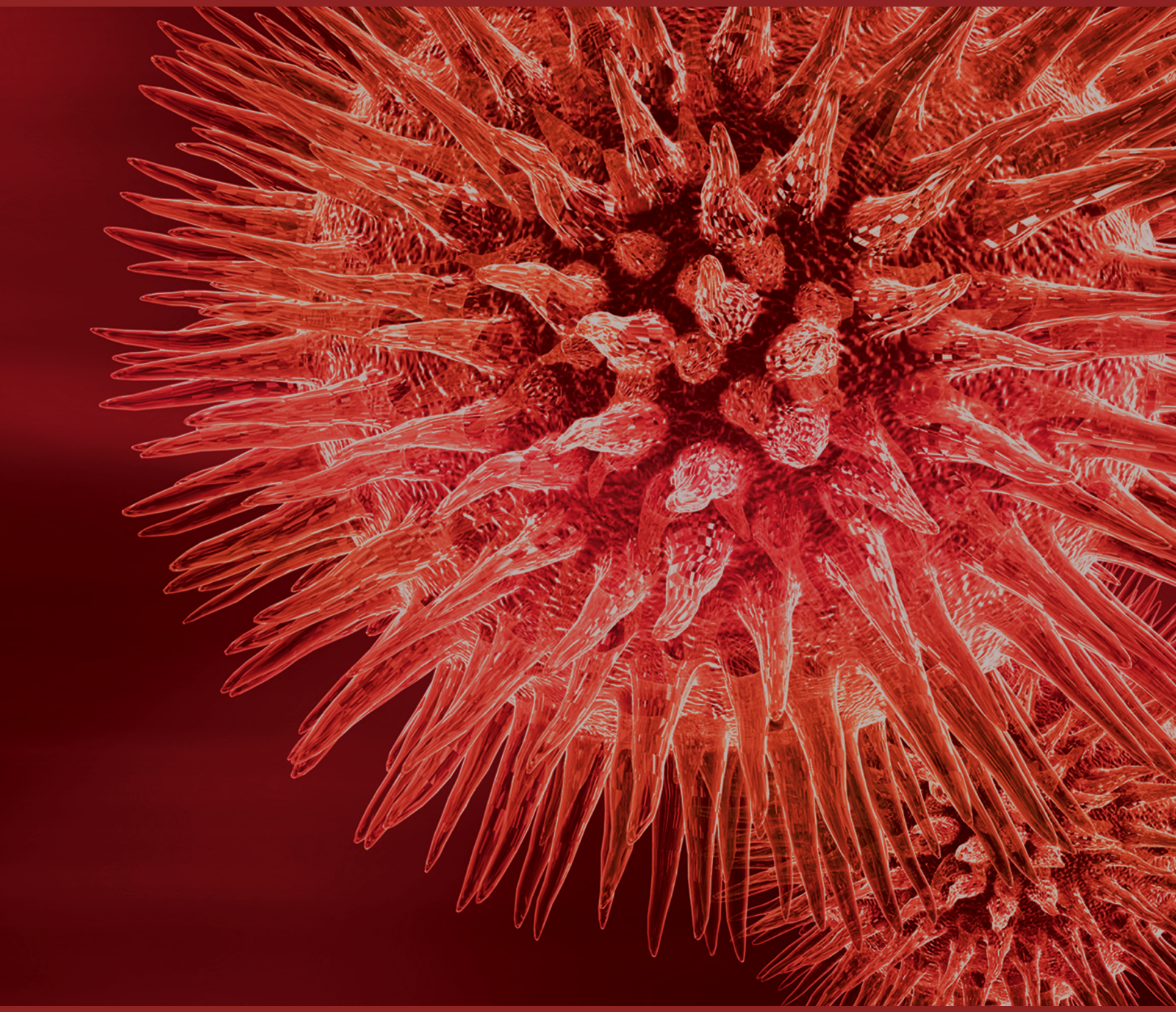
BioMed Research International

Machine Learning in Multimodal Medical Imaging

Guest Editors: Yong Xia, Zexuan Ji, Andrey Krylov, Hang Chang, and Weidong Cai



Machine Learning in Multimodal Medical Imaging

BioMed Research International

Machine Learning in Multimodal Medical Imaging

Guest Editors: Yong Xia, Zexuan Ji, Andrey Krylov, Hang Chang, and Weidong Cai

Contents

Estimation of Response Functions Based on Variational Bayes Algorithm in Dynamic Images Sequences
Bowei Shan
Volume 2016, Article ID 4851401, 9 pages

Hindawi
BioMed Research International
Volume 2017, Article ID 1278329, 2 pages
http://dx.doi.org/10.1155/2017/1278329

Editorial

Machine Learning in Multimodal Medical Imaging

Yong Xia,[1] Zexuan Ji,[2] Andrey Krylov,[3] Hang Chang,[4] and Weidong Cai[5]

[1] *Shaanxi Key Lab of Speech & Image Information Processing (SAIIP), School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China*
[2] *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China*
[3] *Faculty of Computational Mathematics & Cybernetics, Lomonosov Moscow State University, Moscow, Russia*
[4] *Berkeley Biomedical Data Science Center (BBDS), Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[5] *Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Sydney, NSW 2006, Australia*

Correspondence should be addressed to Yong Xia; yxia@nwpu.edu.cn

Received 15 February 2017; Accepted 16 February 2017; Published 5 March 2017

Machine learning techniques have been increasingly applied in the medical imaging field for developing computer-aided diagnosis and prognosis models. Multimodal medical imaging can provide us with separate yet complementary structure and function information of a patient study and hence has transformed the way we study living bodies. Therefore, using machine learning techniques to deal with multimodal medical images is much more challenging due to the diversity of biophysical-biochemical mechanisms. In these years, researchers mainly adapt modern machine learning and pattern recognition techniques such as supervised, unsupervised, semisupervised, and deep learning to solve multimodal medical imaging related problems.

To record the ideas of talents and gather more contributions to these fields, this special issue was launched and supported by this journal. This special issue focuses on the new imaging modalities/methodologies and new machine learning algorithms/applications for the further development in the multimodal medical imaging field, which will provide opportunities for academics and industrial professionals to discuss the latest issues and progresses in the area of multimodal medical imaging. The papers contained in this special issue address the development and application of medical image segmentation, registration, fusion, classification, image restoration, image retrieval, and computer-aided diagnosis.

In "Estimation of Response Functions Based on Variational Bayes Algorithm in Dynamic Images Sequences," B. Shan proposes a nonparametric Bayesian model to estimate the response functions in dynamic medical imaging, in which the nonparametric Bayesian priors are designed to favor desirable properties of the functions and used to improve the estimation of response functions.

In "Two-Layer Tight Frame Sparsifying Model for Compressed Sensing Magnetic Resonance Imaging," S. Wang et al. propose a two-layer tight frame sparsifying model for compressed sensing magnetic resonance imaging (MRI) by sparsifying the image with a product of a fixed tight frame and an adaptively learned tight frame, which is solved by a three-level Bregman numerical algorithm and enables accurate MRI reconstruction from highly undersampled data with efficiency.

In "Many is Better than One: An Integration of Multiple Simple Strategies for Accurate Lung Segmentation in CT Images," Z. Shi et al. present a novel computerized tomography (CT) lung image segmentation method by integrating multiple strategies, including the guided filter to smooth the image, the optimized threshold to get binary image, region-growing strategy to extract thorax regions, and random walk algorithm to segment lung regions and to get the state-of-the-art segmentation accuracy.

In "Pulmonary Nodule Detection Model Based on SVM and CT Image Feature-Level Fusion with Rough Sets," T. Zhou et al. present a pulmonary nodules detection algorithm based on support vector machine (SVM) and CT image

feature-level fusion with rough sets to improve the detection accuracy of pulmonary nodules in CT image. Both the unreasonable feature structure and the nontightness of feature representation are taken into consideration in this pulmonary nodules detection algorithm.

In "Multigrid Nonlocal Gaussian Mixture Model for Segmentation of Brain Tissues in Magnetic Resonance Images," Y. Chen et al. propose a novel segmentation method based on the regional and nonlocal information to overcome the impact of image intensity inhomogeneities and noise in human brain magnetic resonance images.

In "DTI Image Registration under Probabilistic Fiber Bundles Tractography Learning," Z. Guo et al. propose a diffusion tensor imaging (DTI) image registration method under probabilistic fiber bundles tractography learning, where the residual error model is modified with finite sample set and the calculated deformation field is then registered on the DTI images.

In "Automated Segmentation of Coronary Arteries based on Statistical Region Growing and Heuristic Decision Method," Y. Tian et al. propose a fully automated coronary artery segmentation from cardiac data volume based on a statistics region growing together with a heuristic decision to further help cardiovascular radiologists detect and quantify stenosis.

In "Rapid Retrieval of Lung Nodule CT Images Based on Hashing and Pruning Methods," L. Pan et al. propose a new retrieval framework based on a hashing method for lung nodule CT images, which can translate high-dimensional image features into a compact hash code to greatly reduce the retrieval time and memory space. Moreover, a pruning-based decision rule is utilized in this algorithm to improve its retrieval precision.

In "The Classification of Tongue Colors with Standardized Acquisition and ICC Profile Correction in Traditional Chinese Medicine," Z. Qi et al. design a tongue color classification approach using a standardized tongue image acquisition process, color correction, and several machine learning techniques for tongue inspection-based diagnosis in traditional Chinese medicine.

In "Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images," J. Zhang et al. develop a SVM-based diagnostic method for diabetes using standardized tongue images. This work shows the potential of applying digitalized tongue images, which are usually used in traditional Chinese medicine, to the diagnosis of diabetes.

In "A Computer-Aided Analysis Method of SPECT Brain Images for Quantitative Treatment Monitoring: Performance Evaluations and Clinical Applications," X. Zheng et al. introduce and validate a computer-aided analysis method to achieve the quantitative treatment monitoring based on single-photon emission computed tomography (SPECT) images, which can provide a convenient solution to generate a parametric image and derive the quantitative indexes from the longitudinal SPECT brain images for treatment monitoring.

The papers in this special issue provide a useful message of machine learning techniques in dealing with multimodal medical images. This unique and informative collection of papers highlights the direction of related studies. This special issue illustrates the important role that machine learning techniques play in the multimodal medical imaging fields.

Acknowledgments

We would like to express our appreciation to all the authors for their informative contributions and the reviewers for their scientific contribution and support in making this special issue possible.

Yong Xia
Zexuan Ji
Andrey Krylov
Hang Chang
Weidong Cai

Hindawi Publishing Corporation
BioMed Research International
Volume 2017, Article ID 1962181, 11 pages
http://dx.doi.org/10.1155/2017/1962181

Research Article

A Computer-Aided Analysis Method of SPECT Brain Images for Quantitative Treatment Monitoring: Performance Evaluations and Clinical Applications

Xiujuan Zheng,[1,2] Wentao Wei,[1] Qiu Huang,[3] Shaoli Song,[2] Jieqing Wan,[4] and Gang Huang[2]

[1] *Department of Automation, The School of Electrical Engineering Information, Sichuan University, Chengdu, China*
[2] *Department of Nuclear Medicine, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China*
[3] *Med-X Institute, Shanghai Jiaotong University, Shanghai, China*
[4] *Department of Neurosurgery, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China*

Correspondence should be addressed to Xiujuan Zheng; xiujuanzheng@scu.edu.cn

Received 30 September 2016; Revised 4 December 2016; Accepted 28 December 2016; Published 31 January 2017

Academic Editor: Zexuan Ji

The objective and quantitative analysis of longitudinal single photon emission computed tomography (SPECT) images are significant for the treatment monitoring of brain disorders. Therefore, a computer aided analysis (CAA) method is introduced to extract a change-rate map (CRM) as a parametric image for quantifying the changes of regional cerebral blood flow (rCBF) in longitudinal SPECT brain images. The performances of the CAA-CRM approach in treatment monitoring are evaluated by the computer simulations and clinical applications. The results of computer simulations show that the derived CRMs have high similarities with their ground truths when the lesion size is larger than system spatial resolution and the change rate is higher than 20%. In clinical applications, the CAA-CRM approach is used to assess the treatment of 50 patients with brain ischemia. The results demonstrate that CAA-CRM approach has a 93.4% accuracy of recovered region's localization. Moreover, the quantitative indexes of recovered regions derived from CRM are all significantly different among the groups and highly correlated with the experienced clinical diagnosis. In conclusion, the proposed CAA-CRM approach provides a convenient solution to generate a parametric image and derive the quantitative indexes from the longitudinal SPECT brain images for treatment monitoring.

1. Introduction

Single photon emission computed tomography (SPECT) with ^{99m}Tc-ethyl cysteine dimer (^{99m}Tc-ECD) has been widely used to evaluate many types of cerebrovascular diseases and brain disorders by measuring regional cerebral blood flow (rCBF) [1–6]. Moreover, longitudinal ^{99m}Tc-ECD SPECT brain imaging can be adopted to monitor the changes of rCBF to support the treatment plan [7, 8]. The most common approach for interpreting SPECT brain images is visual inspection in daily clinical practice. Usually the relevant structural images, such as CT or MRI images, are preferred for the visual interpretation together with SPECT images. For treatment monitoring, the baseline and follow-up SPECT brain images should be parallelly interpreted under the same condition to figure out the differences. Due to the lack of the quantitative standards, the accuracy and reliability of visual inspection mainly rely on the experience of the physicians, such that only the qualitative results are presented in the reports. The aforementioned drawbacks of vissual insspection prevent the applications of SPECT imaging in distant diagnosis and multicenter studies. The image processing technology can bring solutions to the visual inspection problems and explore the hidden information in the images.

Beside the region of interest (ROI)/volume of interest (VOI) analysis methods, statistical parametric mapping (SPM) has been used in group-wise comparisons of functional brain images to evaluate the responses to the treatments [9, 10]. On the other hand, subtraction analysis is also a useful technology to extract differences in a series

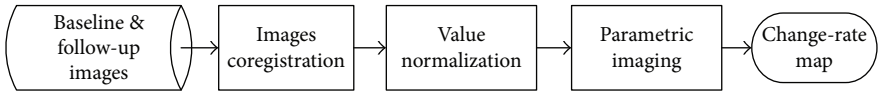


Figure 1: The workflow of the computer-aided analysis method to extract a change-rate map.

of images for individual treatment monitoring. In ictal-interictal SPECT imaging, the subtraction analysis has been proved to benefit the treatment plan of epilepsy patients [11–14]. Recent studies have reported that the quantitative SPECT analysis would be playing an ever-growing role in treatment plan and response monitoring of several disorders related with the central nervous system [15, 16]. These studies encourage us to generate the parametric imaging and extract quantitative indexes from the SPECT images to support the treatment plan. In this study, we introduce a computer-aided analysis (CAA) method inherited from subtraction analysis to quantify the changes of rCBF in longitudinal SPECT images for individual treatment monitoring. The performance of the proposed method would be objectively and systematically evaluated by the computer simulations and the clinical applications.

2. Materials and Methods

2.1. Computer-Aided Analysis Method. When using SPECT imaging in treatment monitoring, the pre- and postscans are usually performed to acquire the baseline and follow-up SPECT images before and after delivering treatments. The obtained baseline and follow-up images from an individual subject are a set of longitudinal SPECT images requiring independent analysis. For interpreting the individual subject's data, a computer-aided analysis (CAA) method is established to process the longitudinal SPECT images via three main steps: coregistration, value normalization, and parametric imaging. The workflow chart is shown in Figure 1.

In the first step, the follow-up SPECT brain images are aligned with the baseline images by the rigid registration algorithm provided by SPM 8.0 software package (http://www.fil.ion.ucl.ac.uk/spm/). If the SPECT brain imaging is performed using SPECT-CT integrated system, the CT and SPECT images can be obtained in the same position and considered as well aligned. In this condition, the baseline and follow-up CT images could be used as the reference and source image, respectively, in the step of coregistration for aligning longitudinal SPECT images.

In the second step, the value normalization is applied on the SPECT images. Before the numerical calculation, the extracerebral voxels are removed by a predefined whole-brain mask. If only the SPECT image is available, then the whole-brain mask can be obtained by segmenting the enhanced SPECT image with Otsu's algorithm. If the corresponding aligned CT image is available, then the whole-brain mask can be more accurately defined by separating the brain tissues from nonbrain tissue in CT images using fuzzy C-means clustering algorithm [17]. The whole-brain mask extracted based on the CT images is then applied on the SPECT images to delineate the brain area.

After deriving the whole-brain area in the SPECT images, the value of each cerebral voxel is normalized by the average voxel value of the reference area that is automatically selected by Z-map approach [18]. In Z-map approach, the Z value of the ith voxel was calculated as in (1). Two Z-maps are, respectively, estimated for baseline and follow-up SPECT brain images. Then, the reference region is the intersection of the $Z < 1$ areas of these two Z-maps.

$$Z_i = \frac{|C_i - \text{mean}|}{\text{SD}}, \tag{1}$$

where C_i is the value of the ith voxel in one SPECT brain image; mean and SD, respectively, denote the average and standard deviation of voxel values of brain area in the SPECT image.

In the third step, the changes in the longitudinal SPECT images are expressed in a parametric image to reflect the disease progress or responses to the treatment. As in subtraction analysis, the difference can be directly obtained by subtracting two aligned normalized images, as

$$D_i = C_i^f - C_i^b, \tag{2}$$

where C_i^f and C_i^b denote the normalized values of the ith voxel in the follow-up and baseline SPECT images, respectively.

Next, the change-rate map can be calculated voxel-by-voxel to reflect the extent of the changes between the baseline and follow-up images. The value of the ith voxel in the estimated change-rate map is denoted by R_i, which can be derived from

$$R_i = \frac{D_i}{C_i^b}. \tag{3}$$

The change-rate map (CRM) is a parametric image which can be fused with aligned SPECT/CT images for visual inspections. The positive voxel value in CRM demonstrates the recovery of hypoperfusion, while the negative value indicates the recovery of hyperperfusion. For the visualization, Gaussian smoothing filter can be applied in the CRM to reduce the impact of noise. Based on the CRM, the changed regions are automatically obtained by thresholding and clustering. Firstly, the voxels with lower change-rate (<20%) were set to 0 in CRM. Then, the K-means clustering algorithm is adopted to recognize the changed regions. The morphological processing is applied to refine and distinguish each region. Considering the SPECT image resolution and partial volume effects, the regions with larger volumes (>120 voxels) are selected as the recovered regions.

For the localization of recovered regions, an atlas of brain lobes, which consists of 12 brain anatomical structures (listed

Table 1: The list of brain anatomical structures in the atlas of brain lobes.

Brain anatomical structures	Values in atlas
Cerebellum anterior lobe	80
Cerebellum posterior lobe	10
Frontal lobe	Left: 50, right: 55
Frontal-temporal space	Left: 110, right: 115
Limbic lobe	Left: 40, right: 45
Medulla	20
Midbrain	100
Occipital lobe	Left: 90, right: 95
Parietal lobe	Left: 120, right: 125
Pons	70
Sublobar	60
Temporal lobe	Left: 30, right: 35

in Table 1), is created from the Talairach Daemon atlas [19, 20] and then translated into MNI (Montreal Neurological Institute) space [21] with dimensions of $91 \times 109 \times 91$ sampled at 2 mm intervals, corresponding to the SPM templates [22]. The SPECT/CT images as well as the obtained CRM are mapped to SPM template by the nonrigid registration algorithm provided by SPM 8.0 software package and then aligned with the atlas of brain lobes. Based on the atlas of brain lobes, the recovered regions could be located in the different brain lobes. The quantitative indexes for the recovered regions, such as the mean and maximum change-rate and the proportion of the recovered regions to the corresponding brain lobes, could be calculated for each detected recovered regions.

In order to facilitate the expression, the proposed approach used to evaluate longitudinal SPECT images through a CRM derived by the CAA method is noted as CAA-CRM approach in the subsequent parts.

2.2. Computer Simulations

2.2.1. Simulated Data. In this study, the performance of the CAA-CRM approach in treatment monitoring was objectively and systematically evaluated by the computer simulations. The longitudinal SPECT images are simulated using predesigned digital brain phantoms. The normal digital brain phantom is a $100 \times 100 \times 82$ matrix representation of the hardware Hoffman phantom [23], whose voxel size is 2.13 mm $\times$ 2.13 mm $\times$ 2.13 mm. In the digital brain phantom, the value of each voxel presents the radioactivity in the corresponding position. Generally, this digital brain phantom is used to simulate the normal brain perfusion images acquired by ^{99m}Tc-ECD SPECT imaging. Comparing with other regular geometrical objects, the sphere is more suitable for simulating the ischemic lesions in perfusion images. For convenience, a sphere is created in the normal digital brain phantom located in the right frontal lobe as a lesion analogue. The diameter and radioactivity of the sphere can be changed for several scales to simulate the varied degrees of hypoperfusion for brain ischemia. The diameter of the sphere was designed in three scales: 8 mm, 16 mm, and 24 mm. In addition, the radioactivity in the sphere was set based on the predefined change-rate scaled in 9 different levels uniformly distributed from 10% to 90%. The simulated brain SPECT images are generated with an injected dose of 25 mCi of ^{99m}Tc-ECD. The abnormal and normal brain images are, respectively, regarded as the baseline and follow-up images obtained in treatment monitoring for brain ischemia.

The system parameters used in computer simulations are set according to the geometry of the dual-head Philips Precedence 6 SPECT/CT scanner. Each detector head is mounted with a low-energy and high-resolution (LEHR) collimator. The two heads rotate in H-mode to obtain 128 projections in total over 360° around the predesigned phantom. Projection data are acquired for 10 minutes with a total count number of 10^8 accompanied with measurement noise that is modeled as an additive Poisson noise. The radioactivity distribution in the brain phantom is reconstructed with a maximum a posteriori (MAP) algorithm with total variation regularization.

