N - l$. α is a ratio between the number of the nonzero entries and the zero entries in the solution X. It is well known that the sparse solution has only less nonzero entries. Therefore, in most practical application, α is very smaller than 1. Here, we selected $\alpha = 0.005$. If the source distribution could be seen as the K-sparse signal, K could be computed by $K = \lceil \alpha N \rceil$. The cost function can be written as follows:

$$f\left(S_i^l\right) = \left\|r_i^l\right\| - \beta\left(\left\|r_i^{l-1}\right\| - \left\|r_i^l\right\|\right)t, \tag{8}$$

where β is a constant. A lot of experiments for different reconstruction models including 2D and 3D experiments were performed to evaluate the impact of β on the
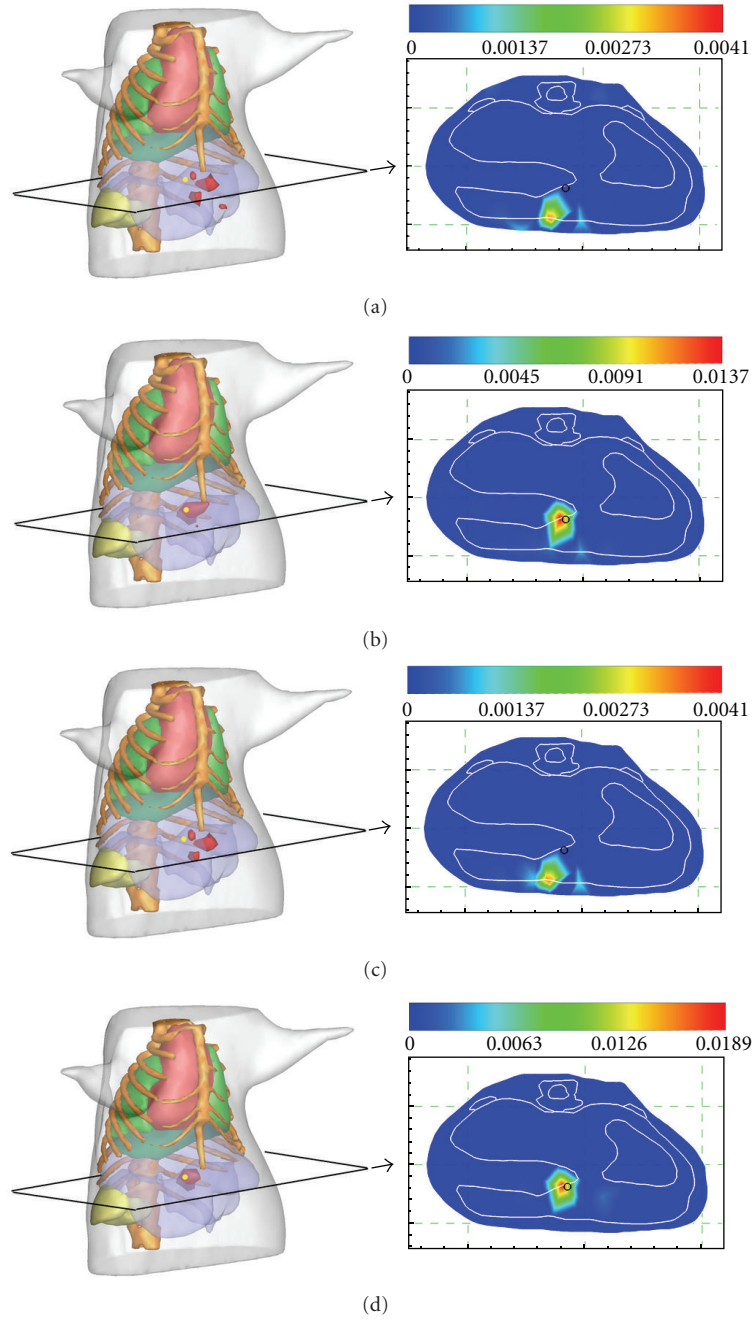
(a)

(b)

(c)

(d)

FIGURE 2: Comparison of reconstruction results. (a), (b), (c) and (d) are the reconstruction results with OMP, FIST-L1 (with regularization parameter set to be 4e-11), FIST-L1 (with regularization parameter set to be 4e-10), and the proposed method, respectively. The results are shown in the form of isosurfaces for 40% of the maximum value (left column) and slice images in $z = 47.29$ mm plane (right column). The small yellow sphere in the iso-surfaces view image and black circles in the slice images denote the real position of the bioluminescent source.