2.2.2. Performance Evaluation. In the performance evaluation, the ground truth of CRM is directly defined based on the phantoms for every simulated lesion size at each change-rate. The estimated CRM is compared to the corresponding ground truth for evaluating its quality. In this study, the indexes of image quality are adopted to objectively and systematically quantify the performance. Denote the value of the ith voxel in the estimated CRM as R_i, and denote the value of corresponding voxel in ground truth as G_i. Thus, the normalized absolute error (NAE), which is the simplest metric for measuring the difference between two images, can be calculated by

$$\mathrm{NAE} = \frac{\sum_{i=1}^{N} \left| R_i - G_i \right|}{\sum_{i=1}^{N} G_i}. \tag{4}$$

As shown in (5), the peak signal to noise ratio (PSNR) is the index to reflect the image quality of obtained CRM comparing with its ground truth.

$$\begin{aligned} \mathrm{PSNR} &= 20 \log_{10} \frac{G_{\max}}{\sqrt{\mathrm{MSE}}}, \\ \mathrm{MSE} &= \frac{1}{N} \sum_{i=1}^{N} \left(R_i - G_i \right)^2, \end{aligned} \tag{5}$$

where $G_{\max}$ is the maximum voxel value of ground truth of CRM; MSE is for mean square error.

The normalized cross-correlation (NCC) is used to quantify the similarity between the obtained CRM and its ground truth. The NCC can be calculated by

$$\mathrm{NCC} = \frac{1}{N-1} \sum_{i=1}^{N} \frac{\left(R_i - \overline{R} \right) \left(G_i - \overline{G} \right)}{\sigma_c \sigma_G}, \tag{6}$$

where $\overline{R}$ and $\overline{G}$ represent the mean values of the obtained CRM and the corresponding ground truth, respectively; σ_c and σ_G denote their standard deviations.

After the recovered region is detected based on CRM, the Dice similarity coefficient (DSC), which is calculated as in (7), is used to measure the accuracy of the recovered regions detection comparing with the ground truth that is the predefined sphere in the phantom.

$$\mathrm{DSC} = 2 \times \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{TP} + \mathrm{FP}}, \tag{7}$$

where TP is true positive, that is, the set of voxels common to the derived recovered region and ground truth; TN is true negative, that is, the set of voxels not labelled as the derived recovered region and ground truth; FN is false negative; and FP is false positive.

Moreover, the change-rates derived from the recovered regions are also used to evaluate the accuracy of recovered region detection. Because of the homogeneity of voxel values in the predefined digital phantom, only the mean change-rate of the recovered region is calculated and compared with the predefined real value by linear regressions.

2.3. Clinical Applications

2.3.1. Clinical Data Acquisition. In this study, ^{99m}Tc-ECD SPECT brain imaging is used in the treatment monitoring of the internal carotid artery (ICA) stenting, which is a common treatment technique for brain ischemia. This study has been approved by the Ethics Committee of Renji Hospital, School of Medicine, Shanghai Jiaotong University. All the SPECT scans are performed in accordance with the guidelines for brain perfusion SPECT using ^{99m}Tc-labelled radiopharmaceuticals [1]. 50 patients in total (7 women, 43 men, and average age 62.9 ± 10.5 years) prescribed ICA stenting are recruited. 27 of them have cerebral infarction, while the rest suffer different degrees of cerebral ischemia. For each patient, the baseline ^{99m}Tc-ECD SPECT imaging is performed within 7 days before surgery, and then the follow-up scan is generally conducted in ~7 days (ranged from 2 to 12 days) after the treatment of ICA stenting. The SPECT imaging was started within 20~30 minutes after the radiotracer injection (around 25 mCi ^{99m}Tc-ECD) using dual-head Philips Precedence 6 SPECT/CT scanner with low-energy and high-resolution collimators. For 41 patients, the CT scans are performed together with SPECT imaging. For the rest of 9 patients, the corresponding CT images are not available. The system resolution is 7.4 mm full width half maximum (FWHM) at 10 cm for SPECT imaging. Three-dimensional SPECT images were reconstructed using Astonish™ technology, which adopts an iterative ordered-subset expectation-maximization (OSEM) algorithm with built-in scatter correction and attenuation correction [24]. For each patient, a pair of baseline and follow-up SPECT images was used to evaluate the therapy of ICA stenting.

2.3.2. Clinical Data Analysis. For the data analysis, the traditional visual inspection and the CAA-CRM approach are both used to assess the recovery levels. For the visual inspections, the baseline and follow-up SPECT brain images were compared by two independent experienced physicians in the same image workstation. When CT images are available, the physicians inspected the SPECT images with the support of the aligned CT images. The hypoperfusion lesions caused by cerebral infarction or ischemia are carefully studied, and then the recovered regions are delineated manually. Furthermore, the overall recovery level for each patient is formally reported in four scales (none, mild recovery, moderate recovery, and severe recovery) based on the physicians' experience. On the other hand, these longitudinal SPECT images are quantitatively analyzed by the CAA-CRM approach. After all automatic processes, a CRM is derived for each patient. Then, the estimated CRM is fused with the original SPECT/CT images. The recovered regions are automatically derived based on the CRM by thresholding and clustering. The threshold is set as 20% according to the results of computer simulations (as mentioned in Section 3). Small regions (less than 120 voxels) are excluded to eliminate the influence of measurement noise. Then, the automatically recovered regions are located by the atlas of brain lobes and the quantitative indexes related to recovery level are estimated. In the further evaluation of the performance of CAA-CRM approach in treatment monitoring, the clinical diagnosis derived by visual inspection is considered as the standard of the classification of recovery groups to validate the results of the CAA-CRM approach.

In the performance evaluation, the automatically recovered regions are firstly compared with manually defined ones. If an automatic recovered region hits the corresponding manual one, it would be considered as a successful detection of the real recovered region. Moreover, McNemar's test [25] for paired automatic and manual recovered regions is used to investigate the concordance of the CAA-CRM approach and traditional visual inspection in the detection of recovered regions. Meanwhile, quantitative indexes, including the mean change-rate, maximum change-rate, and proportion of the recovered regions to the corresponding lobes, are estimated from the CRM. The statistics of these quantitative indexes related to the recovery levels are analyzed in three different recovery groups classified according to clinical reports. Nonparametric one-way ANOVA is also applied in the analysis of variance of different recovery groups.

3. Results

3.1. Results of Computer Simulation. Using CAA-CRM approach, the CRMs are derived based on the simulated data. These estimated CRMs are compared with the corresponding ground truths. The properties of CRM in treatment monitoring are quantified by image quality indexes, including the NAE, PSNR, and NCC. The comparative results of the image quality indexes for different lesion sizes at each change-rate are shown in Figure 2. In Figure 2(a), NAE declines with the increase of change-rate for each lesion size. Moreover, for small-size lesion (ϕ8 mm), the calculated NAE is much higher than larger-size lesions (ϕ16 mm and ϕ24 mm), especially in the condition of low change-rates (10% and 20%). In Figure 2(b), PSNR progressively rises along the increase of change-rate, and it climbs quicker in the case of small-size lesion. Figure 2(c) shows that NCC goes up with the increase of change-rate when the lesion size is comparatively large

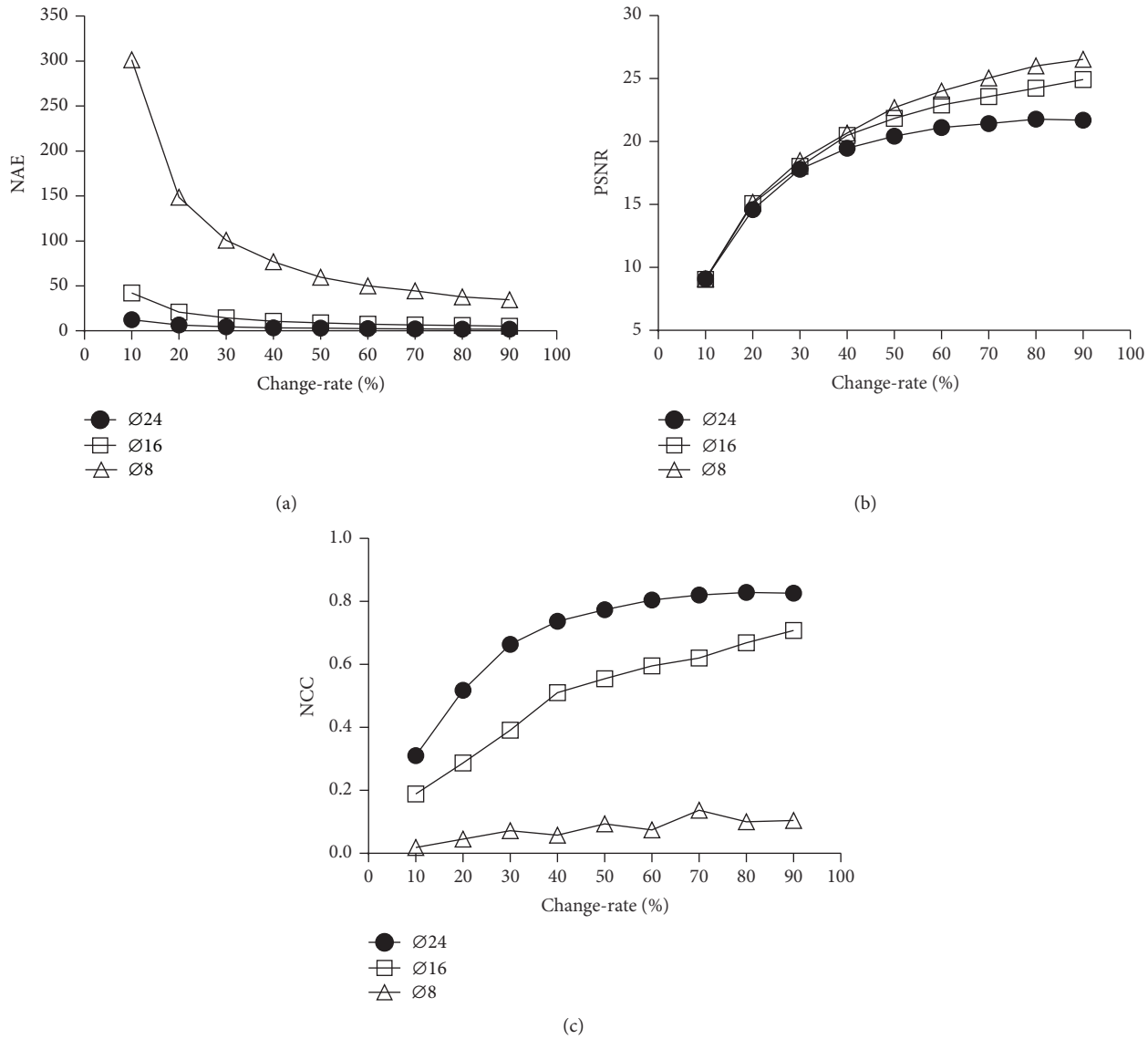


Figure 2: Comparisons of the image quality indexes for three-size lesions (ϕ8 mm: diameter of 8 mm, ϕ16 mm: diameter of 16 mm, and ϕ24 mm: diameter of 24 mm) at nine different change-rates (10% to 90%). (a) The normalized average error (NAE) for measuring the difference between the change-rate map (CRM) and its ground truth; (b) the peak signal-noise ratio (PSNR) for reflecting the image quality of the CRM; (c) the normalized cross-correlation (NCC) for quantifying the similarity between the CRM and its ground truth.

(ϕ16 mm and ϕ24 mm). However, NCC increases slightly for small-size lesion.

DSC and mean change-rate estimates of recovered region are both used to evaluate the performance of the derived CRM in recovered region detection. From the curve chart in Figure 3(a), the DSC of large lesion (ϕ24 mm) maintains a higher value (>0.7). For the medium lesion (ϕ16 mm), the DSC increases quickly with rising of change-rate when the change-rate is less than 40%. Meanwhile, the DSC changes slightly and stays in a high level when the recovery level is higher than 40%. However, the DSC fluctuates with a low value along the increase of recovery level for the small lesion (ϕ8 mm). In Figure 3(b), the linear regressions are plotted for three different lesion sizes. The highly linear relations ($r^2 = 0.99$, $p < 0.0001$) between the change-rates estimates and real values are observed in the condition of larger lesions (ϕ16 mm and ϕ24 mm). It is also found that the estimated mean change-rates are underestimated since the liner regression lines of the estimates and real values are under the line $y = x$ (the black solid line). The underestimation is more serious for the small lesions. It seems that the CAA-CRM approach fails in deriving the acceptable change-rate for small-size lesion with diameter of 8 mm.

Based on the results of compuater simulations, it can be concluded that the CAA-CRM approach has a better performance in the detection of recovered regions and

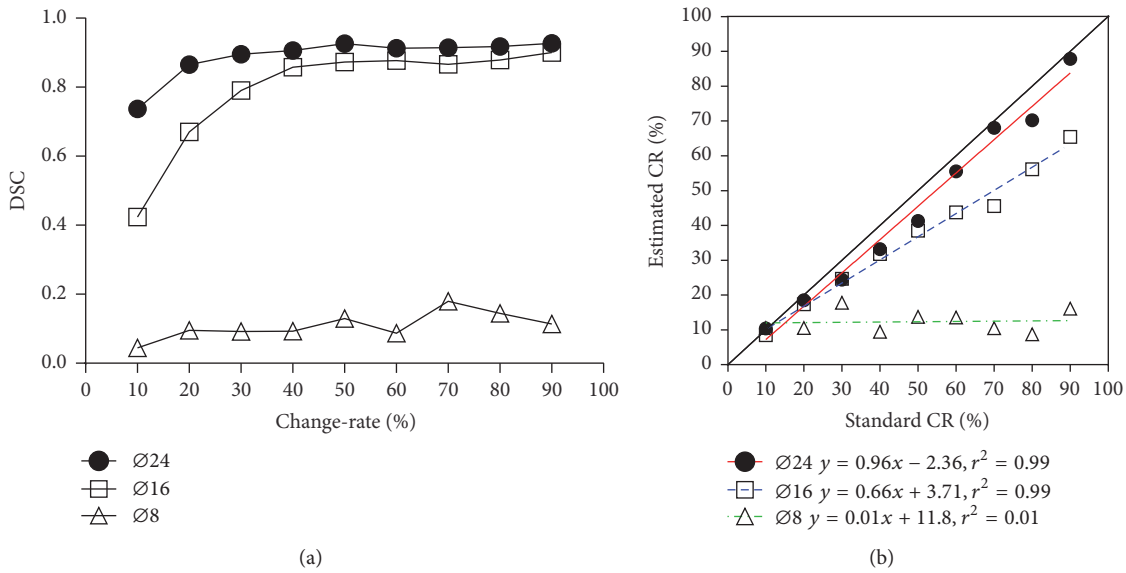


Figure 3: Comparisons of detected recovered regions with the ground truth for three-size lesions (ϕ8 mm: diameter of 8 mm, ϕ16 mm: diameter of 16 mm, and ϕ24 mm: diameter of 24 mm) at nine different change-rates (10% to 90%). (a) The Dice similarity coefficient (DSC) for measuring the accuracy of the recovered regions detection comparing with the ground truth; (b) linear regressions of estimated change-rates (CR) of the recovered regions with the real values.

the quantification of change-rates, when the lesion size is sufficiently large (larger than a sphere with diameter of 8 mm that is closed to the proposed spatial resolution of SPECT images) and the change-rate is high enough (at least not lower than 20%).

3.2. Results of Clinical Application. For the visual inspection, 17 of 50 patients are reported as severe recovery in brain perfusion after ICA stenting, while 22 patients are scaled as moderate recovery. In the remaining 11 patients, there are 8 patients with mild recovery, and 3 patients are diagnosed as no improvement along with the treatment. Considering the population distributions of these 4 groups, the mild and none recovery patients are combined as one group to compare with the other groups. In the visual inspection, 106 manual VOIs are totally delineated to locate the recovered regions for these 50 patients.

Beside the visual inspection, the CAA-CRM approach is applied to monitor the changes based on the longitudinal SPECT images for each patient. Figure 4 illuminates a typical case of severe recovery. In Figure 4(a), there are three transverse slices (the 36th, 40th, and 44th slices) of baseline ^{99m}Tc-ECD SPECT images. The lesion of cerebral infarction can be clearly found in the left parietal lobe, which is pointed by a white arrow. The cerebral ischemia is easily detected for the hypoperfusion regions around the lesion. The corresponding slices of aligned follow-up SPECT images are shown in Figure 4(b). In the clinical report based on visual inspection, the lesion of cerebral infarction had no improvement after the treatment of ICA stenting. However, the cerebral blood flow recovers significantly in the hypoperfusion regions around the lesion of cerebral infarction. The overall recovery level given by physicians is severe recovery. In Figure 4(c), the estimated CRM presents as a parametric image fused with the corresponding baseline SPECT image. In this case, the derived CRM could be used to enhance the visual inspection for physicians in treatment monitoring. In Figure 4(d), the CRM as well as the corresponding SPECT images is mapped to the standard atlas of brain lobes for the convenient localization of recovered regions.

Table 2: The concordance of automatic VOIs with manual ones in the localization of recovered regions.

	FN	TP	FP
Sever	3	41	5
Moderate	3	42	3
Mild/none	1	16	4
Total	7	99	12

FN: false negative, the number of manually recovered regions which were not hit by automatic ones; TP: true positive, the number of manually recovered regions which were hit by automatic ones; FP: false positive, the number of automatically recovered regions which could not find the paired manual ones.

The CAA-CRM approach also has the advantages in automatic and quantitative analysis of recovered regions. For all the patients' SPECT images, the recovered regions are automatically derived, located, and compared to manual ones which are used as the standard for validation. The detailed results are listed in Table 2 to reflect the concordance between the automatic detection and manual detection of recovered regions. From Table 2, there are in total 99 concordat pairs (automatically recovered regions that hit manual ones) and the total concordance rate of localization is 93.4%. For the rest of 19 discordant pairs, there are 12 (12/19, 63.2%)

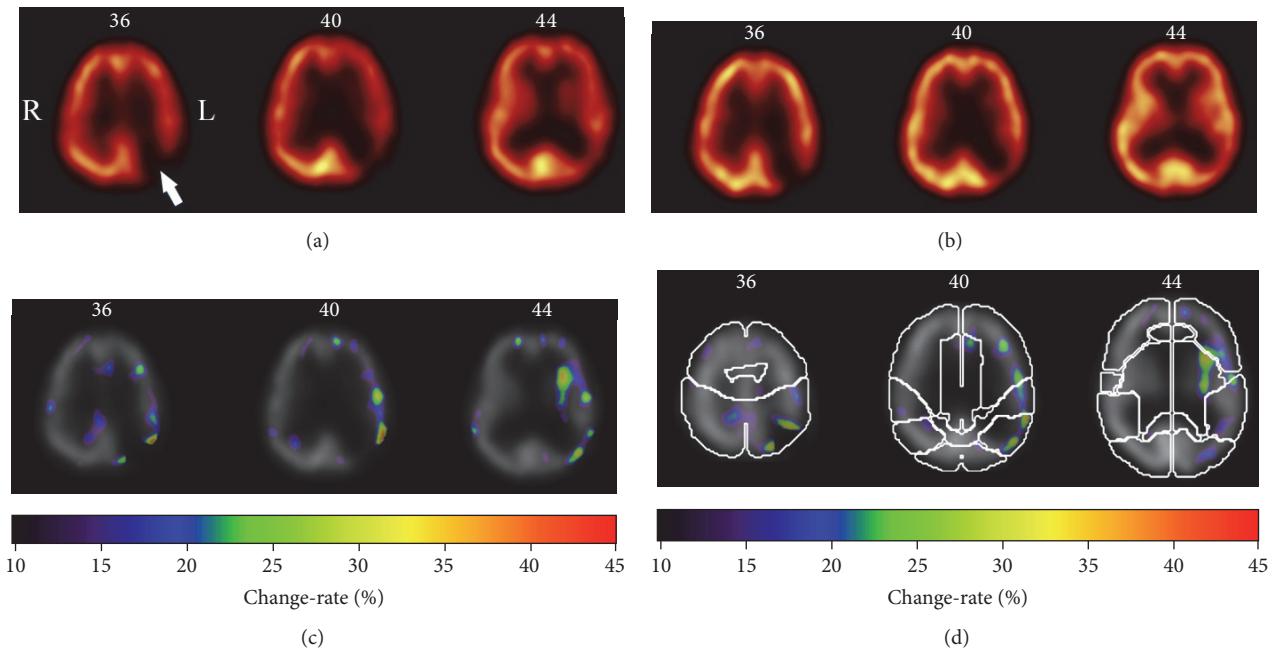


Figure 4: A typical case of treatment monitoring using CAA-CRM approach, where the estimated change-rate map can directly reflect response to treatment. The patient (male, 69-year-old) suffered cerebral infarction in left parietal lobe. While the baseline ^{99m}Tc-ECD SPECT scan was performed 3 days before the treatment of ICA stenting, the follow-up scan was obtained 7 days after the treatment. (a) Three transverse slices (the 36th, 40th, and 44th slices) of baseline SPECT images. The lesion of cerebral infarction is pointed out by white arrow. Around this lesion, the hypoperfusion could be observed. (b) Three corresponding transvers slices of the follow-up SPECT image. There is no improvement in the lesion of cerebral infarction, while the severe recovery is observed for the cerebral ischemia in hypoperfusion regions around the lesion. The recovery level was scaled as severe by the traditional visual inspection in clinical report. (c) Three transverse slices of the derived change-rate map (CRM) fused with baseline SPECT image. The scales of change-rate are presented by rainbow color bar. The warmer color denotes higher change-rate. In this case, the recovered regions can be detected easily and clearly in the change-rate map. (d) The selected transverse slices of the CRM fused with the atlas of brain lobes. The contours of the brain lobes are delineated. The recovered regions can be conveniently localized in the brain area.

pairs where the automatic method recommended recovered regions while the physicians did not, and 7 (7/19, 36.8%) pairs are in the contrary condition. By the conventional criteria of McNemar's test ($p < 0.05$), this difference is considered to be not statistically significant. There is also no significant difference between the automatic and manual approaches in localization of recovered regions, whatever group (severe, moderate, and mild/none groups) is chosen. The results indicate that the CAA-CRM approach and visual inspections of experienced physicians have the concordance in the localization of recovered regions.

After localizing the recovered regions, the mean and maximum change-rates for each patient could be calculated based on the delineated recovered regions and then compared in groups. The group-wise results of statistical comparisons are illuminated by bar graphs in Figure 5. Figure 5(a) illuminates that the mean change-rates of the severe recovery group are mainly higher than those for the moderate group and mild/none group. The similar results can be observed in Figure 5(b) for the maximum change-rate estimates for three groups. The statistical results of nonparametric one-way ANOVA indicate that the significant differences ($p < 0.0001$) exist among the mean/maximum change-rates of three different recovery groups. Beside the indexes of change-rate, the proportion of recovered regions to the corresponding brain lobes, which is calculated as a volume ratio between recovered regions and the corresponding brain lobes, is also a specific index for quantifying the recovery level for treatment monitoring. The comparison of proportion of recovered regions for three recovery groups is shown in Figure 5(c). The tendency of the proportions in groups is similar to that of the change-rate. It is obvious that the better recovery groups have the higher proportions. According to the results of nonparametric one-way ANOVA, three recovery groups had significant effect ($p < 0.001$) on proportion of recovered regions to the brain lobes.

To sum up the results of clinical applications, the higher change-rates and larger recovered regions could correspond to better recovery levels given in the clinical reports. These quantitative indexes derived by the CAA-CRM approach can be used to quantify the response to the treatment.

4. Discussion

In this study, the performance of the CAA-CRM approach is objectively and systematically evaluated by the computer simulations as well as the clinical applications. From the results

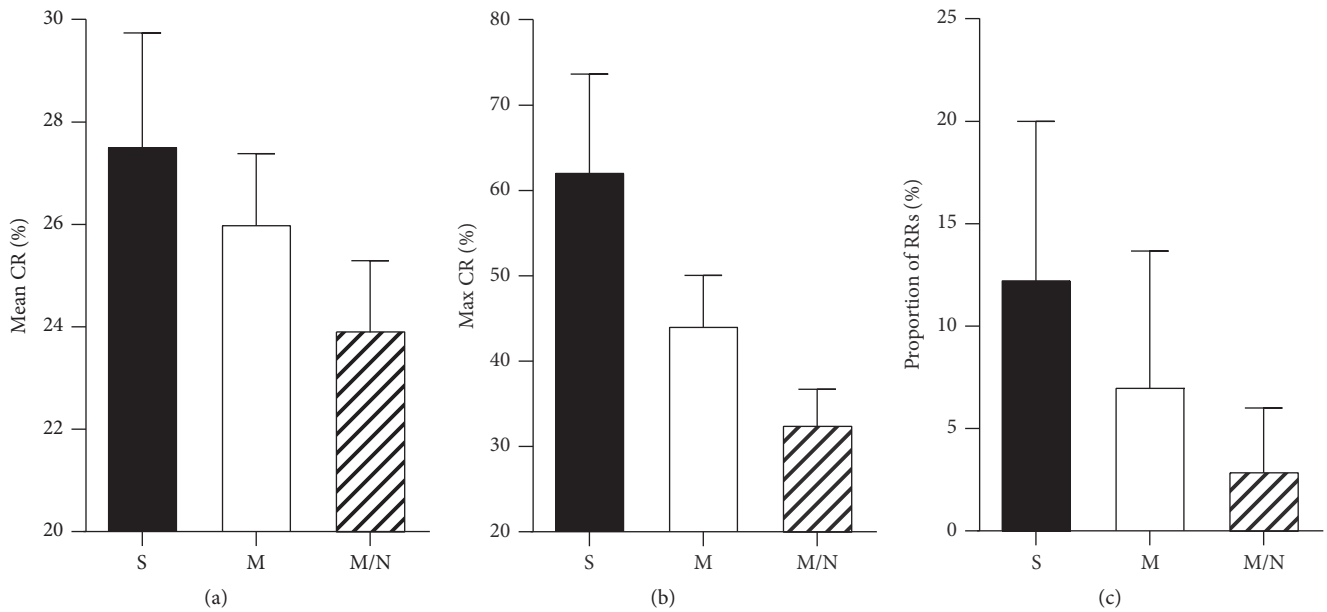


Figure 5: The bar graphs (mean + SD) for comparisons of quantitative indexes for three recovery groups (S: severe, M: moderate, and M/N: mild/none). (a) The mean change-rate (CR) estimates of the recovered regions; (b) the maximum CR estimates of the recovered regions; (c) the proportion of recovered regions (RRs) to the corresponding brain lobes.

of computer simulations, the lesion size and change-rate are considered as two major factors for extracting reliable CRM. According to the changing tendency of image quality indexes of CRM, it can be concluded that the lesions with larger size and higher change-rate could be easier to detect; moreover, the estimates of change-rate could be more accurate as the real values. This conclusion confirmed our experiences from clinical practice. Additionally, the results also clarified that the spatial resolution of SPECT image could be a major limitation of the CAA-CRM approach to get accurate quantitative indexes for quantifying the recovery levels. It has been reported that the quantitative accuracy of the radioactive concentration in SPECT image would deteriorate with the decreasing of target size, especially when the targets are below three times of spatial resolution [26]. For the CAA-CRM approach, the poor quantitative accuracy of original SPECT images would directly result in estimated bias of change-rate and even in the missing recovered regions. As the results shown in Figure 2(c), for the small lesion (ϕ8 mm), whose size is closer to spatial resolution (7.4 mm FWHM at 10 cm), the values of NCC are much lower than those of the larger lesions (ϕ16 mm and ϕ24 mm). The uptrend with increasing change-rate is also very slight. The low value of NCC reflects the poor similarity between the obtained CRM and its ground truth. It means that the obtained CRM could not accurately reflect the small recovered regions when the region size is closed to or even smaller than the spatial resolution of SPECT images.

In the computer simulations, the underestimation of change-rate is observed for all lesions with varied sizes. However, the underestimation is more significant for the small lesion. The bias probably comes from the partial volume effects (PVEs) that are usually related to the spatial resolution. The PVEs could induce the underestimation for quantitative SPECT images, especially when the target is smaller than three times of spatial resolution [26]. This impact could directly propagate into the CRM that is derived based on SPECT images. As illuminated in Figure 3(b), the estimated change-rate is much lower than the predefined change-rate for the small-size lesion (ϕ8 mm). The estimated change-rates for middle-size lesion (ϕ16 mm) have the high linear relation with the predefined values, although the values are only nearly 70% of the real values. It is concluded that the change-rate could be significantly underestimated comparing with the real value, when the sizes of recovered regions are below three times of spatial resolution. Therefore, this underestimation should be kept in mind for clinical applications.

For the clinical applications, the CAA-CRM approach could be used to define the recovered regions and derived quantitative indexes to measure recovery levels. In this study, the thresholding and clustering method is applied to automatically derive the recovered regions. The chosen threshold of change-rate could be the major impact factor in delineating the recovered regions. Furthermore, it could influence the estimations of quantitative indexes from recovered regions [27, 28]. The experimental threshold is set as 20% in the clinical applications. It is chosen based on the results of the computer simulations. As shown in Figure 2, the image quality indexes for the change-rates of 10% and 20% are both much poorer than the other change-rates regardless of the lesion size. The CRM would not accurately reflect the change-rate lower than 20%. This indicates that the significant distortions may exist in the CRM for the voxels with lower change-rates. In this case, the lowest limit for available estimated range of change-rate is required to eliminate the turbulences from

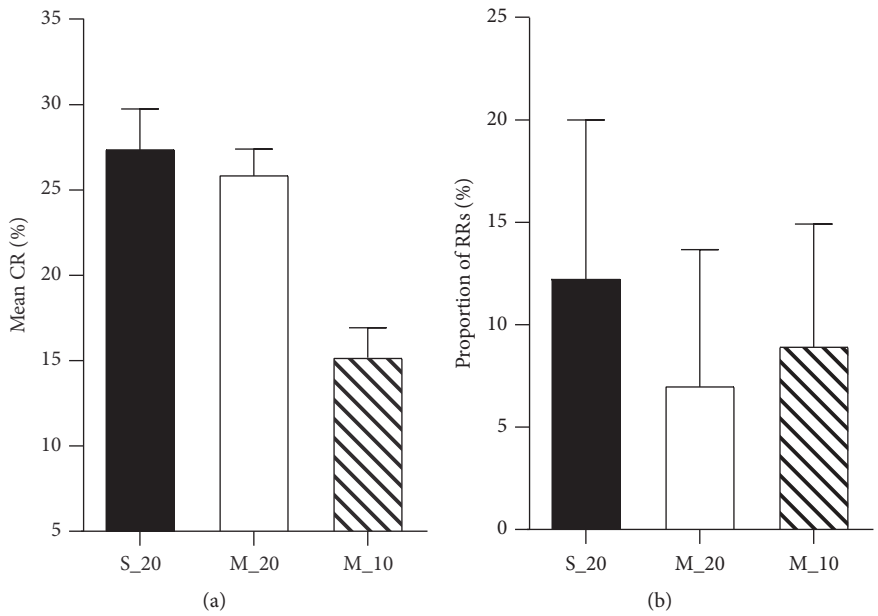


Figure 6: The bar graphs (mean + SD) for comparisons of quantitative indexes for severe and moderate recovery groups with different thresholds (S_20: severe recovery group with the threshold of 20%, M_20: moderate recovery group with the threshold of 20%, and M_10: moderate recovery group with the threshold of 10%). (a) The mean change-rate (CR) estimates of the recovered regions; (b) the proportion of recovered regions (RRs) to the corresponding lobes.

lower change-rate estimates. Meanwhile, the threshold setting should try to retain as much information as possible in the CRM for further quantitative analysis. Hence, the experimental threshold set in this study is chosen as 20% rather than 10%. The thresholds could also be able to reset for different conditions of varied clinical applications.

In the clinical applications, considering the concordance of the CAA-CRM approach in the detection of recovery regions, the false positive (12/99) and false negative (7/99) could be found. The false positive might be caused by using the uniform 20% threshold which might result in the detection of not only the main recovered regions (validated by visual inspections) but also the other minor recovered regions. However, the minor recovered regions might be ignored by the physicians. On the other hand, the uniform threshold could easily introduce the small changed regions (<120 voxels) for a certain case. The small changed regions could be removed or even miss the main recovered regions. This may lead to the false negative. In this condition, the threshold should be adjusted carefully to balance the false positive and false negative for detecting the recovered regions. Moreover, the chosen threshold could further impact on the estimations of the mean change-rate and the proportion of recovered regions to the brain lobes. As shown in Figure 6, when a threshold of 10% instead of 20% is applied in the moderate group, the values of mean change-rate are decreased sharply. Meanwhile, the values of proportion of recovered regions to the brain lobes have increased. From the comparisons, the mean change-rate is negatively related to the proportion of recovered region. Therefore, these two quantitative indexes should be used together to scale the recovery level in treatment monitoring. To an extent the maximum change-rate could not be affected by the chosen threshold in the CAA-CRM approach, so that it becomes an important quantitative index in treatment monitoring.

Because all the algorithms used in the proposed CAA-CRM approach totally rely on image contents, this approach could be extended to analyze other types of SPECT brain images, such as ^{99m}Tc-HMPAO SPECT brain images. The obtained change-rate map could also illuminate the global changes for reflecting the response to treatment. The derived quantitative indexes have the potential to quantify the recovery levels.

5. Conclusion

In this study, a CAA-CRM approach has been introduced to evaluate the longitudinal ^{99m}Tc-ECD SPECT images in treatment monitoring. This approach can provide change-rate map as a parametric image to reflect the changes of rCBF. Computer simulations show the efficacy of the proposed approach in detecting the recovered regions and in quantifying the change-rates for the lesions larger than spatial resolution. In clinical applications, this method is used to assess the treatment of ICA stenting. The results demonstrate that the quantitative indexes derived from CRM are all significantly different among the groups and highly correlated with the experienced clinical diagnosis.

In conclusion, the CAA-CRM approach has the advantages of directly illuminating the global recovery and conveniently quantifying the recovery levels. It could be helpful in

improving the efficiency and accuracy of therapy evaluations using SPECT brain images in clinical routines.

Competing Interests

The authors declare no conflict of interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 81201146, to Xiujuan Zheng, no. 51627807, to Qiu Huang, and no. 81471708, to Shaoli Song). The authors would like to thank the clinicians and the technicians of Renji Hospital for their help with data collection and the experimental procedures. The authors would also like to thank the reviewers of this paper for their constructive suggestions.

References

[1] Ö. Kapucu, F. Nobili, A. Varrone et al., "EANM procedure guideline for brain perfusion SPECT using 99mTc-labelled radiopharmaceuticals, version 2," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 36, no. 12, pp. 2093–2102, 2093.

[2] Y.-A. Chung, O. J. Hyun, J.-Y. Kim, K.-J. Kim, and K.-J. Ahn, "Hypoperfusion and ischemia in cerebral amyloid angiopathy documented by ^{99m}Tc-ECD brain perfusion SPECT," *Journal of Nuclear Medicine*, vol. 50, no. 12, pp. 1969–1974, 2009.

[3] M. Farhadi, S. Mahmoudian, F. Saddadi et al., "Functional brain abnormalities localized in 55 chronic tinnitus patients: fusion of SPECT coincidence imaging and MRI," *Journal of Cerebral Blood Flow and Metabolism*, vol. 30, no. 4, pp. 864–870, 2010.

[4] G. Uruma, K. Hashimoto, and M. Abo, "A new method for evaluation of mild traumatic brain injury with neuropsychological impairment using statistical imaging analysis for Tc-ECD SPECT," *Annals of Nuclear Medicine*, vol. 27, no. 3, pp. 187–202, 2013.

[5] S. T. Donta, R. B. Noto, and J. A. Vento, "SPECT brain imaging in chronic lyme disease," *Clinical Nuclear Medicine*, vol. 37, no. 9, pp. e219–e222, 2012.

[6] L. S. Zuckier and O. O. Sogbein, "Brain perfusion studies in the evaluation of acute neurologic abnormalities," *Seminars in Nuclear Medicine*, vol. 43, no. 2, pp. 129–138, 2013.

[7] J. M. Mountz, H.-G. Liu, and G. Deutsch, "Neuroimaging in cerebrovascular disorders: measurement of cerebral physiology after stroke and assessment of stroke recovery," *Seminars in Nuclear Medicine*, vol. 33, no. 1, pp. 56–76, 2003.

[8] D. G. Amen, D. Highum, R. Licata et al., "Specific ways brain spect imaging enhances clinical psychiatric practice," *Journal of Psychoactive Drugs*, vol. 44, no. 2, pp. 96–106, 2012.

[9] S. Sestini, A. Pupi, F. Ammannati et al., "Predictive potential of pre-operative functional neuroimaging in patients treated with subthalamic stimulation," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 1, pp. 12–22, 2010.

[10] K. Chen, J. B. S. Langbaum, A. S. Fleisher et al., "Twelve-month metabolic declines in probable Alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the Alzheimer's Disease Neuroimaging Initiative," *NeuroImage*, vol. 51, no. 2, pp. 654–664, 2010.

[11] N. J. Kazemi, G. A. Worrell, S. M. Stead et al., "Ictal SPECT statistical parametric mapping in temporal lobe epilepsy surgery," *Neurology*, vol. 74, no. 1, pp. 70–76, 2010.

[12] L. Wichert-Ana, P. M. De Azevedo-Marques, L. F. Oliveira et al., "Ictal technetium-99 m ethyl cysteinate dimer single-photon emission tomographic findings in epileptic patients with polymicrogyria syndromes: a Subtraction of ictal-interictal SPECT coregistered to MRI study," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 35, no. 6, pp. 1159–1170, 2008.

[13] H. W. Lee, S. B. Hong, and W. S. Tae, "Opposite ictal perfusion patterns of subtracted SPECT: hyperperfusion and hypoperfusion," *Brain*, vol. 123, no. 10, pp. 2150–2159, 2000.

[14] G. I. Varghese, M. J. Purcaro, J. E. Motelow et al., "Clinical use of ictal SPECT in secondarily generalized tonic–clonic seizures," *Brain*, vol. 132, no. 8, pp. 2102–2113, 2009.

[15] D. H. Lewis, J. P. Bluestone, M. Savina, W. H. Zoller, E. B. Meshberg, and S. Minoshima, "Imaging cerebral activity in recovery from chronic traumatic brain injury: a preliminary report," *Journal of Neuroimaging*, vol. 16, no. 3, pp. 272–277, 2006.

[16] Y.-W. Chuang, C.-C. Hsu, Y.-F. Huang et al., "Brain perfusion SPECT in patients with Behçet's disease," *Journal of Neuroradiology*, vol. 40, no. 4, pp. 288–293, 2013.

[17] S. Miyamoto, H. Ichihashi, and K. Honda, *Algorithms for Fuzzy Clustering: Methods in C-Means Clustering with Applications*, vol. 229 of *Studies in Fuzziness and Soft Computing*, Springer, Berlin, Germany, 2008.

[18] N. Boussion, C. Houzard, K. Ostrowsky, P. Ryvlin, F. Mauguière, and L. Cinotti, "Automated detection of local normalization areas for ictal-interictal subtraction brain SPECT," *Journal of Nuclear Medicine*, vol. 43, no. 11, pp. 1419–1425, 2002.

[19] J. L. Lancaster, J. L. Summerlin, L. Rainey, C. S. Freitas, and P. T. Fox, "The Talairach Daemon a database server for talairach atlas labels," *NeuroImage*, vol. 5, no. 4, article S633, 1997.

[20] J. L. Lancaster, M. G. Woldorff, L. M. Parsons et al., "Automated Talairach Atlas labels for functional brain mapping," *Human Brain Mapping*, vol. 10, no. 3, pp. 120–131, 2000.

[21] A. C. Evans, D. L. Collins, and B. Milner, "An MRI-based stereotactic atlas from 250 young normal subjects," *Society for Neuroscience Abstracts*, vol. 18, no. 179, p. 408, 1992.

[22] J. A. Maldjian, P. J. Laurienti, R. A. Kraft, and J. H. Burdette, "An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets," *NeuroImage*, vol. 19, no. 3, pp. 1233–1239, 2003.

[23] E. J. Hoffman, P. D. Cutler, T. M. Guerrero, W. M. Digby, and J. C. Mazziotta, "Assessment of accuracy of PET utilizing a 3-D phantom to simulate the activity distribution of [^{18}F]fluorodeoxyglucose uptake in the human brain," *Journal of Cerebral Blood Flow & Metabolism*, vol. 11, supplement 1, pp. A17–A25, 1991.

[24] J. Ye, X. Song, Z. Zhao, A. J. Da Silva, J. S. Wiener, and L. Shao, "Iterative SPECT reconstruction using matched filtering for improved image quality," in *Proceedings of the Nuclear Science Symposium Conference Record (NSS/MIC '10)*, pp. 2285–2287, IEEE, Knoxville, Tenn, USA, November 2006.

[25] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.

[26] D. L. Bailey and K. P. Willowson, "An evidence-based review of quantitative SPECT imaging and potential clinical applications," *Journal of Nuclear Medicine*, vol. 54, no. 1, pp. 83–89, 2013.

[27] M. Vanderhoek, S. B. Perlman, and R. Jeraj, "Impact of the definition of peak standardized uptake value on quantification of treatment response," *Journal of Nuclear Medicine*, vol. 53, no. 1, pp. 4–11, 2012.

[28] P. Tylski, S. Stute, N. Grotus et al., "Comparative assessment of methods for estimating tumor volume and standardized uptake value in ^{18}F-FDG PET," *Journal of Nuclear Medicine*, vol. 51, no. 2, pp. 268–276, 2010.

Hindawi Publishing Corporation
BioMed Research International
Volume 2017, Article ID 7961494, 9 pages
http://dx.doi.org/10.1155/2017/7961494

Research Article

Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images

Jianfeng Zhang,[1] Jiatuo Xu,[1] Xiaojuan Hu,[2] Qingguang Chen,[3] Liping Tu,[2] Jingbin Huang,[1] and Ji Cui[1]

[1]*Basic Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[2]*Shanghai Innovation Center of TCM Health Service, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[3]*Department of Endocrinology, Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*

Correspondence should be addressed to Jiatuo Xu; xjt@fudan.edu.cn

Received 26 August 2016; Revised 24 November 2016; Accepted 12 December 2016; Published 4 January 2017

Academic Editor: Zexuan Ji

Objective. The purpose of this research is to develop a diagnostic method of diabetes based on standardized tongue image using support vector machine (SVM). *Methods.* Tongue images of 296 diabetic subjects and 531 nondiabetic subjects were collected by the TDA-1 digital tongue instrument. Tongue body and tongue coating were separated by the division-merging method and chrominance-threshold method. With extracted color and texture features of the tongue image as input variables, the diagnostic model of diabetes with SVM was trained. After optimizing the combination of SVM kernel parameters and input variables, the influences of the combinations on the model were analyzed. *Results*. After normalizing parameters of tongue images, the accuracy rate of diabetes predication was increased from 77.83% to 78.77%. The accuracy rate and area under curve (AUC) were not reduced after reducing the dimensions of tongue features with principal component analysis (PCA), while substantially saving the training time. During the training for selecting SVM parameters by genetic algorithm (GA), the accuracy rate of cross-validation was grown from 72% or so to 83.06%. Finally, we compare with several state-of-the-art algorithms, and experimental results show that our algorithm has the best predictive accuracy. *Conclusions.* The diagnostic method of diabetes on the basis of tongue images in Traditional Chinese Medicine (TCM) is of great value, indicating the feasibility of digitalized tongue diagnosis.