TABLE 2: Quantitative comparisons of the reconstruction results.

Method	Recon. location center (mm)	Location error (mm)	Recon. time (s)	Maximum recon. value (nW/mm³)
OMP	(15.61, 32.92, 47.48)	2.9605	0.798	0.00412
FIST-L1(1)	(18.26, 31.97, 47.28)	0.3995	16.94	0.01371
FIST-L1(2)	(15.61, 32.92, 47.48)	2.9605	25.44	0.00414
The proposed method	(18.26, 31.97, 47.28)	0.3995	12.20	0.01889

source reconstruction. We found that better results could be obtained when it varied in the interval [1.00, 1.25]. Therefore, in our experiments, its value was taken from the interval and set to be1.05.

In practice, if all of the children of the most promising partial path are added to the search tree at each iteration, the tree might have too many search paths. Therefore, the following pruning strategies are employed a guide on how the tree grows.

The first one is about extensions per path. At each step, it is not necessary to have all of the unopened atoms added to the current optimal path. We can expand the search tree only by the B children which have the highest absolute inner-product with the residue to the selected path. This pruning strategy decreases the upper bound to B^K on the number of paths. Practically, I and B are selected to be much smaller than N, which can drastically decrease the paths involved in the search. Although the number of extensions per path is limited to B, it is also necessary to control the size of path. That is because that adding new paths at each iteration continues increasing required memory. Therefore, to reduce memory requirements, we adopted the "beam search" strategy [30] and limit the maximum number of paths in the tree by the beam width P. When this limit is exceeded, the paths with maximum cost are seen as the worst paths and are removed from the tree until P paths remain. Here, B and P were set to be 4 and 200, respectively. Moreover, since order of nodes along a path is unimportant, amalgamating the equivalent path is also important to improve the search performance. For this purpose, we define a path equivalency notion; $S_1^{l_1}$ and $S_2^{l_2}$ are two paths with different length $l_1 < l_2$. If all atoms of $S_1^{l_1}$ can be found in $S_2^{l_2}$ and these composed the continuous subset of $S_2^{l_2}$, we define the above two paths as being equivalent. Consequently, the insertion of $S_2^{l_2}$ into the tree is unnecessary.

After the growing of the search tree is finished, the linear relationship between the observation Φ^m and the selected PSR can be established as follows:

$$A_{\text{opt}}X = \Phi^m. \tag{9}$$

Since the nodes contained in the optimal path can be much smaller than the number of all nodes N, (9) is an overdetermined linear equation. Therefore, a limited memory variation of the Broyden Fletcher Goldfarb Shanno (LBFGS) [31] was used to directly solve (9).

3. Simulation Studies in the Mouse Atlas

In this subsection, heterogeneous simulations were presented to demonstrate the performance of the proposed method for mouse applications. The experimental data were acquired by a dual-modality BLT/micro-CT system developed in our lab [32, 33]. By using image processing and interactive segmentation technology, heterogeneous model including heart, lungs, liver, bone, spleen, and muscle was built. The optical coefficients for each organ are listed in Table 1 [34]. Here, the torso section with a height of 25 mm was selected as the reconstruction region.

3.1. Reconstruction in a Single Source Case. In the experiment, a spherical source with a 0.6 mm diameter was placed in the liver with the center at (18.24 mm, 31.58 mm, 47.29 mm) as shown in Figure 1(a). The source was modeled as isotropic point sources whose strength was set to be 2 nW/mm^3. As for the forward problem, the FEM was used to obtain the synthetic measurements on the boundary. The atlas model was discretized into a tetrahedral-element mesh with 30892 nodes and 167841 elements. The generated simulated photon distribution on the boundary is presented in Figure 1(b). Then the forward solutions were projected onto a single coarse mesh consisting of 20068 elements and 3098 nodes, which was used for reconstructing the source.

To better illustrate the performance of the proposed method, we compared the proposed method with OMP [25] and FIST-L1 [35, 36]. The former is a typical greedy pursuit method for sparse signal recovery. The latter, as a sparse regularization method, can be viewed as a standard approach to ill-posed linear inverse problems and has been used in fluorescence molecular tomography (FMT) and BLT. Here, the step size in FIST-L1 was computed by using the estimation algorithm introduced in [35]. Since the regularization parameter plays an important role in the regularization methods, we performed two experiments with different regularization parameters that were set to be 4e-11 and 4e-10, respectively. All of the reconstructions were carried out on a personal computer with 3.2 GHz Intel Core2 duo CPU and 2 GB RAM.