1. Introduction

Resulting from a variety of factors, diabetes is a metabolic disorder mainly characterized by chronic high blood glucose. The number of patients diagnosed with diabetes has been increasing at a rapid rate worldwide. Among them, Chinese patients increase at the highest rate. The overall prevalence of diabetes in the adult population of China grew from 0.67% in 1980 to 11.6% by 2010 [1], and it greatly influences people's life.

Traditional Chinese Medicine (TCM) diagnosis is based on the information obtained from four diagnostic processes, that is, looking, listening and smelling, asking, and touching. The most common tasks are taking the pulse and inspecting the tongue [2]. Studies demonstrate that tongue images have relatively strong correlation with diabetes in TCM [3–5]. Traditionally, doctors diagnose diseases and identify patterns by inspecting, describing, and experiences. Thus, the results are easily affected by doctors' own professional skills and surrounding environment. Without objective assessment criteria, the precision of pattern identification and the repeatability of verification are unclear. Thus, a workable solution is to apply computer techniques and image processing, guided by TCM theory, to make tongue diagnosis standard, objective, and quantitative.

Fortunately, computational methods for digital image processing techniques in tongue have been developed, which achieve promising results [6–8]. Based on the results above, various machine learning methods have been used in tongue manifestation recognition or classification, such as support vector machine (SVM) [9–11], k Nearest Neighbor (k-NN)

[12, 13], Naive Bayes [11], Decision Tree [11], and Neural Network [9, 14]. Throughout all mentioned works on the inspection, the popular machine learning algorithms, such as k-NN and SVM, are still the first choice in current literature [15]. Though the identification and classification of tongue image have made certain achievements in the past researches, there still existed some issues, firstly, a standard lighting source environment is needed in the tongue image collection, and an effective method is necessary in the tongue image analysis. Moreover, for a successful SVM model, the selection of its kernel parameter and optimization of the data set is of great importance.

In this study, tongue images were collected by the TDA-1 digital tongue instrument, which can create a stable light source environment for tongue image acquisition. Research on tongue image mainly focuses on two parts: tongue body and tongue coating. The segmentation of these two parts is an important step in tongue diagnosis since it provides a premise for analyzing the color and texture. This study adopted division-merging method and chrominance-threshold method to distinguish tongue body from tongue coating and then to obtain the parameters of each part [16].

Among commonly used data mining methods, SVM [17] is widely applied due to its excellency in generalization and nonlinear function fitting, and it also presents a lot of advantages in dealing with small sample studies [18]. For a successful SVM model, kernel parameters of SVM are the most important factors affecting the prediction accuracy. Therefore, with its kernel parameters optimized by genetic algorithm (GA), we designed a genetic-based SVM (GA-SVM) model, which was adopted to establish a diagnosis model for diabetes on the basis of tongue images. In our study, techniques for collecting and analyzing information were employed to achieve interdisciplinary research and application in TCM, which will promote the construction of assessment system supported by information acquired from TCM four diagnostic methods in the future for medical treatment in communities and individual health management.

2. Materials and Methods

2.1. Subjects. This study included 827 subjects with informed consent from outpatients and physical examination centers at Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine (SHUTCM) and TCM Hospital of Baoshan Area in Shanghai, from July 2013 to June 2015. Among them, 296 (159 males and 137 females, average age: 58.35 ± 12.99) were diagnosed with diabetes; 531 (191 males and 340 females, average age: 62.37 ± 8.13) had no diabetes.

2.2. Inclusion and Exclusion Criteria. Inclusion criteria for diabetes group signed with informed consent were from World Health Organization (WHO) in 1999 [19], which include (1) symptoms of high blood sugar and random plasma glucose ≥11.1 mmol/L (200 mg/dL); (2) fasting plasma glucose level ≥7.0 mmol/L (126 mg/dL); (3) plasma glucose ≥ 11.1 mmol/L (200 mg/dL) two hours after a 75 g oral glucose load as in a glucose tolerance test.

Figure 1: TDA-1 tongue instrument.

Exclusion criteria include (1) those diagnosed with other severe diseases such as tumor and diseases of immune and hematological systems; (2) those who cannot make a clear description or cooperate with the imaging collection due to mental disorders; (3) those who refuse to sign informed consent.

2.3. Tongue Image Collection and Analysis Methods

2.3.1. Collection Instrument. Developed by the research team handling intelligent processing of TCM diagnosis information in SHUTCM, the TDA-1 digital tongue instrument was applied to collect images. This apparatus (Figure 1) is made up with an Eolane digital camera, a LED3 light, a removable collection ring, a circuit board, and a hand. The main technical parameters are Charge-coupled device: Eolane A12; light: cold white LED light; color temperature: 6466 K; luminance: 23541 ux [20].

2.3.2. Collection Methods. The TDA-1 tongue instrument was used to collect tongue images in the morning before breakfast. With build-in light on, parameters were set into a manual mode with an aperture of 13, a shutter speed of 1/60 s, microlens, and no flash. Subjects were asked to take a seat, look at the front horizontally, and extend tongue with tongue tip hanging naturally at a 60-degree angle to the horizontal line. The chin was supported by the inferior margin of the collection ring so that the face was closely attached to the ring. With flat surface, 1/2 to 2/3 part of the tongue was protruded. Then, the OK button was pressed. After closing the build-in light, the collected images were checked. If these images did not conform to the requirements mentioned above, collect it again.

2.3.3. Analysis Methods. To obtain parameters of tongue images, tongue body and coating need to be separated, and the color and texture of both body and coating need

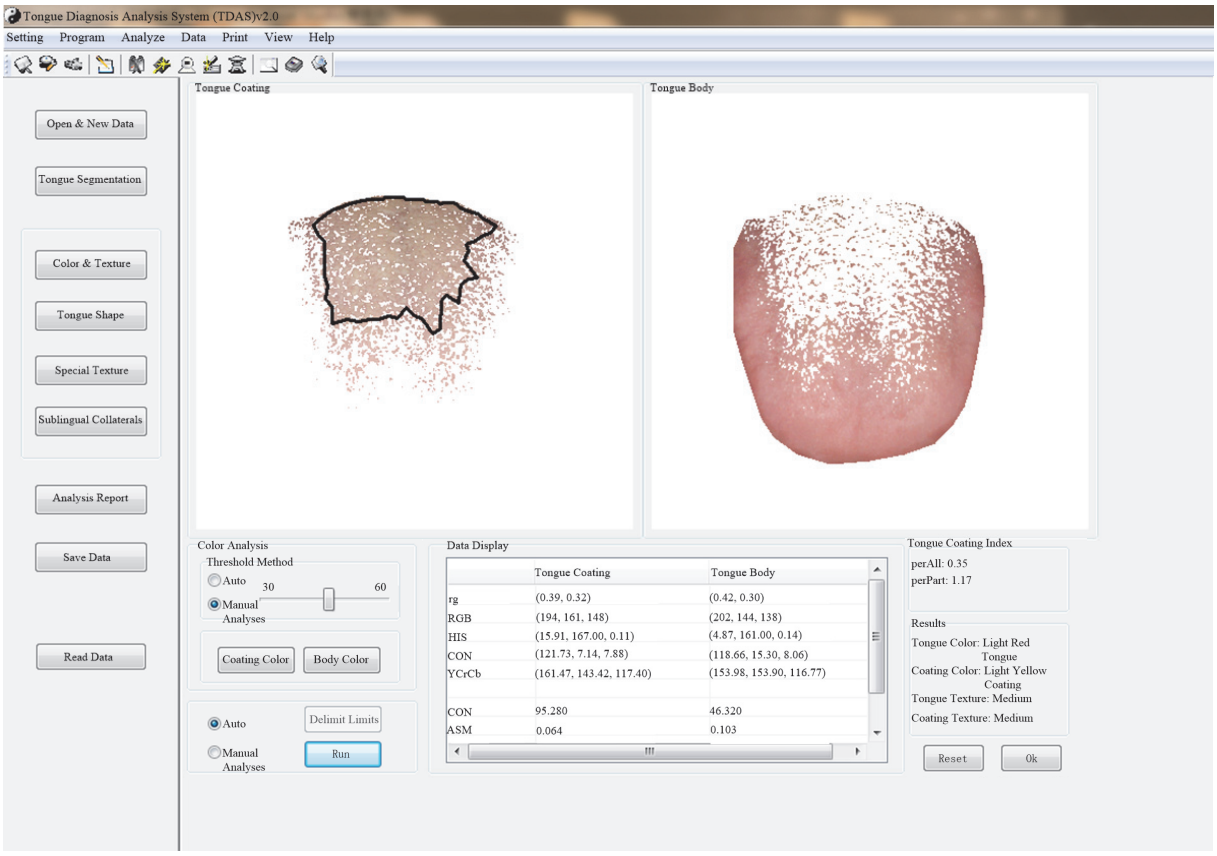


Figure 2: Tongue diagnosis analysis system.

to be analyzed and recognized [16]. The method has been developed into a Tongue Diagnosis Analysis System (TDAS) by the TCM diagnosis Intelligent Information Processing Laboratory of Shanghai University of TCM (Figure 2). The upper part of the system is the toolbar, including settings, projects, analysis, data, and print; the left side is the module of tongue manifestation analysis, including new data, tongue manifestation segmentation and results. The middle part is tongue body and tongue coating picture. Manual and automatic analyses can be used in color and texture. Because manual analysis can be based on the experience of TCM experts to choose different points and boundaries, we choose the combination of manual and automatic analyses according to experience. The data view window shows the color of the tongue body, the tongue coating, and texture parameters. The bottom right part shows the results of tongue manifestation analysis (Figure 2, light red tongue; light yellow coating).

Segmentation of Tongue Body and Coating. The identification of tongue body and coating is an important procedure in tongue diagnosis since it is the premise for analyzing the color and texture features of body and coating. The biggest distinction between tongue body and coating lies in the color: with red as the dominant hue, the color of the tongue body can be presented as light white, light red, crimson red, and purple. For tongue coating, the color can be white, yellow, grey, and black. Due to different color attributes and value ranges, division-merging algorithm based on the color of body and coating was adopted. For images with typical color of body and coating, this method achieves good results. But when tongue coating is thin, the color of tongue body overlaps with that of tongue coating. In this case, this approach fails to separate two colors. Therefore, chrominance-threshold method was adopted too. In this study, the division-merging algorithm and chrominance-threshold method were combined to separate the tongue body and tongue coating. Detailed algorithm was referred to the relevant literature [16]. The tongue images after segmentation were shown in Figure 3.

Acquisition of Features in Color and Texture. After separating the area of tongue body and coating, RGB color values of the pixels in tongue body and coating were calculated and then the values of total pixels were averaged. Considering the visualization of color and the feasibility and practicability of classification, we transformed RGB chroma space into LAB and HIS [21]. The texture of tongue contains rough tongue, tender tongue, greasy coating, and rough coating. Among them, tender tongue body and greasy coating are fine and smooth. The changes of texture mainly lie in the variations of gray level. Thus, this study applied gray scale differential algorithm to describe the texture of body and coating. The obtained parameters included contrast (CON), angle second moments (ASM), entropy (ENT), and mean [22].

Figure 3: The segmentation of tongue body and coating.

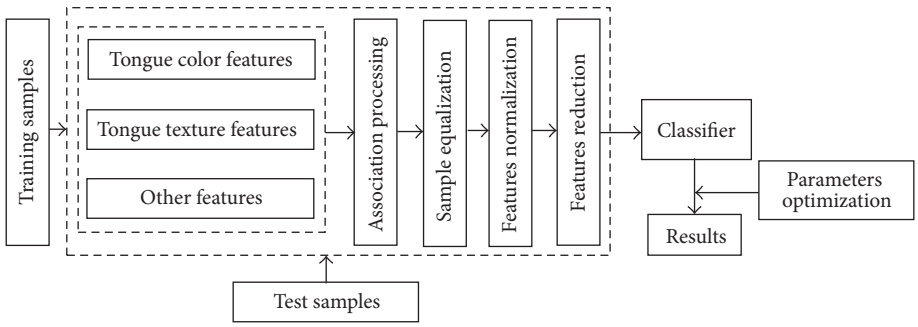


Figure 4: Map of classification model.

2.4. Study Design and Setting. The data, mainly including color feature and texture feature, were derived from tongue images of diabetes group and nondiabetes group. These feature parameters were input as independent variables in the preprocessing. Whether the subject has diabetes or not (dependent variable) was considered as the classified variable. 80% specimens were used for training, while 20% for test. The model is illustrated in Figure 4.

2.4.1. Sample Equalization. The sample size of diabetes group and nondiabetes group differs a lot. To avoid the influences of unequal sample size on classification model, Synthetic Minority Oversampling Technique (SMOTE) was adopted to equalize samples. SMOTE, proposed by Chawla et al. in *Artificial Intelligence* in 2002 [23], is a solution based on oversampling. In this study, the equalization was achieved by DMwR package in R Language.

2.4.2. Feature Normalization. Due to the differences of feature parameters in magnitude orders, to eliminate the negative effects of these differences, the value ranges were scaled and mapped into the range between −1 and 1.

2.4.3. Dimension Reduction of Features. The increase in the number of variables will make the SVM more complicated. In addition, variables may have relevance between each other. Thus, in our study, principal component analysis (PCA), the classic approach to reduce dimensions, was adopted to process the acquired raw features of tongue images so that the information integrity can be maintained as much as possible in the process of dimension reduction.

2.4.4. Optimization of Kernel Parameters for SVM by GA. GA, first proposed by American professor Holland in 1962 [24], is a computational model for optimization with parallel search that simulates genetic mechanism and biological evolution in nature. In the study, the penalty parameter c and kernel function parameter g were optimized by GA. The accuracy of training sample prediction was considered as the fitness function value of GA. The process of algorithm is shown in Figure 5 [25, 26].

2.4.5. Development Platform. The study was performed in MATLAB platform by the toolbox of LIBSVM-FarutoUltimate [27] that adds some auxiliary functions on the basis of LIBSVM [28].

3. Results

3.1. Results of Sample Equalization. In SMOTE, by confirming the frequency of sampling, the samples of both groups were equalized. The result is shown in Table 1.

3.2. Results of Dimension Reduction of Features. There were 23 input parameters, which include personal information (gender, age, and BIM) and parameters of tongue color

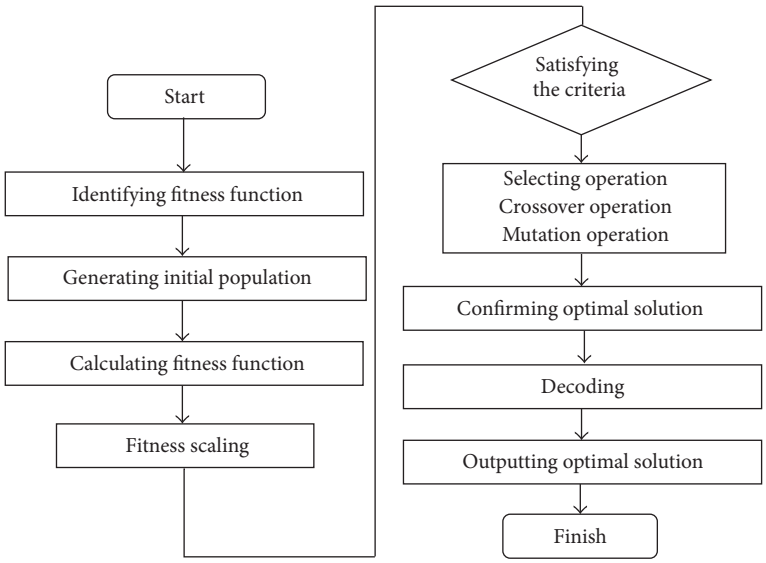


Figure 5: Flow chart of GA-SVM.

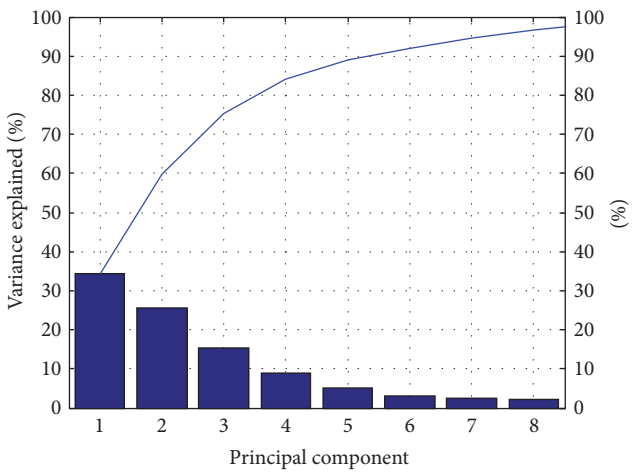


Figure 6: Result of PCA.

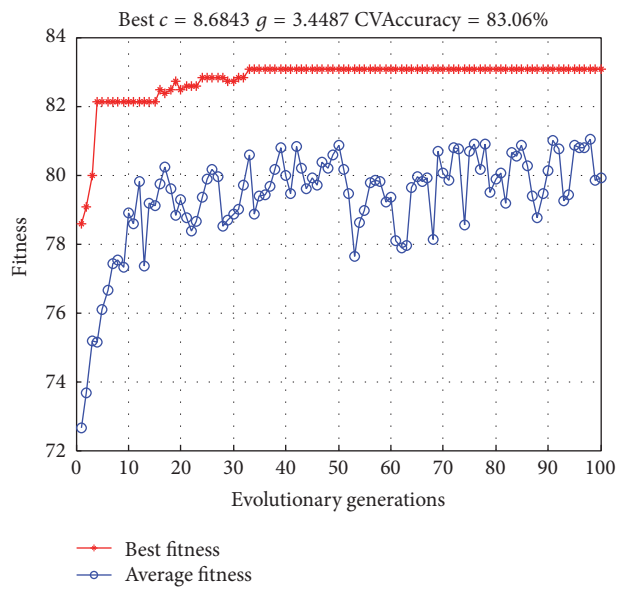


Figure 7: Fitness curve of SVM parameters optimized by GA.

Table 1: Samples before and after equalization.

Samples	Diabetes group	Nondiabetes group
Original	296	531
Equalized	531	531

and texture. PCA was applied to reduce the dimensions of raw data on the condition that the 95% information was maintained. The result is shown in Figure 6.

3.3. Optimized SVM Parameters. In the training process of SVM model, the penalty parameter c and kernel function parameter g were optimized by GA. With population size set as 20, evolutionary generations as 100, and other parameters of LIBSVM toolbox as the default, the precision of sample tests with 10-fold cross-validation was considered as fitness and the accuracy of cross-validation in the training process grew from 72% or so to 83.06%, which is shown in Figure 7.

3.4. Results of Prediction with GA-SVM Model. We established three GA-SVM classifiers on different datasets which are raw data, normalized data, and normalized data after PCA, respectively. The result shown in Table 2 demonstrates that the classifier on normalize data after PCA yields a better accuracy than other two datasets, which is 1.89% higher than that of raw data at 79.72%.

A receiver operating characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifiers system. The curve is created by plotting the true

Table 2: SVM classification before and after data processing.

Datasets	Accuracy of cross-validation (%)	Accuracy of training samples (%)	Accuracy of test samples (%)	Running time (s)
Raw data	81.65	100.00	77.83	817.86
Normalized data	84.35	99.53	78.77	747.40
Normalized data with PCA	83.06	99.88	79.72	465.52

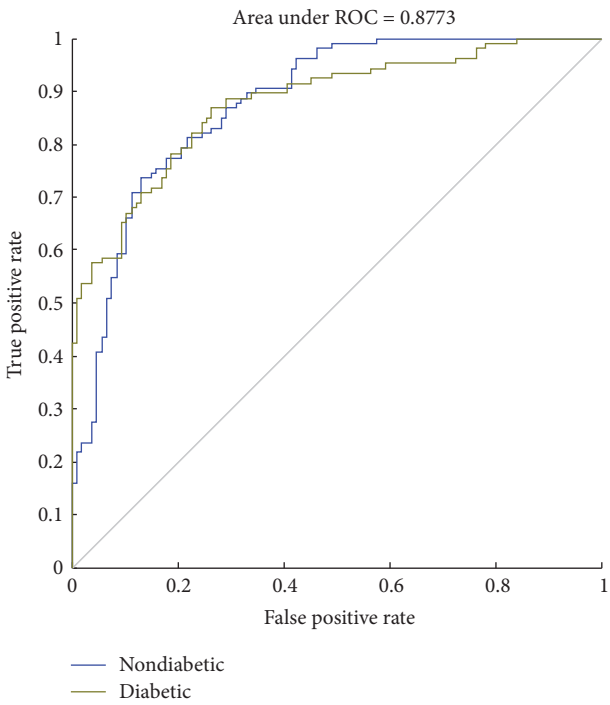


Figure 8: The ROC curve (raw data).