The qualities of the reconstruction were quantitatively assessed in terms of location error and the maximum reconstructed intensity. The location error was defined as Euclidean distance between S_{real} and S_{recons}, where S_{real} and S_{recons} were the real locations of the source center and the location of the node with the maximum reconstructed value, respectively. The visual effects of the reconstruction results are presented in the form of slice images and iso-surfaces, as shown in Figure 2. Additionally, the detailed information about parameters and the final quantitative reconstruction results are summarized in Table 2.

We found that the reconstructed positions by L1 regularization with an optimal regularization parameter and the proposed method were identical. Specifically, the reconstructed center was (18.26 mm, 31.97 mm, 47.28 mm) with a location error of 0.3995 mm from the actual source, whereas the location error by OMP was 2.9605 mm. The performance of OMP was inferior to the other two methods. L1 regularization performed well and obtained satisfactory source localizations and maximum reconstruction value. However, the selection of the regularization parameter had a great impact on the reconstruction results. As for the proposed method, it performed slightly better than L1 regularization with manually optimized regularization parameter in terms of maximum reconstruction value. Moreover, it was also an efficient reconstruction method.

The above experiments were performed without noise. In order to evaluate the sensitivity of the proposed method to various noise levels, six cases were carried out where the measurements were added to 5%, 10%, 15%, 20%, 25% and 30% Gaussian noise, respectively. We also made
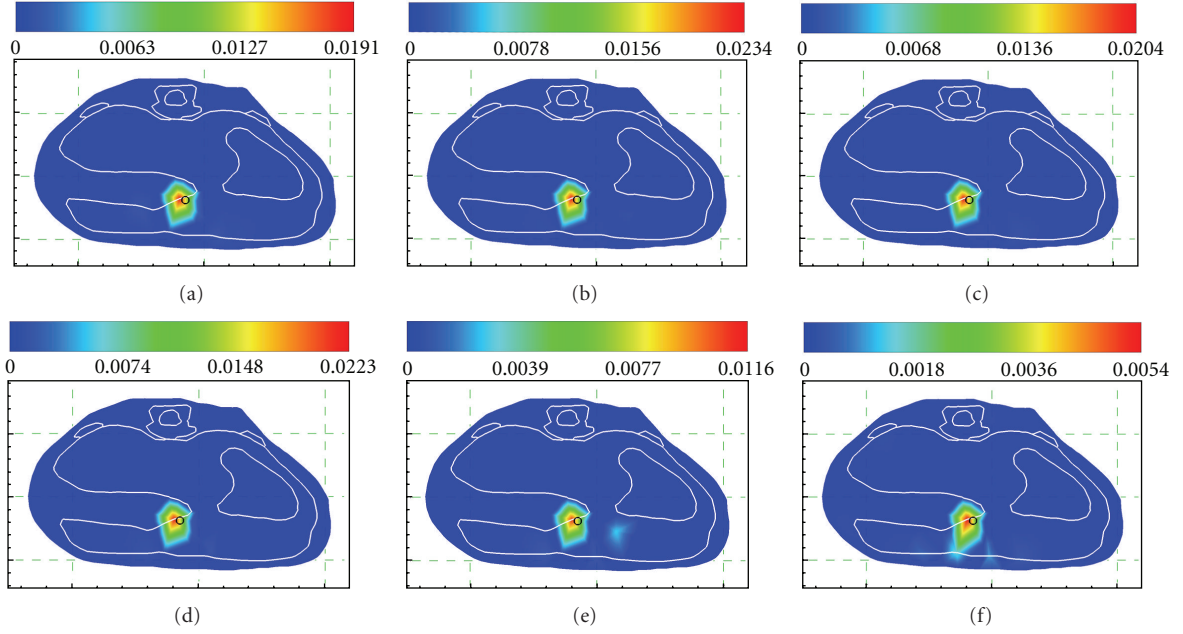
FIGURE 3: The proposed method for different noise levels of 0%, 5%, 10%, 15%, 20%, 25%, and 30%.

TABLE 3: Quantitative results in a single source case with different noise levels.

| Method	Noise level	Recon. location center (mm)	Location error (mm)	Recon. Time (s)	Maximum recon. value (nW/mm³)
FIST-L1	5%	(18.26, 31.97, 47.28)	0.3995	24.43	0.0122
	10%	(18.26, 31.97, 47.28)	0.3995	24.99	0.0085
	15%	(18.26, 31.97, 47.28)	0.3995	24.77	0.0050
	20%	(15.61, 32.92, 47.48)	2.9605	24.82	0.0032
	25%	(15.61, 32.92, 47.48)	2.9605	24.94	0.0025
	30%	(15.61, 32.92, 47.48)	2.9605	24.60	0.0023
The proposed method	5%	(18.26, 31.97, 47.28)	0.3995	13.41	0.0192
	10%	(18.26, 31.97, 47.28)	0.3995	14.15	0.0234
	15%	(18.26, 31.97, 47.28)	0.3995	16.37	0.0204
	20%	(18.26, 31.97, 47.28)	0.3995	18.09	0.0223
	25%	(18.26, 31.97, 47.28)	0.3995	18.95	0.0116
	30%	(18.26, 31.97, 47.28)	0.3995	46.61	0.0054
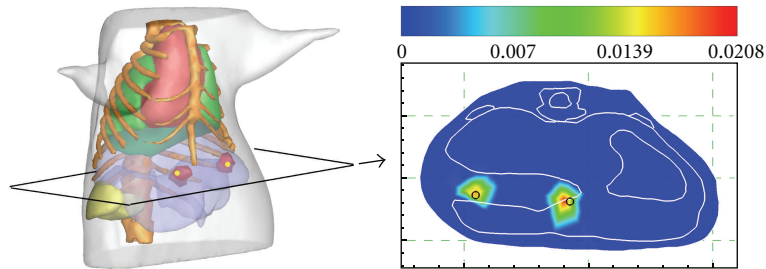
FIGURE 4: Reconstruction results in double source case. The results are shown in the form of iso-surface for 40% of the maximum value (left column) and slice image in $z = 47.29$ mm plane (right column). The small yellow sphere in the iso-surface view image and black circles in the slice image denote the real position of the bioluminescent source.

TABLE 4: Reconstruction results in double source case.

Source number	Actual position (mm)	Recon. location center (mm)	Location error (mm)	Maximum recon. value (nW/mm^3)
1	(18.24, 31.58, 47.29)	(18.26, 31.97, 47.28)	0.3995	0.0204
2	(18.74, 39.15, 47.06)	(18.75, 39.20, 46.86)	0.2064	0.0208
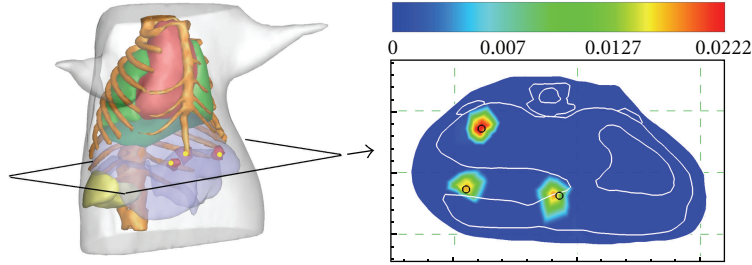
FIGURE 5: Reconstruction results in multisource case. The results are shown in the form of iso-surface for 40% of the maximum value (left column) and slice image in $z = 47.29$ mm plane (right column). The small yellow sphere in the iso-surface view image and black circles in the slice image denote the real position of the bioluminescent source.

a comparison between the proposed method and FIST-L1 with the regularization parameter that was set to be 4e-11. The reconstruction results under different noise levels are compiled in Table 3 and Figure 3 showing that the proposed method was robust to measurement noise.

3.2. Double Source Case and MultiSource Case. Dual source setting was also considered in order to evaluate the proposed method. Two sources have the same size as one used in the single source case. Their strength and position were set to be 2 nW/mm^3 and (18.24 mm, 31.58 mm, 47.29 mm), 1 nW/mm^3 and (18.74 mm, 39.15 mm, 47.06 mm), respectively. The reconstruction results are shown in Figure 4 and Table 4.