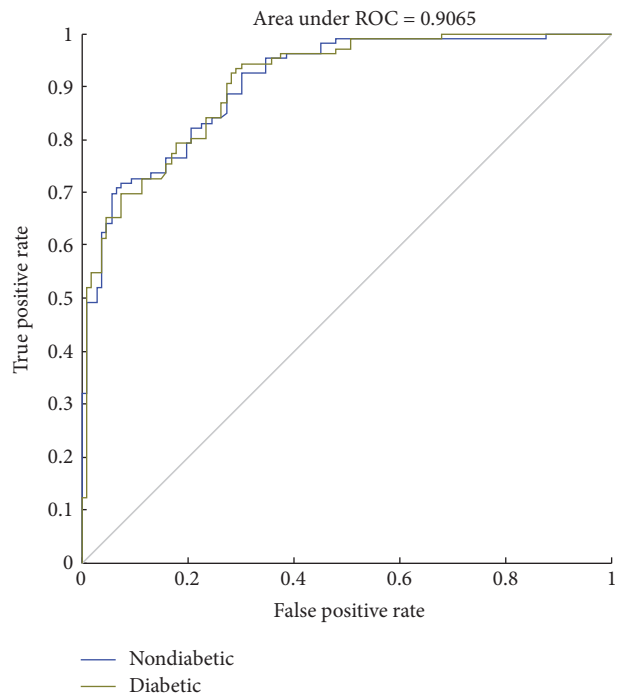


Figure 9: The ROC curve (normalized data).

positive rate (TPR) against the false positive rate (FPR) at various threshold settings. Sensitivity is also known as TPR, which means the probability that true judgement is made for having diabetes. Specificity is equal to true negative rate, which means the probability that true judgement is made for not having the disease. The area under ROC curve (AUC) is most commonly used as precision index. When the sensitivity and specificity reached 1, the area under ROC curve is obtaining a perfect precision. The best possible prediction method would generate a point on the upper left corner (0, 1) in ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives) [29].

In this study, we used sensitivity, specificity, ROC, and AUC to assess the performance of classifiers. As shown in Figures 8–10, the ROC curves in three figures are for the classifiers using different datasets. Blue curves represent the ROC curve of the nondiabetic class, while other curves represent the ROC curve of the diabetic class. The AUC values for the classifiers using different datasets are 0.8773, 0.9065, and 0.9037, which indicates that the classifier is effective in distinguishing the two classes of objects. We can know from Table 3 that the performance of third classifier (normalized data with PCA) has higher sensitivity and specificity than that of the first two (raw data and normalized data). Although the AUC of the third classifier is lower than the second by 0.3%, it substantially saves the average training time as shown in Table 2.

Table 3: Specificity, sensitivity, and AUC of SVM classification before and after data processing.

Datasets	Specificity (%)	Sensitivity (%)	AUC
Raw data	81.05	75.21	0.8773
Normalized data	82.80	75.63	0.9065
Normalized data with PCA	83.16	76.92	0.9037

3.5. Comparison with Other Algorithms. In order to evaluate efficacy for established GA-SVM model, three distinct prediction models, k-NN, Naive Bayes, and Backpropagation Neural Network (BP-NN), were employed to compare with GA-SVM model. The three classification model using k-NN, Naive Bayes, and BP-NN methods are established in MATLAB. As shown in Table 4, accuracy, specificity, and AUC of GA-SVM models are higher than other algorithms,

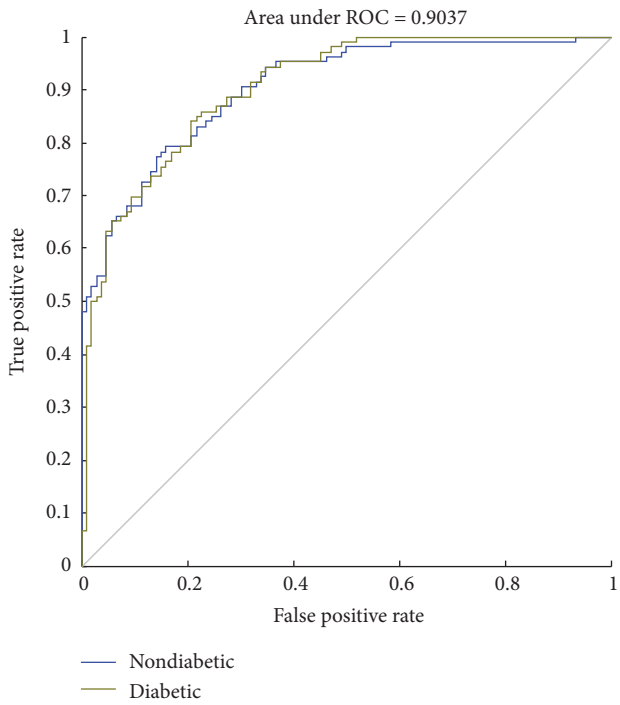


Figure 10: The ROC curve (normalized data with PCA).

Table 4: Result compared with other algorithms.

Algorithms	Accuracy (%)	Specificity (%)	Sensitivity (%)	AUC
k-NN	78.77	80.18	77.36	0.8471
Naive Bayes	75.94	78.30	73.58	0.8248
BP-NN	75.00	73.58	76.42	0.8285
GA-SVM	79.72	83.16	76.92	0.9037

except for sensitivity a little lower than k-NN. Combined with the results above, it can be concluded that our algorithm has a better established classification model for tongue manifestation.

4. Discussion

Over the past 3000 years, tongue diagnosis has been proved to be one of the most valuable and the most extensively applied TCM diagnostic approaches in clinical practice. The color, moisture, size, shape, and texture of tongue reveal the overall health condition and dysfunctions of specific organs. Tongue color and coating have long been key parameters in differentiating diseases. In this research, we took photos of tongues with the TDA-1 tongue instrument, which segmented the tongue body and coating. After obtaining color parameters of RGB, HIS, and LAB and texture parameters of CON, ASM, ENT, and MEAN, with these data, a diabetes diagnosis model was established on the basis of SVM.

Sample imbalance is a problem that must be faced in many computational problems in medical research. For unbalanced samples between two groups, SMOTE was adopted to equalize the data. Due to the differences of feature parameters in magnitude orders, to eliminate the negative effects of these differences, the value ranges of input variables were scaled and mapped into the range between −1 and 1. After processing, the accuracy of test samples was slightly increased from 77.83% to 78.77%, and the AUC was increased from 0.8773 to 0.9065.

In addition, in the training process of SVM model, the penalty parameter c and kernel function parameter g in SVM model were optimized by GA. As shown in Figure 7, the accuracy of cross-validation in training process grew from 72% or so to 83.06%, indicating the significance of parameter optimization in improving the precision of classification.

To handle multiple input variables, PCA was adopted to decrease the number of variables from 29 to 8 on the condition that 95% information was retained. The results show the accuracy of classification was not reduced and it substantially saved the training time.

With standardized tongue image parameters, we developed a novel model to search the optimal values of SVM parameters, to increase the accuracy of prediction. In order to evaluate efficacy for the established GA-SVM model, k-NN, Naive Bayes, and BP-NN model were applied to our datasets. From Table 4, the results show that the GA-SVM model performs the best, implying that the hybrid system has a high potential to dramatically increase the predictive accuracy when integrating GA with traditional SVM model. As shown in Figure 7, the accuracy of cross-validation in training process has greatly increased after SVM parameters optimized by GA. The diabetes is of varied TCM syndromes for its clinical manifestations, such as the deficiency of qi and yin syndrome, yin-deficiency and fire-hyperactivity syndrome, and different syndromes manifest with different tongue image features. However it is difficult for the traditional statistical methods to identify the diabetes through its tongue image automatically, so it is essential to find a method which is suitable for the diagnosis of diabetes via its tongue images. By comparison with other algorithms and internal validation of the model, it is indicated that the SVM classification model we established had a fair effect in this study.

5. Conclusion

In this paper, with tongue images, SVM was used to establish the classification model for diabetes, which achieves good classification results. It indicates the feasibility of using the information science method to carry out TCM diagnosis. Data preprocessing and parameter optimization directly impact the results. Feature dimension reduction is a double-edged sword. On the one hand, it can accelerate the training speed and avoid overfitting; on the other hand, it may cause the loss of useful information. GA can find the optimal option without going through the whole search space and can also be used for feature selection in other studies.

With information collecting and analyzing techniques, this interdisciplinary study researches on the informatization of TCM and its application and provides a reference for designing more effective data analysis and processing algorithms. In future researches, other pieces of TCM diagnostic

information can be integrated to improve the precision of classification.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (no. 81373556), National Science and Technology Support Program during the Twelfth Five-Year Plan Period (no. 2012BAI37B06), "Shu Guang" Project Supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation (no. 12SG36), and Basic Subject Innovation Fund Supported by Shanghai Municipal Education Commission (no. A2-P1600317).

References

[1] X. Zhao, Z. Zhen, J. Guo et al., "Assessment of the reporting quality of placebo-controlled randomized trials on the treatment of Type 2 diabetes with traditional Chinese medicine in mainland China: a PRISMA-compliant systematic review," *Medicine*, vol. 95, no. 3, Article ID e2522, 9 pages, 2016.

[2] R. Kanawong, T. Obafemi-Ajayi, T. Ma, D. Xu, S. Li, and Y. Duan, "Automated tongue feature extraction for ZHENG classification in Traditional Chinese Medicine," *Evidence-based Complementary and Alternative Medicine*, vol. 2012, Article ID 912852, 14 pages, 2012.

[3] J. Xu, J. T. Xu, Y. H. Zhu et al., "Exploration on correlation between glycometabolism and digital tongue picture in diabetic patients," *Shanghai Journal of Traditional Chinese Medicine*, vol. 48, no. 11, pp. 11–17, 2014.

[4] H. He, S. Yan, L. Yang et al., "Correlation of tongue and laboratory parameters in 5930 cases of type 2 diabetes," *Journal of Traditional Chinese Medicine*, vol. 54, no. 23, pp. 2031–2034, 2013.

[5] P.-C. Hsu, Y.-C. Huang, J. Y. Chiang, H.-H. Chang, P.-Y. Liao, and L.-C. Lo, "The association between arterial stiffness and tongue manifestations of blood stasis in patients with type 2 diabetes," *BMC Complementary and Alternative Medicine*, vol. 16, no. 1, article 324, 7 pages, 2016.

[6] J. Ning, D. Zhang, C. Wu, and F. Yue, "Automatic tongue image segmentation based on gradient vector flow and region merging," *Neural Computing and Applications*, vol. 21, no. 8, pp. 1819–1826, 2012.

[7] M. Zhu, J. Du, and C. Ding, "A comparative study of contemporary color tongue image extraction methods based on HSI," *International Journal of Biomedical Imaging*, vol. 2014, Article ID 534507, 2014.

[8] N. D. Kamarudin, C. Y. Ooi, T. Kawanabe, and X. Mi, "Tongue's substance and coating recognition analysis using HSV color threshold in tongue diagnosis," in *SPIE Proceedings of the 1st International Workshop on Pattern Recognition*, Tokyo, Japan, May 2016.

[9] R. Kanawong, T. Obafemi-Ajayi, J. Yu, D. Xu, S. Li, and Y. Duan, "ZHENG classification in Traditional Chinese Medicine based on modified specular-free tongue images," in *Proceedings of the 2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '12)*, pp. 288–294, IEEE, Philadelphia, Pa, USA, October 2012.

[10] M.-C. Hu, M.-H. Cheng, and K.-C. Lan, "Color correction parameter estimation on the smartphone and its application to automatic tongue diagnosis," *Journal of Medical Systems*, vol. 40, no. 1, article 18, 11 pages, 2016.

[11] C. H. Siu, Y. He, and D. T. C. Thach, "Machine learning for tongue diagnosis," in *2007 6th International Conference on Information, Communications and Signal Processing, ICICS*, sgp, December 2007.

[12] B. Zhang, X. Wang, J. You, and D. Zhang, "Tongue color analysis for medical application," *Evidence-based Complementary and Alternative Medicine*, vol. 2013, Article ID 264742, 11 pages, 2013.

[13] B. Huang and N. Li, "Pixel based tongue color analysis," in *Medical Biometrics*, vol. 4901 of *Lecture Notes in Computer Science*, pp. 282–289, Springer, Berlin, Germany, 2007.

[14] L. Zhi, D. Zhang, J.-Q. Yan, Q.-L. Li, and Q.-L. Tang, "Classification of hyperspectral medical tongue images for tongue diagnosis," *Computerized Medical Imaging and Graphics*, vol. 31, no. 8, pp. 672–678, 2007.

[15] C. Zhao, G.-Z. Li, C. Wang, and J. Niu, "Advances in patient classification for traditional chinese medicine: a machine learning perspective," *Evidence-based Complementary and Alternative Medicine*, vol. 2015, Article ID 376716, 18 pages, 2015.

[16] J. Xu, L. Tu, Z. Zhang, L. Zhang, and C. Zhou, "The region partition of quality and coating for tongue image based on color image segmentation method," in *Proceedings of the IEEE International Symposium on IT in Medicine and Education (ITME '08)*, pp. 817–821, Xiamen, China, December 2008.

[17] V. N. Vapnik, *The nature of statistical learning theory*, Springer-Verlag, New York, NY, USA, 1995.

[18] W. Liu, J. Wang, and C. Zhang, "Study of bioelectrical impedance analysis methods for visceral fat estimation using SVM," *Journal of Electronic Measurement And Instrument*, vol. 25, no. 7, pp. 648–653, 2011.

[19] World Health Organization, *Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications Report of A who Consultation. Part 1 Diagnosis and Classification of Diabetes Mellitus*, World Health Organization, Geneva, Switzerland, 1999.

[20] J. T. Xu, Z. F. Zhang, and L. P. Tu, "Miniaturized handheld tongue picture acquisition device," CN. 201020543888.0, 2010.

[21] J. T. Xu, C. L. Zhou, Z. F. Zhang et al., "Computerized analysis and recognition of tongue and its coating color in tongue diagnosis," *Journal of Shanghai University of Traditional Chinese Medicine*, vol. 18, no. 3, pp. 43–47, 2004.

[22] J. T. Xu, Y. Sun, Z. F. Zhang et al., "Analysis and discrimination of tongue texture characteristics by difference statistics," *Journal of Shanghai University of Traditional Chinese Medicine*, vol. 17, no. 3, pp. 55–58, 2003.

[23] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

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

[25] C.-H. Wu, G.-H. Tzeng, Y.-J. Goo, and W.-C. Fang, "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy," *Expert Systems with Applications*, vol. 32, no. 2, pp. 397–408, 2007.

[26] X.-L. Liu, X.-S. Zhao, F. Lu, and W.-B. Sun, "A GA-SVM based model for throwing rate prediction in the open-pit cast blasting," *Journal of the China Coal Society*, vol. 37, no. 12, pp. 1999–2005, 2012.

[27] Y. Li, "A toolbox with implements for support vector machines based on libsvm," Software, 2011, http://www.matlabsky.com.

[28] C. C. Chang and C. J. Lin, LIBSVM-a library for support vector machines, Version 3.12, https://www.csie.ntu.edu.tw/~cjlin//libsvm/.

[29] J. Jaworek-Korjakowska, "Computer-aided diagnosis of micro-malignant melanoma lesions applying support vector machines," *BioMed Research International*, vol. 2016, Article ID 4381972, 8 pages, 2016.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 3510807, 9 pages
http://dx.doi.org/10.1155/2016/3510807

Research Article

The Classification of Tongue Colors with Standardized Acquisition and ICC Profile Correction in Traditional Chinese Medicine

Zhen Qi,[1] Li-ping Tu,[1] Jing-bo Chen,[2] Xiao-juan Hu,[3] Jia-tuo Xu,[1] and Zhi-feng Zhang[1]

[1] *Department of Basic Medical College, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Pudong New Area, Shanghai 201203, China*
[2] *College of Computer Science and Technology, Jilin University, Jilin 130012, China*
[3] *Shanghai Collaborative Innovation Center of Health Service in Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Shanghai 201203, China*

Correspondence should be addressed to Jia-tuo Xu; xjt@fudan.edu.cn

Received 9 August 2016; Revised 20 October 2016; Accepted 30 October 2016

Academic Editor: Zexuan Ji

Background and Goal. The application of digital image processing techniques and machine learning methods in tongue image classification in Traditional Chinese Medicine (TCM) has been widely studied nowadays. However, it is difficult for the outcomes to generalize because of lack of color reproducibility and image standardization. Our study aims at the exploration of tongue colors classification with a standardized tongue image acquisition process and color correction. *Methods*. Three traditional Chinese medical experts are chosen to identify the selected tongue pictures taken by the TDA-1 tongue imaging device in TIFF format through ICC profile correction. Then we compare the mean value of $L^*a^*b^*$ of different tongue colors and evaluate the effect of the tongue color classification by machine learning methods. *Results*. The $L^*a^*b^*$ values of the five tongue colors are statistically different. Random forest method has a better performance than SVM in classification. SMOTE algorithm can increase classification accuracy by solving the imbalance of the varied color samples. *Conclusions*. At the premise of standardized tongue acquisition and color reproduction, preliminary objectification of tongue color classification in Traditional Chinese Medicine (TCM) is feasible.

1. Introduction

Tongue inspection, an important diagnosis tool in Traditional Chinese Medicine (TCM), is essential for clinical syndrome differentiation and therapeutic evaluation. The traditional tongue inspection in TCM mainly focuses on the visual description other than the objective quantification and as a result limits the development of diagnostic methods and techniques in TCM to a certain extent.

However, with the development of computer techniques, application of digital image processing techniques in tongue diagnosis has been widely studied, such as segmentation of the tongue image [1–3], separation of tongue substance and tongue coating [4, 5], and analysis of tongue image [6].

Recently, on the basis of the above research results, more and more researchers attempt to use machine learning methods in tongue image recognition or classification. Their research can be grouped into four categories, shape [7–9], texture [10], color [11–15], and comprehensive feature analysis of tongue images [16, 17].

The majority pays more attention to the machine learning for the fuzziness and uncertainty of tongue images due to the uneven personal experiences of observers and different illumination conditions. According to two different color measurement systems, studies on the objective classification of tongue color are mainly implemented through the visible reflection spectrum [12, 13] and the commonly used pixelwise or color space, respectively [14, 15].

There are only a few studies on former measurement system used in the tongue color classification. Most researchers try to recognize the tongue color on the basis of the pixelwise or color space. For example, Yamamoto et al. observed

that the tongue color change in different periods in the device-independent International Commission on Illumination (CIE) 1976 $L^*a^*b^*$ color space varied between individuals [18].

On the other hand, some researchers attempt to observe the difference of the tongue color identification outcomes accompanied with different medical practitioners. Oji et al. once studied the color discrimination and tongue color diagnosis in 68 traditional Japanese medical practitioners by the Farnsworth-Munsell 100 Hue test and 84 tongue images, respectively, and they found out that overall color discrimination became worse with aging [19].

Another group of Japanese researchers had chosen 10 experienced Kampo medicine physicians to identify the tongue images acquired by the DS01-B tongue color information acquisition system. They used K-means clustering algorithm to quantify tongue body and color coating information and concluded that clinically important tongue color differences in Kampo medicine can be visualized by applying machine learning to tongue images that are taken under stable conditions [15].

Machine learning methods used in tongue image classification include KNN [20, 21], Naive Bayes [22], Decision Tree [22], SVM [20, 22–25], Neutral Network [26], Graphical Models [22, 27], and Principle Component Analysis (PCA) [28]. KNN and SVM are the most commonly used methods.

Although in the past researches have made significant achievements in tongue classification and identification, there are some issues left to be discussed. First, there exists some inconsistency between the identified tongue images and the actual tongue colors observed by clinical TCM practitioners. It is because color correction or reproduction on tongue images needs to be performed. Consequently, it is difficult for the identified outcomes to generalize to the clinics. Second, tongue's objective quantization parameters vary from each research institute, making it hard to communicate between institutes. Moreover, tongue images are neither normalized nor unified. It is extremely hard to compare those models of cognition on the same platform and choose the best classification scheme.

Thus, our study would focus on the standardization of tongue collection environment [29] and color correction to overcome the above issues [30].

2. Materials and Methods

2.1. Subjects of Our Study. Since there was no available data by the literature retrieval that would indicate the effective sample size to study the classification of tongue color, our study would be carried out using tongue images sampled from First People's hospital in Taicang, Shuguang hospital, and Longhua hospital affiliated to Shanghai University of Traditional Chinese Medicine during 2011 to 2014. A total of 2230 tongue photos were collected.

2.2. Tongue Imaging Device and Methods. TDA-1 handheld tongue imaging device consists of image acquisition system, LED illuminator, and removable collecting ring. It is equipped with Charge Coupled Device as its photosensitive components, namely, a small-sized Eolane digital camera (Altek A12, China). The camera can capture color images with resolution of 2048 × 1536 pixels. We can manually adjust white balance, exposure time, exposure compensation, ISO speed, metering modes, flash mode, and so on. To obtain a more realistic color feature for the tongue image, LED illuminator (Kingbright KA-3021HVR4D1Z1S-C1-SH, Japan) with 6447 k color temperature, 98 color rendering index (RI), and 2413 Lux illumination is placed in the middle part of the telescopic cylinder. The whole system can create a stable light source environment for tongue image acquisition. In order to evenly illuminate the dorsal surface of the tongue, the concave reflective material is used in the back of telescopic cylinder to form a uniform irradiation light. The removable collecting ring can fix lower jaw position to ensure tongue is in a proper location in the process of tongue image collection. It can also keep out natural sunlight or unnecessary external source of light that will have disturbance on tongue and create a stable illumination environment; see Figure 1.

The operation procedure of TDA-1 tongue imaging device for tongue image collection is as follows. (1) The participants would have their mouth rinsed with water 10 minutes before imaging collection. (2) Sterilize the removable collecting ring with alcohol wipes, turn on the TDA-1 device and set up the parameters (manual mode, aperture of 13, shutter of 1/60 s, sunshine, macro lens, and nonflashlight), and then turn on the built-in light. (3) The participants would be asked to look at the front horizontally with a sitting position. And his/her jaw would be hold by the inside face of lower edge of the collecting ring which is close to the face area of the participants. After a short stabilization period of the light source (usually 5–10 s), the participants would be asked to open their mouth and protrude their tongue with a relaxed tongue body, flat tongue surface, and naturally pointing-down tongue tip. In the meantime, press the collection button when the tongue image in the preview window is appropriate. Three–five-minute rest would be needed for them if we need another collection.

2.3. Tongue Image Color Correction Method. In order to solve the chromatic aberration of tongue images caused by the different collection time, ICC profile correction was used in this study, and the digital ColorChecker SG was chosen. The detailed tongue image color correction based on ICC profile is as follows. First, an ICC profile is made before each batch of tongue image acquisition; the TDA-1 tongue imaging device is placed right above the digital ColorChecker SG (produced by X-rite Macbeth) in the darkroom; then take a photo of the digital ColorChecker SG and import it into the ProfileMaker software (Gretagmacbeth ProfileMaker Pro 5.0.5). Make the ICC profile of the ColorChecker SG correspond to the stable tongue acquisition light source D65. Second, import this ICC profile into Photoshop CS 4.0; the specific path is "C:\WINDOWS\system 32\spool\drivers\color." Import the tongue images that need color correction into Photoshop CS 4.0, assign the profile, and choose the inserted ICC profile for color correction of the original tongue images. To evaluate the effect of the ICC profile correction, we used the TDA-1 tongue imaging device to obtain 5 photos of the same SG color

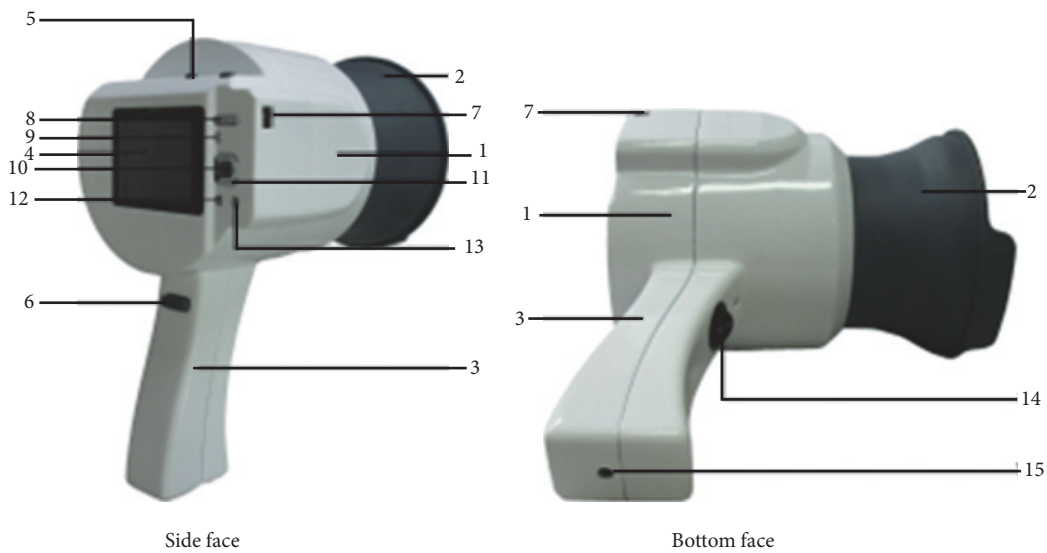


Figure 1: Structure chart of the TDA-1 tongue imaging device. Note: 1: telescopic cylinder, 2: removable collecting ring, 3: hand shank, 4: control panel, 5: Camera power, 6: camera button, 7: USB interface, 8: zoom button, 9: playback button, 10: OK button, 11: four-way navigation buttons, 12: MENU button, 13: SCENE button, 14: external power switch, and 15: external power connector.

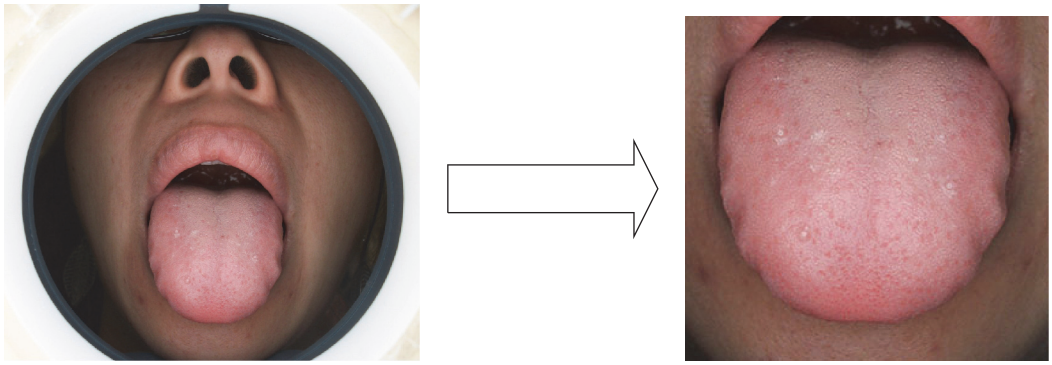
checker. And in each photo, 13 color-blocks (including D7, E7, F7, G7, H7, J7, D8, E8, F8, G8, H8, I8, and J8) corresponding to the human complexion would be chosen to carry out ICC profile correction and the Imatest software would be used to evaluate the color value difference ($\Delta E = [(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2]^{1/2}$).

2.4. Tongue Image Selection Method. First, the tongue body area of each collected tongue photos was extracted as a size of 300 × 300 in order to make the clipped tongue image close to the actual size of the tongue in clinics, and this operation was implemented by the function of tongue extraction in the software "TCM Tongue Diagnosis and Analysis System (TDAS) V2.0" (developed by our experiment team). The picture which mainly included the tongue body in its middle part was manually framed in a square, and the clipped tongue image would be automatically saved as a size of 300 × 300 by clicking the save function (Figure 2). Then look through those extracted tongue body images on the same computer monitors (Lenovo AIO has resolution of 1440 × 900); tongue pictures would be excluded if they fail to satisfy the requirements due to wrong position of tongue protrusion, underexposure or overexposure, blur, invisible tongue body covered by tongue coating, and so on. Tongue images which had petechia or ecchymosis or prick or barb on the tongue surface would also be discarded since they can be disturbance in color recognition of tongue surface.

2.5. Tongue Image Identification Method. The color of tongue images in TIFF format after color correction would be recognized by three TCM diagnosis professionals. Those that received the agreement from all three professionals would be included in this study. According to the literature's reporting, tongue color would be classified into five groups, namely, pale tongue, light red tongue, red tongue, crimson tongue, and purplish tongue (including light purple tongue and blue and purple tongue) [15, 19].

2.6. The Selection and Analysis of Characteristic Parameter for the Color of Tongue Images. Given the conformance and intuition between different color space's description on tongue color and the regular TCM diagnostic description for tongue color, $L^*a^*b^*$ color space would be used in this study. The L^* value represents the luminosity, which means the lightness of color, ranging from 0 to 100 (absolute black to absolute white). The a^* value runs in range of +127 to −128 (red to green) and the b^* value suggests the range from yellow to blue with the value between +127 and −128 (yellow to blue). Besides, in order to avoid the variation in tongue color classification index caused by different image segmentation techniques, the mean value of L^*, a^*, and b^* before or after the color correction would be acquired by manual selection of tongue tip, tongue left side, and tongue right side in each tongue image with the color sampler tool of 5 × 5 mm in Photoshop CS 4.0 (Figure 3).

2.7. Statistical Method and Data Mining Method. t-test would be used in the comparison of ΔE before and after the ICC color correction. Dunnett's T3 multiple comparison test in ANOVA would be used for the pairwise comparison of L^*, a^*, and b^* value of different color groups in SPSS18.0. Consider the idea that SVM is the most commonly used supervised machine learning method in tongue diagnosis [31], while Random forest is rarely used in this area according to the reports; besides the tongue color classification in this study is a multiclassification research; meanwhile the samples are limited and imbalanced, and the feature number for the studied images is relatively small; in order to obtain a better



Original tongue image acquired by TDA-1 device

After process of clipping

Figure 2

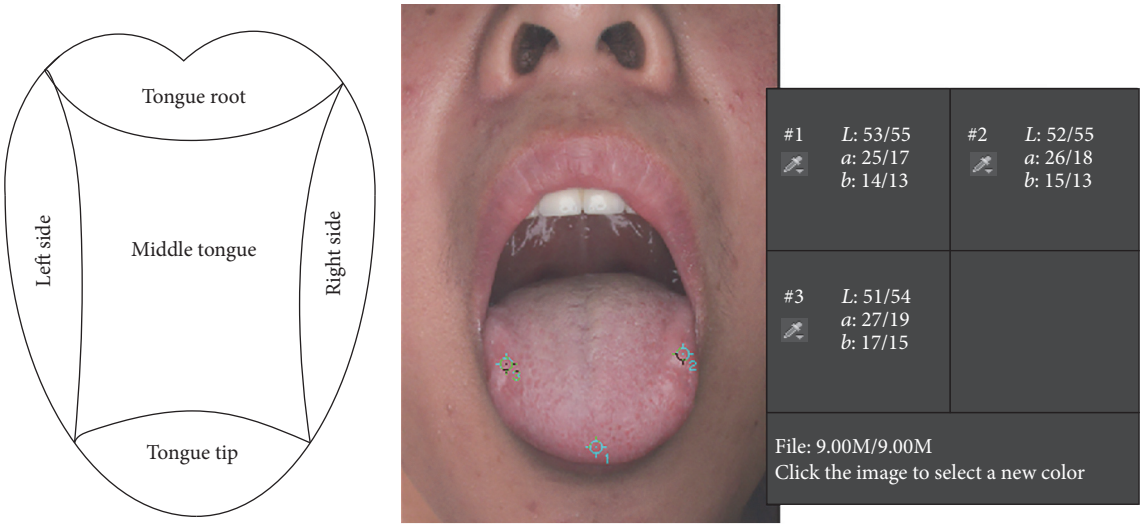


Figure 3

classification accuracy, we will use SVM and Random forest in the modeling and analytical test for the experimental data in WEKA software [32].

3. Results

3.1. Results for the Evaluation of ICC Profile Correction. The color value difference (ΔE) of the 65 color-blocks in the SG color checker before ICC profile correction is 11.79 ± 1.67, it would decrease to 4.50 ± 2.41 after ICC profile correction, and it was statistically evident between them ($p < 0.05$). As we know, the value of ΔE indicates the total difference of varied colors, and a bigger ΔE value means a larger color difference. Generally speaking, when ΔE is more than 12.0, it indicates that the two colors are entirely different; while it is less than 1.5, the difference is beyond the human vision's sensation [33]. Therefore, from the above outcomes, the color difference is obviously reduced after the ICC profile correction.

3.2. Tongue Color's Distribution in the $L^*a^*b^*$ Color Space before and after the Correction. Tongue color's distribution in the $L^*a^*b^*$ color space before and after the correction by ICC profile was demonstrated in Figures 4 and 5. Those figures indicate that the total dispersion of tongue color was reduced after correction, especially for the light red tongue's distribution in the $L^*a^*b^*$ color space. It means that the tongue colors have a more centralized and balanced distribution in the $L^*a^*b^*$ color space, so that it would be more beneficial for the TCM practitioners to recognize.

3.3. The Comparison for the Feature Parameter of Intergroup Tongue Colors. The comparison result for different tongue colors is shown in Table 1.

The results in Table 1 indicate the following aspects. (1) As for the L^* value, most colors have a statistical significance in their pairwise comparison ($p < 0.05$) except for the comparison between purplish tongue and light red tongue/pale tongue as well as the comparison between the red tongue and crimson tongue. For the L^* value suggesting the lightness of a certain color, from Table 1 we could conclude that the lightness of the red or crimson tongue is relatively lower than other three colors, and the light red tongue has lower lightness

Table 1: Comparison for the different tongue color's L^*, a^*, and b^* value ($\overline{x} \pm s$).

Group	Number ($n = 729$)	L^*	a^*	b^*
Light red	478	$53.31 \pm 5.90^{\blacktriangle\star *}$	$20.63 \pm 3.08^{\blacktriangle *\blacksquare}$	$7.6 \pm 3.93^{\star *\blacksquare}$
Pale	45	$56.18 \pm 6.06^{\bullet\star *}$	$16.93 \pm 1.75^{\bullet\star *}$	$8.69 \pm 3.48^{\star *\blacksquare}$
Crimson	35	$48.07 \pm 6.12^{\bullet\blacktriangle\blacksquare}$	$21.08 \pm 3.14^{\blacktriangle *\blacksquare}$	$2.83 \pm 3.65^{\bullet\blacktriangle *\blacksquare}$
Red	107	$48.43 \pm 6.57^{\bullet\blacktriangle\blacksquare}$	$24.21 \pm 3.43^{\bullet\blacktriangle\star\blacksquare}$	$5.32 \pm 4.14^{\bullet\blacktriangle\star\blacksquare}$
purplish	63	$52.46 \pm 6.53^{\star *}$	$16.47 \pm 2.67^{\bullet\star *}$	$-1.39 \pm 4.22^{\bullet\blacktriangle\star *}$

$^{\bullet}$Compared with light red tongue $p < 0.05$.
*Compared with crimson tongue $p < 0.05$.
$^{\blacksquare}$Compared with purplish tongue $p < 0.05$.
$^{\blacktriangle}$Compared with pale tongue $p < 0.05$.
*Compared with red tongue $p < 0.05$.

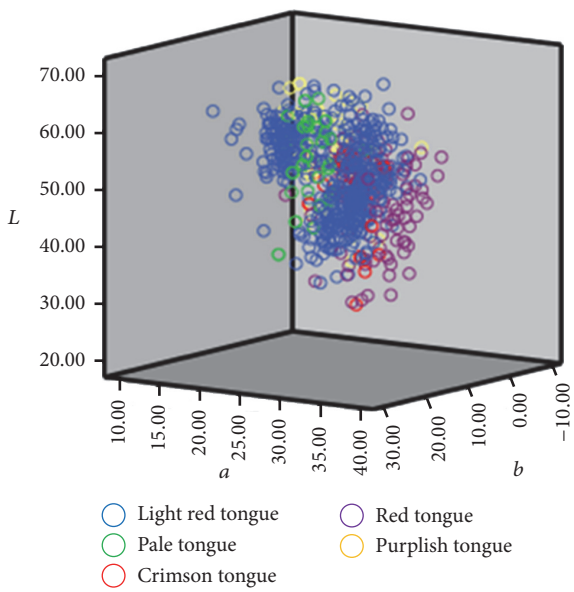


Figure 4: $L^*a^*b^*$ color space distribution for varied tongue colors (before color correction).

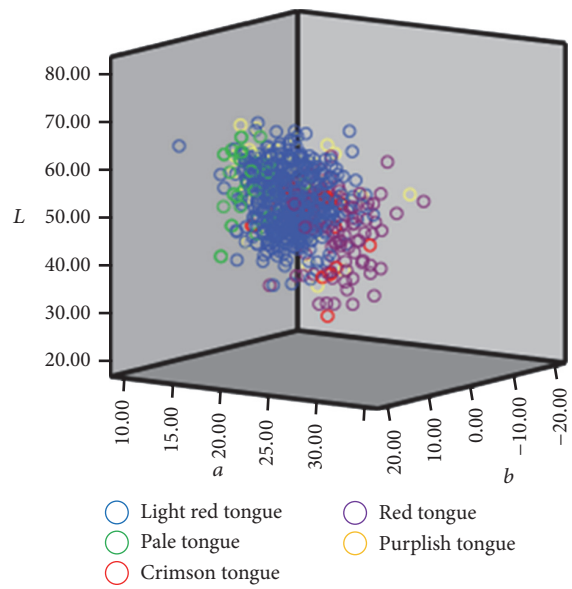


Figure 5: $L^*a^*b^*$ color space distribution for varied tongue colors (after color correction).

than the pale tongue. There was no obvious difference in lightness between the purplish tongue and the light red tongue/the pale tongue, and so did the red tongue versus the crimson tongue. (2) As for the a^* value, except for the light red tongue versus the crimson tongue and the pale tongue versus the purplish tongue, the rest colors are statistically evident for their pairwise comparison. And the a^* mean value of different tongue colors shows that the red tongue had the highest red component, next is light red tongue/crimson tongue, and pale tongue/purplish tongue is the lowest in the red component, while no obvious difference was found in both pale tongue versus purplish tongue and light red tongue versus crimson tongue for the red component. (3) As for the b^* value, each tongue color is statistically significant ($p < 0.05$) with the rest of tongue colors except the light red tongue versus the pale tongue, and the mean value of b^* of those varied colors shows that the blue component was decreased from purplish tongue to crimson tongue and then to red tongue. We can also tell that light red tongue or pale tongue has the lowest blue component and there is no obvious difference between them.

The results in Table 1 indicate that the difference of those five tongue colors is more obvious in a^* and b^* value. The distribution of different tongue colors in a^*b^* plane was demonstrated in Figure 6. Even though there exist certain overlaps at the boundaries between some colors in the a^*b^* plane color space distribution due to the absence of L^* value, general tendency can be found in the distribution of varied tongue colors in the a^*b^* plane color space, namely, from the pale tongue to the light red tongue and then to the red tongue; their color is gradually changed. In other words, it is a progress of gradual increase of a^* value and gradual decrease of b^* value, which indicates the gradual increase of red component and gradual decrease of yellow components in their colors. Besides, the purplish tongue has both the lowest a^* value and the lowest b^* value (negative value), which indicates that both the red component and yellow component of this tongue color are reduced while its blue component is augmented. For the crimson tongue, its a^*

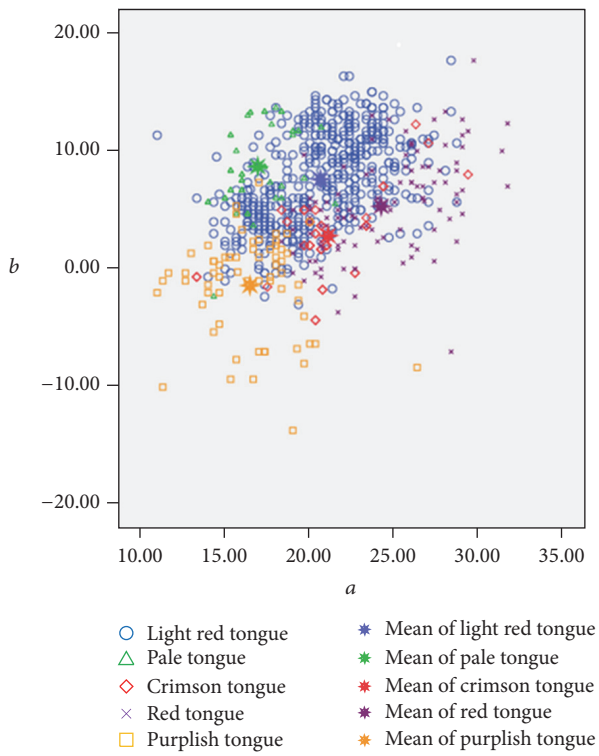


Figure 6: Distribution of respective average tongue colors in a^*b^* plane color space. Note: (1) a^* value: purplish tongue < pale tongue < light red tongue < crimson tongue < red tongue, which indicates that the red components are decreasing. (2) b^* value: purplish tongue < crimson tongue < red tongue < light red tongue < pale tongue, which indicates that the blue components are decreasing, while the yellow component is increasing.

value ranks only second to the red tongue's and its b^* value is next to the lowest (purplish tongue), which suggests a relatively higher red component and a relatively lower yellow component among the five tongue colors.

3.4. The Results for Data Mining Modeling. WEKA software was used in the data mining modeling. The selected 728 tongue images would be studied, which included 478 cases of light red tongue, 45 cases of pale tongue, 35 cases of crimson tongue, 107 cases of red tongue, and 63 cases of purplish tongue. All original data would be processed in the ".ARFF" format under the software's requirement.

In order to fulfill the evaluation of effectiveness for the established model, the studied tongue images would be divided into two groups by 10-fold cross-validation: one is the training set for the training of the best model and the other is testing set for the evaluation of the classification effect. In order to solve the problem of imbalance of tongue color samples, SMOTE algorithm would be used in this study to the sample amplification of 2390 cases. Furthermore, SVM and Random forest would be applied, respectively, for analytical test and the evaluation of the classification accuracy of model.

The results are shown in Tables 2, 3, 4, 5, and 6. Table 2 demonstrates that the overall classification accuracy for tongue color is relatively high. Random forest has a higher classification accuracy than LibSVM either before or after the sample amplification. The classification accuracy of both Random forest and SVM is increased after the sample amplification. The accuracy of SVM increases from 74.59% to 79.83%, while the accuracy of Random forest increases from 78.71% to 84.94%.

Furthermore, the classification accuracy of SVM and Random forest for different tongue colors was shown in Table 7. We can see that, without sample amplification, Random forest achieved relatively higher classification accuracy for varied tongue colors than SVM. However, after sample

Table 2: Comparison of classification accuracy.