Multiple sources setting simulation experiment was also presented to further demonstrate the ability of the proposed method. Based on the setting in double source case, the third source with the same size and shape was added. Its strength and position were set to be 1.5 nW/mm^3 and (23.60 mm, 37.94 mm, 47.45 mm). The final reconstruction results are presented in Figure 5 and Table 5. The result of two group experiments indicated that the sources can be accurately distinguished by using the proposed method.

4. *In Vivo* Experiment Validation

Besides the numerical simulations with mouse atlas, an *in vivo* experiment was carried out on a mouse to further test the proposed method. The experiment was also performed with a dual-modality BLT/micro-CT system developed in our lab [32, 33, 37]. A nude hairless mouse (Nu/Nu, Laboratory Animal Center, Peking University, China) was used in this experiment. To simulate a known bioluminescent source, a home-made cylindrical light source about 3 mm long and 1.5 mm in diameter was implanted into the abdomen of the mouse in this experiment. The source was made of a catheter filled with luminescent liquid and emitted a red luminescent

light that had a similar emission spectrum with a firefly luciferase-based source.

Before the beginning of the experiment, the CCD (VersArray, Princeton Instruments, Trenton, NJ, USA) was cooled to 110° by using liquid nitrogen. The mouse was anesthetized and placed in a mouse holder. The mouse holder was set to rotate to 0°, 90°, 180° and 270°. At each of four positions, the mouse was photographed by the CCD camera. After the optical data were acquired, the mouse was scanned by using the micro-CT to obtain the anatomical maps which could provide structural information for the source reconstruction. Then the CT data were segmented into five regions represent muscle, lungs, heart, liver, and kidneys, respectively, as shown in Figure 6(a). The heterogeneous model including five tissues was discretized into the mesh containing 11917 tetrahedral elements and 2557 nodes. The optical parameters for different tissues were calculated based on the literature as listed in Table 6 [34, 38]. The optical data was registered with the volumetric mesh, and measured data were mapped onto the surface of the mesh. The result of mapped photon distribution is shown in Figure 6(b).

It took about 8 seconds to complete the reconstruction using the proposed method. The final results are presented in Figure 7, where the reconstruction source center is (37.17 mm, 38.82 mm, and 20.92 mm) with a deviation about 2 mm to the actual center. As can be seen in the reconstruction results, the proposed methods could obtain satisfactory bioluminescent source localizations.

5. Discussion and Conclusion

In this paper, we have proposed a new method based on the semigreedy for bioluminescence tomography. The reconstruction results of the simulations on a mouse atlas demonstrate that the proposed reconstruction method is able to accurately and stably localize bioluminescent source from whole body, even with noisy measurements. The *in vivo* experiment further shows its performance.

TABLE 5: Reconstruction results in multiple source case.

Source number	Actual position (mm)	Recon. location center (mm)	Location error (mm)	Maximum recon. value (nW/mm³)
1	(18.24, 31.58, 47.29)	(18.26, 31.97, 47.28)	0.3995	0.0195
2	(18.74, 39.15, 47.06)	(18.75, 39.20, 46.86)	0.2064	0.0226
3	(23.60, 37.94, 47.45)	(23.60, 37.97, 47.65)	0.2023	0.0300

(a)   (b)
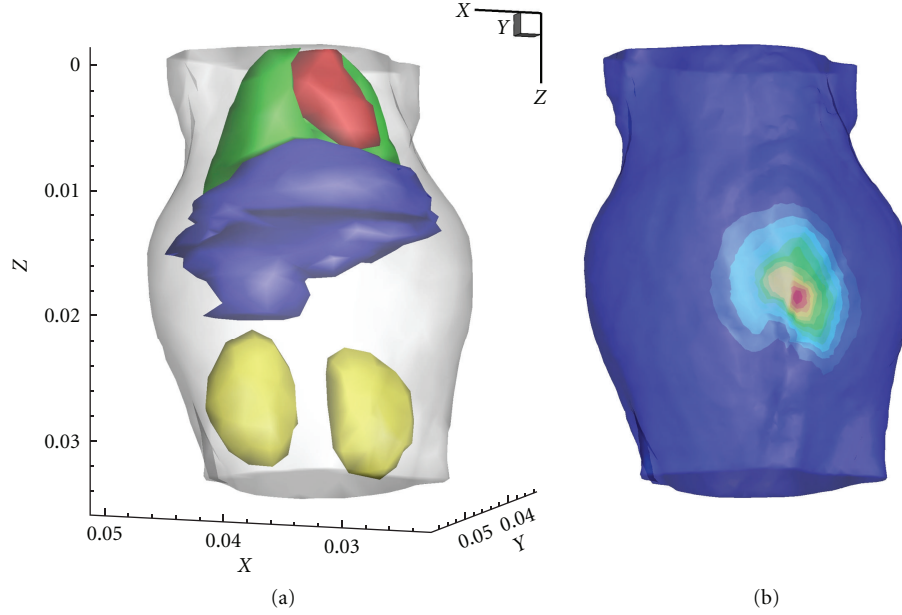
FIGURE 6: *In vivo* heterogeneous model. (a) The torso of the model. (b) The mapped photon distribution on the mouse surface.

TABLE 6: Optical parameters for each organ in the heterogeneous model [34, 38].

	Muscle	Heart	Lungs	Liver	Kidneys
μ_a (mm^{-1})	0.008	0.138	0.456	0.829	0.150
μ'_s (mm^{-1})	1.258	1.076	2.265	0.735	2.507

The PSR strategy can significantly improve reconstruction qualities. However, in most cases, empirically selecting such small and appropriate region is unconvenient, even available. In this study, The optimal PSR problem is cast into a search for the correct support of source distribution among a number of dynamically evolving candidate subsets. In view of the characteristics in BLT sparse distribution, only the columns that contribute to the source reconstruction are chosen automatically by using semi-greedy method. The constraint imposed on the source space reduces the ill posedness of the problem and computational cost.

It is noted that *in vivo* experiment is not as accurate as simulations. Some reasons can be explained for this phenomenon. First of all, the error was generated, when the energy distribution was mapped from 2D images to a 3D mouse surface. Secondly, there were only five main segmented tissues used to build a heterogeneous model while others simply were regarded as the muscle, which also led to errors. Finally, the accuracy of the photon propagation model was very important for source reconstruction. The diffusion

approximation was used due to its moderate computational efficiency and explicit physical meaning. However, it has some limitations in certain regions, such as void or more absorptive regions. Therefore, the error brought on by the DA model is inevitable. As discussed above, our future work will focus on studying more accurate forward models to describe photon propagation in biological tissues and improving the experimental procedures and imaging system to further promote the reconstruction quality.
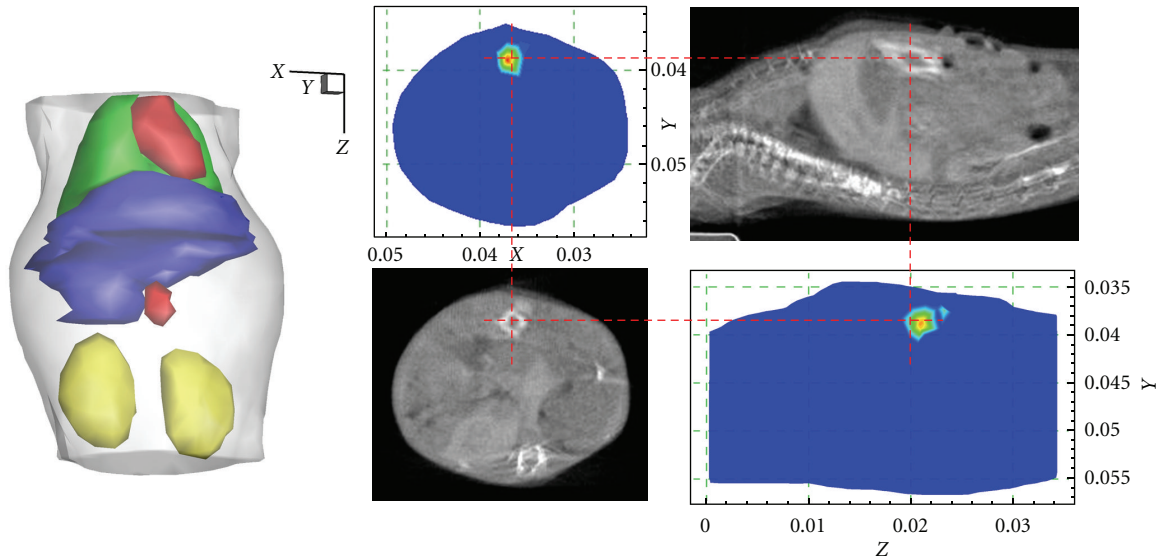
Acknowledgments

FIGURE 7: The reconstruction results. The results are shown in the form of isosurface for 40% of maximum value (left column) and slice images (right column). The transverse view of the results and the comparisons with the corresponding CT slices. The cross of the red lines denotes the actual source center.