Classification accuracy (%)	samples	LibSVM	Random forest
Before sample amplification	728	74.59	78.81
After sample amplification	2390	79.83	84.94

Table 3: Classification results of LibSVM before amplification.

	Pale	Light red	Crimson	Red	Purplish
Pale	**14**	30	0	0	1
Light red	6	**450**	0	15	7
Crimson	0	20	**2**	11	2
Red	0	57	3	**47**	0
Purplish	1	29	2	1	**30**

Table 4: Classification results of LibSVM after amplification.

	Pale	Light red	Crimson	Red	Purplish
Pale	**444**	21	1	3	9
Light red	75	**284**	26	46	47
Crimson	0	7	**404**	56	11
Red	0	21	113	**337**	7
Purplish	8	16	10	5	**439**

Table 5: Classification results of Random forest before amplification.

	Pale	Light red	Crimson	Red	Purplish
Pale	**32**	11	0	0	2
Light red	12	**420**	3	28	15
Crimson	0	4	**24**	5	2
Red	0	35	7	**62**	3
Purplish	1	23	2	2	**35**

Table 6: Classification results of Random forest after amplification.

	Pale	Light red	Crimson	Red	Purplish
Pale	**439**	30	0	1	8
Light red	41	**338**	21	43	35
Crimson	0	13	**408**	46	11
Red	0	28	45	**399**	6
Purplish	4	18	6	4	**446**

Table 7: Classification accuracy of different tongue colors (%).

Amplification	LibSVM		Random forest	
	NO	YES	NO	YES
Pale	31.1	92.9	71.1	91.8
Light red	94.1	59.4	87.9	70.7
Crimson	5.7	84.5	68.6	85.3
Red	43.9	70.5	57.9	83.5
Purplish	47.6	91.8	55.5	93.3

Table 8: Auc for different machine learning methods.

Auc	LibSVM	Random forest
Before amplification	0.69	0.89
After amplification	0.87	0.98

amplification both SVM and Random forest's classification accuracy for the light red tongue were decreased, but the whole classification accuracy for the rest tongue colors is obviously improved.

3.5. Diagnostic Value for Different Classifiers. Tongue weighted average area under the ROC curve (Auc) would be calculated for both SVM and Random forest (Table 8). The Auc for both classifiers is approximately between 0.69 and 0.98 whether before or after the amplification. This is a fairly good diagnostic value in terms of the diagnostic significance of area. And the Auc of the two classified models is 0.69 and 0.89, respectively, before the amplification but increases to 0.87 and 0.98 after the amplification. Therefore, the Auc of the Random forest is higher than LibSVM no matter whether the samples were amplified or not. Besides, the Auc of the two classification model is largely increased after the sample amplification. Combined with the accuracy above, it can be concluded that the Random forest has a better established classification model for the tongue, and the sample amplification can solve the imbalance of different tongue colors effectively. Therefore, our method has great diagnostic significance in clinics.

4. Discussion

Along with the achievements obtained in the digitization and objectification of tongue image diagnosis in TCM, many other researches about prediction of health state or abnormity of physical indicators in western medicine have been done recently [20, 22–25, 28, 34, 35]. Though they have successfully set up certain connection between the objective tongue feature parameter and the health state, it is difficult for them to be accepted and admitted by the TCM clinicians for their abstraction in the tongue diagnosis. Therefore, our study successfully establishes the connection between the visualized tongue color diagnosed by TCM experts and the abstract objective feature parameters through the exploration on the classification of different tongue colors in TCM.

Tongue color, generally, is clinically classified into categories as pale, light red, red, crimson, blue, purple, and others. For there is no strict boundaries among them, the color recognition of each observer is also varied. In this study, tongue color is classified into light red tongue, pale tongue, red tongue, crimson tongue, and purplish tongue on the basis of relevant literature report.

ICC profile is a kind of color management method based on the ICC standard by the International Color Consortium. Its basic idea is to choose a reference color space which has nothing to do with the equipment to describe the equipment's characteristic; therefore, it can make it up to the loss of color information caused by equipment. So far, this color management method has been admitted by a lot of operating systems and equipment manufacturers, and it has become an internationally recognized method for color correction. From this study, it can be concluded that ICC profile correction could not only decrease color value difference (ΔE) evidently but also reduce the dispersion of different tongue colors effectively, which would make the identified tongue images more consistent on the coherence and authenticity with the images seen in the clinic by the TCM practitioners.

Moreover, the outcomes indicate the a^* value of the light red tongue was between the pale/purplish tongue and the red tongue ($p < 0.05$), while its b^* value was obviously higher than the red tongue, crimson tongue, and purplish tongue ($p < 0.05$). Preliminarily objective quantification could be carried out for the light red tongue in the future. As for the pale tongue, a^* value is significantly lower than the red tongue and the light red/crimson tongue ($p < 0.05$), and it has a higher b^* value than the red tongue, crimson tongue, and purplish tongue ($p < 0.05$); combined with its higher L^* value than the light red tongue ($p < 0.05$), the red component of the pale tongue is significantly decreased, corresponding to the general understanding of Traditional Chinese Medicine. The red tongue has the highest a^* value than the rest of the four colors ($p < 0.05$) and a lower L^* value than the light red tongue, pale tongue, and purplish tongue ($p < 0.05$); that is to say, its red component is relatively higher. With the red color being deeper, the brightness dimmed. As for the crimson tongue, its L^* value was obviously lower than the light red tongue, pale tongue, and purplish tongue ($p < 0.05$), its a^* value was between pale/purplish tongue and red tongue ($p < 0.05$), and its b^* value is obviously lower than red tongue, light red tongue, and pale tongue ($p < 0.05$). The crimson tongue had a mediate red component but relatively small brightness. For the purplish tongue, its a^* value is obviously lower than the light red tongue, red tongue, and crimson tongue ($p < 0.05$), and its b^* value is the lowest compared to the other four colors ($p < 0.05$). This shows the decreasing of the red component and the yellow component and the tendency to be blue and purple. Totally, a^* value became gradually smaller from the red tongue to the crimson/light red tongue and to the pale/purplish tongue, which brought the decreasing of the red component. And the L^* value became gradually smaller from the pale tongue to light red tongue, to purplish tongue, and to red tongue/crimson tongue, which caused the gradually dim brightness. As for the b^* value, the purplish tongue was the lowest compared to the others, which indicated its highest blue component.

From the perspective of $L^*a^*b^*$ color space, the main characteristics index of pale tongue and red tongue were L^* and a^* value; namely, the pale tongue had a higher L^* value and a lower a^* value than normal tongue color, while the red tongue is the opposite. The main characteristics index of the crimson tongue is lower L^* value than the normal tongue color. And the purplish tongue shows the lowest b^* value and the decreasing of a^* value.

This experiment suggests the different characteristics in $L^*a^*b^*$ color space for varied colors. For the imbalance and limitation of the studied samples, some index of the five tongue color in the $L^*a^*b^*$ color space has shown no significant difference, such as the L^* value for the red and crimson tongue, a^* value of the pale tongue and purplish tongue, a^* value of the light red tongue and crimson tongue, and b^* value of the light red tongue and pale tongue. Therefore, it is feasible for the standardization and formulation of medical reference range for tongue colors in TCM and it is necessary to increase the sample size of other abnormal tongue colors in the further study.

In addition, we also study the tongue color classification in the application of machine learning methods. Comparing to the commonly used SVM algorithm, Random forest has a better result on the tongue color classification despite using imbalanced sample for different tongue colors (Tables 2, 7, and 8). Moreover, SMOTE algorithm could improve both the whole accuracy of tongue color classification and the accuracy of those abnormal tongue color classification effectively by sample amplification (Tables 2, 3, 4, 5, 6, 7, and 8). Consistency of the tongue image collection process and the use of the color correction ensured the uniformity and authenticity of the tongue pictures' quality. The inconformity of tongue image segmentation during the tongue image processing procedure would lead to the difference of tongue color classification index, so an agreement should be reached in the future on the tongue image processing method. Only 3 experts are chosen to recognize the tongue color in this study; on account of the difference of the subjective experience of varied observers and the uncertainty of the classification standard, the consensus on the TCM expert diagnosis system of the tongue images should be reached in the future.

Last but not least, color is a low-level feature for tongue images. The identification and extraction of other tongue features like texture or fissures are based on the color information of the digital tongue image. Therefore, we initially researched the objective classification of tongue color. Given the idea that TCM normally combines color, texture, and fissures for diagnosis rather than the single factor color, studies on the objective and quantitative classification of texture and fissures are still required in the future.

5. Conclusions

In this study, preliminary objectification of tongue color classification is feasible on the basis of standardized tongue acquisition and color correction. Data mining methods of SVM and Random forest can help us to evaluate the classification results effectively. Random forest has a better performance in the classification accuracy of abnormal tongue colors than SVM. SMOTE algorithm can improve the classification accuracy by solving the imbalance of the studied samples. Our research would contribute to the automatic diagnostic system of tongue image in TCM.

Competing Interests

All the authors declare that they have no conflict of interests.

Authors' Contributions

Zhen Qi and Li-ping Tu contributed equally to this work.

Acknowledgments

The research was conducted through the National Natural Science Foundation of China (no. 81373556), National Key Technology R&D Program (no. 2012BAI37B06),"Shu Guang" Project Supported by Shanghai Municipal Education Commission and Shanghai Education Development Foundation (no. 12SG36), Basic Subject Innovation Fund Supported by Shanghai Municipal Education Commission (no. A2-P1600317), and Science and Technology Development Program of Jilin Province (no. 20160414009GH, no. 20140101206JC-21).

References

[1] Z. Liu, J.-Q. Yan, D. Zhang, and Q.-L. Li, "Automated tongue segmentation in hyperspectral images for medicine," *Applied Optics*, vol. 46, no. 34, pp. 8328–8334, 2007.

[2] J. Ning, D. Zhang, C. Wu, and F. Yue, "Automatic tongue image segmentation based on gradient vector flow and region merging," *Neural Computing and Applications*, vol. 21, no. 8, pp. 1819–1826, 2012.

[3] M. F. Zhu and J. Q. Du, "A novel approach for color tongue image extraction based on random walk algorithm," *Applied Mechanics and Materials*, vol. 462-463, pp. 338–342, 2014.

[4] N. D. Kamarudin, C. Y. Ooi, T. Kawanabe, and X. Mi, "Tongue's substance and coating recognition analysis using HSV color threshold in tongue diagnosis," in *1st International Workshop on Pattern Recognition*, vol. 10011 of *Proceedings of SPIE*, 5 pages, Tokyo, Japan, May 2016.

[5] K. H. Kim, J. H. Do, H. Ryu et al., "Tongue diagnosis method for extraction of effective region and classification of tongue coating," in *Proceedings of the 1st Workshops on Image Processing Theory, Tools and Applications (IPTA '08)*, pp. 1–7, Sousse, Tunisia, November 2008.

[6] X. Zhiming, Y. Bao-an, and C. Xin, "Semantic object generation in tongue image analysis," in *Innovations and Advanced Techniques in Computer and Information Sciences and Engineering*, T. Sobh, Ed., pp. 41–46, Springer, Dordrecht, The Netherlands, 2007.

[7] T. Obafemi-Ajayi, R. Kanawong, D. Xu, S. Li, and Y. Duan, "Features for automated tongue image shape classification," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '12)*, pp. 273–279, Philadelphia, Pa, USA, October 2012.

[8] B. Huang, J. Wu, D. Zhang, and N. Li, "Tongue shape classification by geometric features," *Information Sciences*, vol. 180, no. 2, pp. 312–324, 2010.

[9] B. Zhang and H. Zhang, "Significant geometry features in tongue image analysis," *Evidence-Based Complementary and Alternative Medicine*, vol. 2015, Article ID 897580, 8 pages, 2015.

[10] Q. Li, Y. Wang, H. Liu, Z. Sun, and Z. Liu, "Tongue fissure extraction and classification using hyperspectral imaging technology," *Applied Optics*, vol. 49, no. 11, pp. 2006–2013, 2010.

[11] Y.-G. Wang, J. Yang, Y. Zhou, and Y.-Z. Wang, "Region partition and feature matching based color recognition of tongue image," *Pattern Recognition Letters*, vol. 28, no. 1, pp. 11–19, 2007.

[12] L. Zhi, D. Zhang, J.-Q. Yan, Q.-L. Li, and Q.-L. Tang, "Classification of hyperspectral medical tongue images for tongue diagnosis," *Computerized Medical Imaging and Graphics*, vol. 31, no. 8, pp. 672–678, 2007.

[13] Q. Li and Z. Liu, "Tongue color analysis and discrimination based on hyperspectral images," *Computerized Medical Imaging and Graphics*, vol. 33, no. 3, pp. 217–221, 2009.

[14] Y. Wang, Y. Zhou, J. Yang, and Y. Wang, "JSEG based color separation of tongue image in traditional Chinese medicine," in *Proceedings of the 9th Iberoamerican Congress on Pattern Recognition, Progress in Pattern Recognition, Image Analysis and Applications (CIARP '04)*, A. Sanfeliu, J. F. Martínez Trinidad, and JA. Carrasco Ochoa, Eds., pp. 503–508, Springer, Puebla, Mexico, October 2004.

[15] T. Kawanabe, N. D. Kamarudin, C. Y. Ooi et al., "Quantification of tongue colour using machine learning in Kampo medicine," *European Journal of Integrative Medicine*, 2016.

[16] X. Zhang, J. Zhang, G. Hu, and Y. Wang, "Preliminary study of tongue image classification based on multi-label learning," in *Proceedings of the 11th International Conference on Advanced Intelligent Computing Theories and Applications (ICIC '15)*, pp. 208–220, Fuzhou, China, August 2015.

[17] Y. Jiao, X. Zhang, L. Zhuo, M. Chen, and K. Wang, "Tongue image classification based on Universum SVM," in *Proceedings of the 3rd International Conference on BioMedical Engineering and Informatics (BMEI '10)*, pp. 657–660, IEEE, Yantai, China, October 2010.

[18] S. Yamamoto, Y. Ishikawa, T. Nakaguchi et al., "Temporal changes in tongue color as criterion for tongue diagnosis in Kampo medicine," *Forschende Komplementärmedizin*, vol. 19, no. 2, pp. 80–85, 2012.

[19] T. Oji, T. Namiki, T. Nakaguchi et al., "Study of factors involved in tongue color diagnosis by Kampo medical practitioners using the Farnsworth-Munsell 100 hue test and tongue color images," *Evidence-based Complementary and Alternative Medicine*, vol. 2014, Article ID 783102, 2014.

[20] B. Zhang, X. Wang, J. You, and D. Zhang, "Tongue color analysis for medical application," *Evidence-Based Complementary and Alternative Medicine*, vol. 2013, Article ID 264742, 11 pages, 2013.

[21] B. Huang and N. Li, "Pixel based tongue color analysis," in *Proceedings of the 1st International Conference Medical Biometrics (ICMB '08)*, D. Zhang, Ed., pp. 282–289, Springer, Hong Kong, January 2008.

[22] H. Siu Cheung, H. Yulan, and T. Doan Thi Cam, "Machine learning for tongue diagnosis," in *Proceedings of the 6th International Conference on Information, Communications & Signal Processing*, pp. 1–5, Singapore, December 2007.

[23] Z. Yang and N. Li, "Automatic classification of tongueprints in healthy and unhealthy individuals," in *Proceedings of the 1st International Conference on Medical Biometrics (ICMB '08)*, D. Zhang, Ed., pp. 137–144, Hong Kong, 2008.

[24] R. Kanawong, T. Obafemi-Ajayi, J. Yu, D. Xu, S. Li, and Y. Duan, "ZHENG classification in Traditional Chinese Medicine based on modified specular-free tongue images," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW '12)*, pp. 288–294, IEEE, Philadelphia, Pa, USA, October 2012.

[25] Z. Gao, L. Po, W. Jiang, X. Zhao, and H. Dong, "A novel computerized method based on support vector machine for tongue diagnosis," in *Proceedings of the 3rd IEEE International Conference on Signal-Image Technologies and Internet-Based Systems (SITIS '07)*, pp. 849–854, IEEE, Shanghai, China, December 2007.

[26] J. H. Jang, J. E. Kim, K. M. Park, S. O. Park, Y. S. Chang, and B. Y. Kim, "Development of the digital tongue inspection system with image analysis," in *Proceedings of the 2002 IEEE Engineering in Medicine and Biology 24th Annual Conference and the 2002 Fall Meeting of the Biomedical Engineering Society (BMES/EMBS '02)*, vol. 2, pp. 1033–1034, Houston, Tex, USA, October 2002.

[27] H. Zhang, K. Wang, D. Zhang, B. Pang, and B. Huang, "Computer aided tongue diagnosis system," in *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 7, pp. 6754–6757, IEEE Engineering in Medicine and Biology Society, Shanghai, China, 2005.

[28] S. Yamamoto, N. Tsumura, T. Nakaguchi et al., "Principal component vector rotation of the tongue color spectrum to predict 'Mibyou' (disease-oriented state)," *International Journal of Computer Assisted Radiology and Surgery*, vol. 6, no. 2, pp. 209–215, 2011.

[29] J.-T. Xu, Z.-F. Zhang, Z.-J. Yan et al., "Analysis of tongue color under natural daylight based on chromatic aberration correction," *Journal of Chinese Integrative Medicine*, vol. 7, no. 5, pp. 422–427, 2009.

[30] J. T. Xu and Z. F. Zhang, "The experimental observation of tongue digital image acquisition conditions," *Chinese Journal of Basic Medicine in Traditonal Chinese Medicine*, vol. 13, no. 1, pp. 23–27, 2007 (Chinese).

[31] C. Zhao, G.-Z. Li, C. Wang, and J. Niu, "Advances in patient classification for traditional chinese medicine: a machine learning perspective," *Evidence-Based Complementary and Alternative Medicine*, vol. 2015, Article ID 376716, 2015.

[32] C. Meng and Z. Chun Lei, "The comparison study of LIBSVM, LIBLINER, SVMmulticlass," *Elextronic Technology*, vol. 6, no. 1, pp. 1–5, 2015 (Chinese).

[33] L. Wu-hui, *Graphic Arts Chromatology*, Chemical Industry Press, Beijing, China, (Chinese) 2004.

[34] W. Su, Z.-Y. Xu, Z.-Q. Wang, and J.-T. Xu, "Objectified study on tongue images of patients with lung cancer of different syndromes," *Chinese Journal of Integrative Medicine*, vol. 17, no. 4, pp. 272–276, 2011.

[35] D. Zhang, B. Pang, N. Li, K. Wang, and H. Zhang, "Computerized diagnosis from tongue appearance using quantitative feature classification," *The American Journal of Chinese Medicine*, vol. 33, no. 6, pp. 859–866, 2005.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 3162649, 10 pages
http://dx.doi.org/10.1155/2016/3162649

Research Article

Rapid Retrieval of Lung Nodule CT Images Based on Hashing and Pruning Methods

Ling Pan,[1] Yan Qiang,[1] Jie Yuan,[2] and Lidong Wu[3]

[1] *College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, China*
[2] *Shanxi Provincial People's Hospital, Taiyuan 030012, China*
[3] *University of Texas at Tyler, 3900 University Blvd., Tyler, TX 75799, USA*

Correspondence should be addressed to Yan Qiang; qiangyan@tyut.edu.cn

Received 19 July 2016; Accepted 18 October 2016

Academic Editor: Yong Xia

The similarity-based retrieval of lung nodule computed tomography (CT) images is an important task in the computer-aided diagnosis of lung lesions. It can provide similar clinical cases for physicians and help them make reliable clinical diagnostic decisions. However, when handling large-scale lung images with a general-purpose computer, traditional image retrieval methods may not be efficient. In this paper, a new retrieval framework based on a hashing method for lung nodule CT images is proposed. This method can translate high-dimensional image features into a compact hash code, so the retrieval time and required memory space can be reduced greatly. Moreover, a pruning algorithm is presented to further improve the retrieval speed, and a pruning-based decision rule is presented to improve the retrieval precision. Finally, the proposed retrieval method is validated on 2,450 lung nodule CT images selected from the public Lung Image Database Consortium (LIDC) database. The experimental results show that the proposed pruning algorithm effectively reduces the retrieval time of lung nodule CT images and improves the retrieval precision. In addition, the retrieval framework is evaluated by differentiating benign and malignant nodules, and the classification accuracy can reach 86.62%, outperforming other commonly used classification methods.

1. Introduction

The early diagnosis and treatment of lung cancer patients can help improve their survival rate [1]. However, with the development and improvement of various medical image scanning technologies, especially computed tomography (CT), the number of medical images is growing explosively every year. Hence, in the early medical screening process, reviewing lung lesions is an extremely labor-intensive job for radiologists. In addition, when reviewing and analyzing lesions, radiologists mainly rely on their diagnostic experience, and the diagnosis tends to be highly subjective. Moreover, clinical statistical studies show that the same radiologist, at different times, under different states of physical fatigue, may come up with a different diagnosis for the same CT image. Therefore, it is necessary to retrieve similar lung nodule CT images to improve diagnostic efficiency. By obtaining similar images from a CT image repository of pulmonary nodules, the anamnesis and successful treatments of these images can be viewed as clinical references for the case under consideration, which can lessen the reliance on a physician's clinical diagnostic experience to a certain degree.

Given the explosive growth of the number of current lung images and advantage of medical image retrieval for physicians' diagnosis of lung lesions, in this paper, a novel retrieval framework based on a hashing and pruning algorithm for lung nodule CT images is proposed. When retrieving similar lung nodule CT images, it not only can reduce the memory space required but also further shorten the retrieval time and improve precision with a pruning-based similarity measure method.