References

[1] V. Ntziachristos, J. Ripoll, L. V. Wang, and R. Weissleder, "Looking and listening to light: the evolution of whole-body photonic imaging," *Nature Biotechnology*, vol. 23, no. 3, pp. 313–320, 2005.

[2] J. K. Willmann, N. van Bruggen, L. M. Dinkelborg, and S. S. Gambhir, "Molecular imaging in drug development," *Nature Reviews Drug Discovery*, vol. 7, no. 7, pp. 591–607, 2008.

[3] A. M. Loening, A. M. Wu, and S. S. Gambhir, "Red-shifted Renilla reniformis luciferase variants for imaging in living subjects," *Nature Methods*, vol. 4, no. 8, pp. 641–643, 2007.

[4] G. Wang, W. Cong, K. Durairaj et al., "In vivo mouse studies with bioluminescence tomography," *Optics Express*, vol. 14, no. 17, pp. 7801–7809, 2006.

[5] G. Wang, Y. Li, and M. Jiang, "Uniqueness theorems in bioluminescence tomography," *Medical Physics*, vol. 31, no. 8, pp. 2289–2299, 2004.

[6] H. Dehghani, S. C. Davis, and B. W. Pogue, "Spectrally resolved bioluminescence tomography using the reciprocity approach," *Medical Physics*, vol. 35, no. 11, pp. 4863–4871, 2008.

[7] C. Kuo, O. Coquoz, T. L. Troy, H. Xu, and B. W. Rice, "Three-dimensional reconstruction of in vivo bioluminescent sources based on multispectral imaging," *Journal of Biomedical Optics*, vol. 12, no. 2, Article ID 024007, 2007.

[8] H. Dehghani, S. C. Davis, S. Jiang, B. W. Pogue, K. D. Paulsen, and M. S. Patterson, "Spectrally resolved bioluminescence optical tomography," *Optics Letters*, vol. 31, no. 3, pp. 365–367, 2006.

[9] J. Feng, K. Jia, G. Yan et al., "An optimal permissible source region strategy for multispectral bioluminescence tomography," *Optics Express*, vol. 16, no. 20, pp. 15640–15654, 2008.

[10] M. A. Naser and M. S. Patterson, "Algorithms for bioluminescence tomography incorporating anatomical information and reconstruction of tissue optical properties," *Biomedical Optics Express*, vol. 1, no. 2, pp. 512–526, 2010.

[11] C. Qin, X. Yang, J. Feng et al., "Adaptive improved element free Galerkin method for quasi-or multi-spectral bioluminescence tomography," *Optics Express*, vol. 17, no. 24, pp. 21925–21934, 2009.

[12] Y. Lv, J. Tian, W. Cong et al., "A multilevel adaptive finite element algorithm for bioluminescence tomography," *Optics Express*, vol. 14, no. 18, pp. 8211–8223, 2006.

[13] Y. Lu, H. B. Machado, A. Douraghy, D. Stout, H. Herschman, and A. F. Chatziioannou, "Experimental bioluminescence tomography with fully parallel radiative-transfer-based reconstruction framework," *Optics Express*, vol. 17, no. 19, pp. 16681–16695, 2009.

[14] H. Gao and H. Zhao, "Multilevel bioluminescence tomography based on radiative transfer equation part 1: 11 regularization," *Optics Express*, vol. 18, no. 3, pp. 1854–1871, 2010.

[15] X. He, J. Liang, X. Wang et al., "Sparse reconstruction for quantitative bioluminescence tomography based on the incomplete variables truncated conjugate gradient method," *Optics Express*, vol. 18, no. 24, pp. 24825–24841, 2010.

[16] Y. Lu, H. B. Machado, A. Douraghy, D. Stout, H. Herschman, and A. F. Chatziioannou, "Experimental bioluminescence tomography with fully parallel radiative-transfer-based reconstruction framework," *Optics Express*, vol. 17, no. 19, pp. 16681–16695, 2009.

[17] A. Cong, W. Cong, Y. Lu, P. Santago, A. Chatziioannou, and G. Wang, "Differential evolution approach for regularized bioluminescence tomography," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 9, pp. 2229–2238, 2010.

[18] M. Freiberger, C. Clason, and H. Scharfetter, "Total variation regularization for nonlinear fluorescence tomography with an augmented Lagrangian splitting approach," *Applied Optics*, vol. 49, no. 19, pp. 3741–3747, 2010.

[19] J. Feng, C. Qin, K. Jia et al., "An adaptive regularization parameter choice strategy for multispectral bioluminescence

tomography," *Medical Physics*, vol. 38, no. 11, pp. 5933–5944, 2011.

[20] A. D. Klose, V. Ntziachristos, and A. H. Hielscher, "The inverse source problem based on the radiative transfer equation in optical molecular imaging," *Journal of Computational Physics*, vol. 202, no. 1, pp. 323–345, 2005.

[21] M. Schweiger, S. R. Arridge, M. Hiraoka, and D. T. Delpy, "The finite element method for the propagation of light in scattering media: boundary and source conditions," *Medical Physics*, vol. 22, no. 11 I, pp. 1779–1792, 1995.

[22] S. Ahn, A. J. Chaudhari, F. Darvas, C. A. Bouman, and R. M. Leahy, "Fast iterative image reconstruction methods for fully 3D multispectral bioluminescence tomography," *Physics in Medicine and Biology*, vol. 53, no. 14, pp. 3921–3942, 2008.

[23] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[24] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[25] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[26] S. Koenig, M. Likhachev, Y. Liu, and D. Furcy, "Incremental heuristic search in AI," *Artificial Intelligence Magazine*, vol. 25, no. 2, pp. 99–112, 2004.

[27] N. B. Karahanoglua and H. Erdogan, "Compressed sensing signal recovery via A* Orthogonal Matching Pursuit," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '11)*, pp. 3732–3735, 2011.

[28] S. Koenig, M. Likhachev, and D. Furcy, "Lifelong Planning A* ," *Artificial Intelligence*, vol. 155, no. 1-2, pp. 93–146, 2004.

[29] R. Dechter and J. Pearl, "Generalized best-first search strategies and the optimality of A* ," *Journal of the ACM*, vol. 32, no. 3, pp. 505–536, 1985.

[30] R. Zhou and A. Hansen, "Beam-stack search: integrating backtracking with beam search," in *Proceedings of the The International Conference on Automated Planning & Scheduling (ICAPS '05)*, pp. 90–98, 2005.

[31] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming B*, vol. 45, no. 3, pp. 503–528, 1989.

[32] J. Feng, K. Jia, C. Qin et al., "Three-dimensional bioluminescence tomography based on bayesian approach," *Optics Express*, vol. 17, no. 19, pp. 16834–16848, 2009.

[33] X. Ma, J. Tian, C. Qin et al., "Early detection of liver cancer based on bioluminescence tomography," *Applied Optics*, vol. 50, no. 10, pp. 1389–1395, 2011.

[34] Y. Lv, J. Tian, W. Cong et al., "Spectrally resolved bioluminescence tomography with adaptive finite element analysis: methodology and simulation," *Physics in Medicine and Biology*, vol. 52, no. 15, pp. 4497–4512, 2007.

[35] D. Han, J. Tian, S. Zhu et al., "A fast reconstruction algorithm for fluorescence molecular tomography with sparsity regularization," *Optics Express*, vol. 18, no. 8, pp. 8630–8646, 2010.

[36] D. Han, J. Tian, C. Qin, K. Liu, and X. Ma, "A fast reconstruction method for fluorescence molecular tomography based on improved iterated shrinkage," in *Proceedings of the Medical Imaging : Biomedical Applications in Molecular, Structural, and Functional Imaging (SPIE '011)*, Lake Buena Vista, Fla, USA, February 2011.

[37] B. Zhang, X. Yang, C. Qin et al., "A trust region method in adaptive finite element framework for bioluminescence

tomography," *Optics Express*, vol. 18, no. 7, pp. 6477–6491, 2010.

[38] G. Alexandrakis, F. R. Rannou, and A. F. Chatziioannou, "Tomographic bioluminescence imaging by use of a combined optical-PET (OPET) system: a computer simulation feasibility study," *Physics in Medicine and Biology*, vol. 50, no. 17, pp. 4225–4241, 2005.