The remainder of this paper is organized as follows. Section 2 introduces previous work related to lung nodule image retrieval and current popular retrieval methods. Section 3 describes the proposed retrieval framework in detail. Experimental results are presented with some discussion in Section 4. Section 5 concludes the paper and discusses future work.

2. Related Work

Recent years have witnessed the growing popularity of medical image retrieval, and there are many significant results in the field of lung imaging. Oliveira and Ferreira [2] proposed a bag-of-tasks method combining texture features and registration techniques to retrieve lung cancer images. Ng et al. [3] presented an improved hierarchical spatial descriptor and binary descriptor to retrieve similar lung nodule CT images from the perspective of spatial similarity. Aggarwal et al. [4] studied the detection and classification of lung nodules with content-based medical image retrieval. However, given the large number of lung CT images generated every year, effective medical image retrieval is still a difficult challenge.

At present, hashing-based methods for image retrieval can solve the storage and efficiency problems that traditional image retrieval methods may encounter [5]. Further, these hashing methods can transform high-dimensional image data into a low-dimensional binary space by utilizing the constructed hash functions [6]. It is precisely because of these advantages that many scholars focus on researching hashing-based image retrieval technology. Locality-sensitive hashing [7], a pioneering work, can generate compact binary codes with a random threshold. Further, in many hashing methods, such as those in [8–10], principal component analysis (PCA) is a common method for preprocessing image data. The simplest of these approaches is PCAH: after using PCA to reduce the dimensionality of the image data, "0" is viewed as the boundary, and both sides, respectively, correspond to "0" and "1." Moreover, according to whether label information is used to construct hash functions, hashing methods can be categorized as unsupervised hashing [8, 9, 11], semisupervised hashing [10], or supervised hashing [12, 13]. Additionally, the core of these hashing methods is the minimization of the error when translating the image data into binary space.

Although many hashing and improved hashing methods have been presented, only a few researchers have applied them to medical image retrieval. Liu et al. [14] presented an image retrieval framework for digital mammograms with anchor graph hashing and improved its search accuracy by fusing different features. Jiang et al. [15] used a joint kernel-based supervised hashing algorithm with a small amount of supervised information to compress breast histopathological images into 10-bit hash codes and identified actionable and benign tumors based on the retrieval results. Zhang et al. [16] built a histopathological image retrieval framework using a supervised hashing with kernel (KSH) method and validated the retrieval performance on breast microscopic tissues images.

In our proposed retrieval method for lung nodule CT images, partitioning the dataset with a clustering algorithm is the foundation of the pruning algorithm. The KSH method is then used to translate the images in each cluster into short hash codes and form the hash code database. During retrieval, a pruning algorithm is employed to shorten the retrieval area and further improve the retrieval speed and precision. We use other state-of-the-art hashing methods to validate the proposed pruning algorithm and compare it with other commonly used classification methods to demonstrate the performance of our retrieval framework.

3. Description of the Retrieval Framework and Pruning Algorithm

The retrieval framework for lung nodule CT images proposed in this paper consists of two main parts, the learning phase and query phase, as shown in Figure 1. The aim of the learning phase is to build a hash code database. First, we use the extracted visual and medical features to represent each lung nodule CT image. Binary codes are then obtained and stored in a hash code database using the KSH method. In the query phase, given a query lung nodule CT image, we first extract the same features that were extracted in the learning phase and translate them into binary code with the constructed hash functions. Next, similar images are retrieved from the hash code database while using the pruning algorithm. The retrieval results can be used to recognize benign or malignant nodule.

3.1. Lung Nodule Feature Extraction. Feature extraction plays an important role in image retrieval and can transform high-dimensional nodule images into a lower-dimensional space while retaining the essential content of the image. Good features can help physicians to distinguish lung nodules efficiently [17–19]. To facilitate analysis and research on lung lesions, we extract lung nodule features based on grayscale, morphology, and texture.

Grayscale features are the most basic characteristics of an image of lung nodules, and the grayscale difference can highlight the corresponding organization and structure. The proposed method extracts three gray level characteristics, grayscale mean, variance, and entropy, where grayscale entropy reflects the grayscale information contained in the nodule image, and is defined as

$$H = -\sum_{i=0}^{k-1} H(i) \log H(i), \tag{1}$$

where $H(i)$ represents the probability density of a different greyscale value and k is the number of available gray level values.

Morphological features are the most intuitive visual features and are helpful for identifying tumors. We describe the morphological features of lung nodule mainly using invariant moments, medical signs, and geometric features. We employ the seven invariant moments proposed by Hu to describe the shape characteristics of pulmonary nodules. The calcification area, calcification degree, cavitary area, and cavitary ratio are calculated and represent the medical sign information of the lung nodules. The geometric features consist of lung nodule perimeter, area, maximum diameter, rectangle, and roundness, where roundness describes the degree of deviation of the nodule region from a circular shape, defined as

$$F = \frac{4\pi A}{L^2}, \tag{2}$$

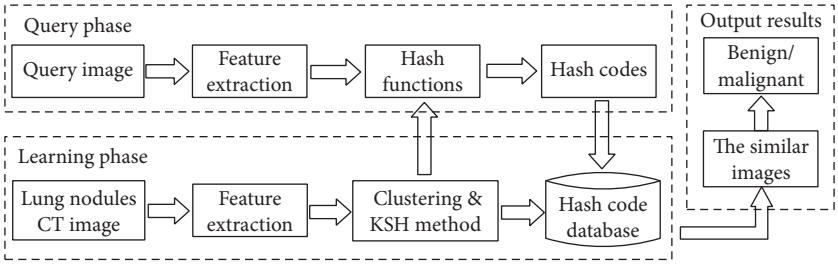


Figure 1: Retrieval framework for lung nodule CT images with hashing and pruning algorithms.

where A is the area of the nodule region and L is the perimeter.

Texture features can provide important information on the health of the lung. For example, the structure of diseased tissue is more chaotic and rough than healthy tissue [18]. Here, the gray level co-occurrence matrix is used to extract texture features. This is the most widely used texture analysis method in medical imaging. The computed features include 14 characteristic values, such as contrast, angular second moment, entropy, and inverse difference moment, which are defined as

$$\begin{aligned} \mathrm{CON} &= \sum_{n=0}^{k-1} n^2 \left\{ \sum_{|i-j|=n} G(i,j) \right\}, \\ \mathrm{ASM} &= \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \{G(i,j)\}^2, \\ \mathrm{ENT} &= -\sum_{i=0}^{k-1} \sum_{j=0}^{k-1} G(i,j) \log G(i,j), \\ \mathrm{IDM} &= -\sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \frac{G(i,j)}{1+(i-j)^2}. \end{aligned} \tag{3}$$

We also calculate the mean and variance of these 14 feature values.

A detailed description of the extracted multiple features is given in Table 1. By extracting the lung nodule features from grayscale, morphology, and texture, we utilized a 104-dimensional feature vector to uniquely represent each lung nodule CT image, constructed as follows:

$$x = \left[\underbrace{f_1, f_2, f_3}_{\text{gray}}, \underbrace{f_4, \ldots, f_{10}, f_{11}, \ldots, f_{15}, f_{16}, \ldots, f_{20}}_{\text{morphology}}, \underbrace{f_{21}, f_{22}, \ldots, f_{104}}_{\text{texture}} \right]. \tag{4}$$

3.2. Building the Hash Code Database. In order to achieve the rapid retrieval of lung nodule CT images using the proposed pruning algorithm, it is necessary to partition the dataset before constructing the hash code database using the hashing method. As shown in Figure 2, the construction of a hash code database includes two parts: clustering and hashing. In the first part, a spectral clustering algorithm is used to partition our training dataset into several clusters so that the distribution of lung nodule CT images in each cluster is near uniform. In addition, when retrieving a query image, the retrieval scope can be narrowed according to the distance between the query image and cluster centers. In the second part, the KSH method is employed to construct hash functions for the whole training dataset and obtain the hash code database. Furthermore, the uniform distribution of lung nodule CT images in clusters could reduce the instability of retrieval performance caused by the uneven distribution of images.

3.2.1. Partitioning the Lung Nodule Dataset. Spectral clustering is a clustering algorithm based on spectral graph theory that can identify a sample space with arbitrary shapes and converge to the global optimal solution. Furthermore, the obtained clustering results outperform traditional clustering approaches, such as k-means or single linkage clustering [20, 21].

In this paper, given a training dataset of lung nodule CT image $\chi = \{x_1, \ldots, x_n\} \in R^d$, where n is the number of images and d is the dimension of extracted features, a graph G is first built to represent these data. The vertices V in the graph represent lung nodule CT images, and the weights of edges E represent the similarity of any two lung nodule CT images. An undirected weighted graph $G = (V, E)$ based on the similarity of the images then can be obtained. Thus, the clustering problem is converted into a graph partitioning problem on G.

The main step of spectral clustering is to construct graph partitions based on graph Laplacian matrix L. Here, we use the normalized Laplacian matrix, which is defined as

$$L = D^{-1/2} (D - W) D^{-1/2}, \tag{5}$$

where W is a similarity matrix, defined as

$$W_{ij} = \exp^{-\|x_i - x_j\|^2 / 2\delta^2}, \tag{6}$$

Table 1: Description of lung nodule feature extraction.

Feature	Descriptors	Representation
Gray	3 gray features (gray mean, variance, and entropy)	f_1, f_2, f_3
Morphology	7 invariant moments features (Hu moment invariants) 5 geometrical features (perimeter, area, roundness, rectangularity, and maximum diameter) 4 medical signs (calcification area and degree, cavitary area, and degree) and Fourier descriptor	$f_4, \ldots, f_{10}$, $f_{11}, \ldots, f_{15}$, $f_{16}, \ldots, f_{19}, f_{20}$
Texture	14 texture features from GLCM (contrast, angular second moment, entropy, inverse difference moment, etc.) with 4 different angles as well as the mean, variance of them	$f_{21}, f_{22}, \ldots, f_{104}$

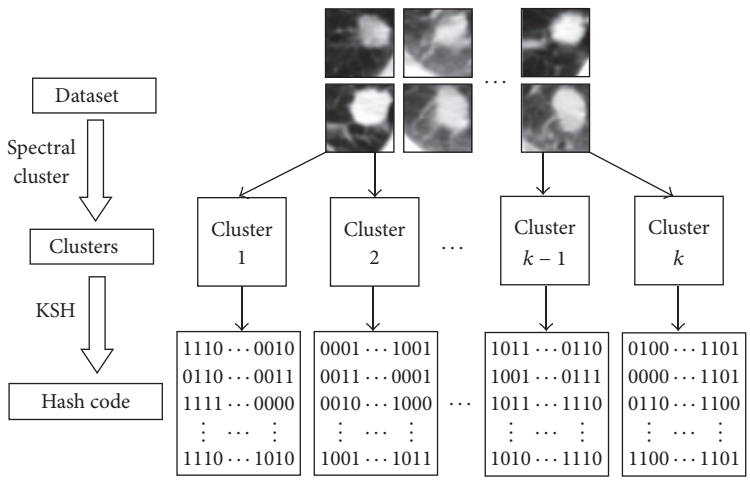


Figure 2: Hash code database construction.

where δ is a parameter and D is a diagonal matrix obtained from W, defined as

$$D_{ij} = \begin{cases} \sum_{j=1}^{n} W_{ij}, & i = j \\ 0, & i \neq j. \end{cases} \tag{7}$$

The classical k-means method is then utilized to cluster the eigenvectors of Laplacian matrix L. Using the above steps, we can acquire several clusters in which the lung nodule CT images are similar to each other. This is the foundation for achieving a pruning algorithm and is helpful for improving the retrieval speed as well as precision.

3.2.2. Construction of Hash Functions. One of the factors affecting the performance of a hashing method is the ability to preserve the similarity of any two images in the original feature space. Hence, the key to a hashing method is to construct appropriate hash functions and maintain the similarity within the hash code. KSH is a supervised hashing method that uses a limited amount of supervised information for learning hash functions, and the retrieval results are better than other unsupervised hashing methods as well as some supervised hashing methods.

Given all the lung nodule CT images in training dataset χ, we need to construct a group of hash functions $H = \{h_1(x), \ldots, h_r(x)\}$, each of which will generate a single hash bit. In addition, if the length of the hash code is r, then r hash functions will be constructed. A hash function is defined as

$$h(x) = \operatorname{sgn}(f(x)) = \operatorname{sgn}\left(\sum_{j=1}^{m} \kappa\left(x_{(j)}, x\right) a_j - b\right), \tag{8}$$

where $\kappa(x, y) = \exp(-\|x - y\|/2\sigma^2)$is a Gaussian kernel function for solving the problem of the linear inseparability of lung nodule images, $\{x_{(1)}, \ldots, x_{(m)}\}$ are samples randomly selected from the training dataset χ to support kernel computation, $\{a_1, \ldots, a_m\}$ are a group of coefficients, $\operatorname{sgn}(x)$ is a sign function outputting 1 for positive input and -1 for negative input, and $b \in R$ is a bias defined as

$$b = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \kappa\left(x_{(j)}, x_i\right) a_j. \tag{9}$$

As the differences in lung nodule images are not apparent, as they are in natural images, the coefficient vector $\mathbf{a} = [a_1, \ldots, a_m]^T$ is vital for generating distinguishable hash functions. Here, supervised information, that is, a label matrix, is utilized to solve this problem. During the image preprocessing, we mark each lung nodule CT image with a label 1 or 0 based on whether it shows benign or malignant lesions. The l $(m < l \leq n)$ images from the training dataset are then randomly selected to construct a label matrix $\mathbf{S} \in R^{l \times l}$. The construction process is shown in Figure 3.

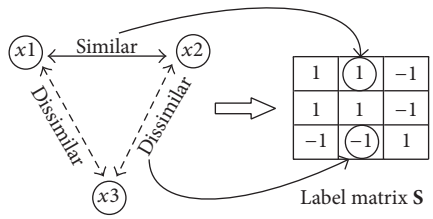


FIGURE 3: Construction of a label matrix.

When retrieving a query image, the Hamming distance is a commonly used method to measure the similarity between the query image and the database images. However, it is difficult to directly compute this distance because of its complex formula. The research in [13] explains the corresponding relation between the Hamming distance and code inner product. Hence, the objective function can be defined using a code matrix formed by the selected l samples and label matrix to solve $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_r]$ for hash functions H as follows:

$$\min_{\mathbf{A}} Q(\mathbf{A}) = \left\| \sum_{k=1}^{r} \operatorname{sgn}\left(\overline{K}_l \mathbf{a}_k\right) \left(\operatorname{sgn}\left(\overline{K}_l \mathbf{a}_k\right)\right)^T - r\mathbf{S} \right\|_F^2, \quad (10)$$

where $\|\cdot\|_F$ is the Frobenius norm and $\overline{K}_l$ represents the kernel computation for the l images involved in label matrix $\mathbf{S}$ and is expressed as

$$\overline{K}_l = \begin{bmatrix} \kappa(x_{(1)}, x_1) - \frac{1}{n}\sum_{i=1}^{n}\kappa(x_{(1)}, x_i) & \cdots & \kappa(x_{(m)}, x_1) - \frac{1}{n}\sum_{i=1}^{n}\kappa(x_{(m)}, x_i) \\ \vdots & \ddots & \vdots \\ \kappa(x_{(1)}, x_l) - \frac{1}{n}\sum_{i=1}^{n}\kappa(x_{(1)}, x_i) & \cdots & \kappa(x_{(m)}, x_l) - \frac{1}{n}\sum_{i=1}^{n}\kappa(x_{(m)}, x_i) \end{bmatrix} \in \mathbb{R}^{l \times m}. \quad (11)$$

Thus, by minimizing the error between the code matrix and label matrix, the hash functions with supervised information can be acquired and used to encode each lung nodule CT image.

3.3. Retrieval Process for Lung Nodule CT Images. The aim of the proposed method is to achieve rapid retrieval for lung nodule CT images with higher precision. Hence, a pruning algorithm is presented. The retrieval procedure with pruning algorithm is illustrated in Figure 4.

Given a query image, the retrieval process includes the following three steps: (1) determining candidate clusters, that is, selecting some clusters as the candidate clusters to which the query image may belong, (2) encoding the query image, that is, compressing the extracted relevant features into binary codes with the constructed hash functions, and (3) calculating similarity, that is, computing and sorting the code inner products between the query image and images in the candidate clusters and returning the similar images according to the similarity.

However, when sorting the code inner products, if the length of hash code is r, the value range of code inner product is $[-r, r]$, and it is impossible to directly sort the images that have the same code inner products. In order to solve this problem, we designed a decision rule: compare the distances between the query image and the clusters that these similar images belong to and return the image with a smaller distance. The whole pruning process is shown in Algorithm 1.

Algorithm 1 (pruning algorithm for image retrieval).

Input. Query lung nodule image q, cluster centers $\{\mu_1, \ldots, \mu_k\}$, number of candidate clusters m $(0 < m < k)$, hash functions $H = \{h_1(x), \ldots, h_r(x)\}$, hash codes $\{\text{code}_{\mu_1}, \ldots, \text{code}_{\mu_k}\}$ in clusters, and number of returned similar images p.

Output. Similar images $\{x_1, \ldots, x_p\}$.

Step 1. Calculate the distance between q and each cluster center using

$$d = \|q - \mu_i\|^2, \quad (i = 1, \ldots, k). \quad (12)$$

Step 2. Rank and select the top m clusters as the candidate clusters using

$$\text{cluster}_1, \ldots, \text{cluster}_m \longleftarrow \text{sort}\left(d_i, \text{“abscend”}\right). \quad (13)$$

Step 3. Compress the query image into code with H.

Step 4. Calculate the code inner products between q and the images in the candidate clusters as follows:

$$\text{sim} = \text{code} \circ \text{code}_{\mu_i}, \quad (i = 1, \ldots, m). \quad (14)$$

Step 5. Rank the code inner products sim as follows:

$$x_1, \ldots, x_p \longleftarrow \text{sort}\left(\text{sim}, \text{“descend”}\right). \quad (15)$$

Step 6. If equal $(\text{sim}_i, \text{sim}_j)$, then compare the corresponding distances d_i and d_j in Step 1.

Step 7. If $d_i > d_j$, then return image x_j first.

Step 8. Repeat Steps 6-7 until p similar images are returned in order.

4. Experimental Results and Discussion

In this section, we first introduce the dataset used in the evaluation and the extracted multiple features. Next, we discuss

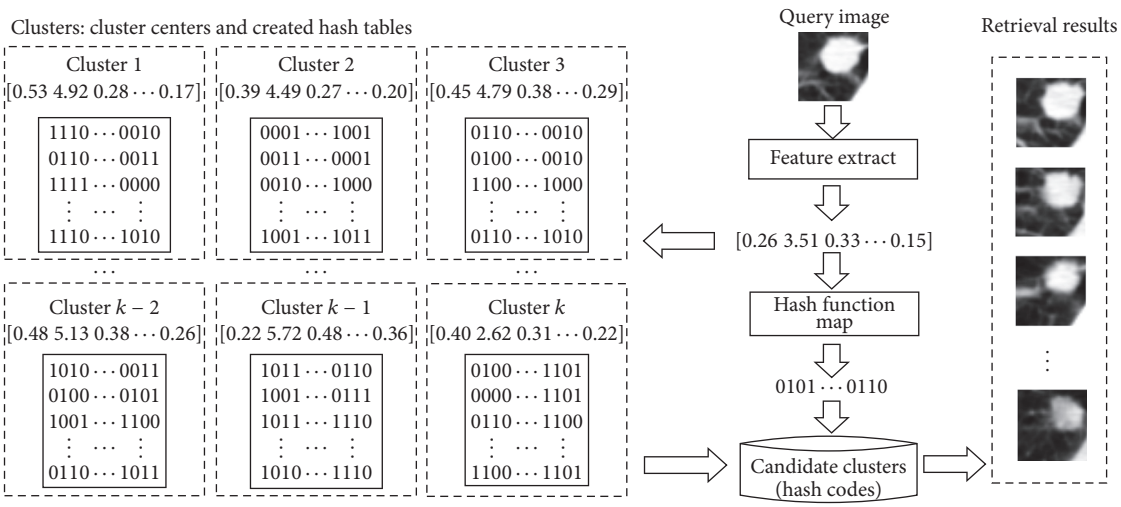


Figure 4: Procedure of lung nodule CT image retrieval with pruning.

Table 2: Distribution of cases and slices over the lung tumor types involved in the dataset.

Lung tumor type	Cases	Slices
Benign	250	1054
Malignant	350	1396

the parameter settings in our proposed retrieval method. We then validate our pruning algorithm on different hashing methods. Finally, we evaluate the performance of our retrieval framework by comparing it with other commonly used classification methods. All our experiments were implemented in MATLAB R2014b on a workstation with Intel Core i7-4770 CPU 3.40 GHz and 8 GB of RAM.

4.1. Image Dataset. The image data used in our experiments are from the LIDC dataset [22]. The LIDC dataset contains 1,018 cases, each of which includes a set of chest CT images and an associated XML file that records some relevant information about the lung nodules (such as whether they are benign or malignant). There are a total of 7,371 nodules labeled at least by one radiologist and 2,669 of these nodules are marked "nodule ≥ 3 mm." Here, in order to ensure that the training dataset does not influence the testing dataset (i.e., that no images belonging to the same case appear in both the training dataset and the testing dataset), the dataset in this research is constructed from 600 cases. Further, we randomly selected 70% of them as training data and the remaining 30% as testing data. Here, slices with unclear nodules in each case were discarded, resulting in a total of 2,450 slices in our dataset. The detailed contents of this dataset are listed in Table 2.

This study is aimed at lung nodules, so the first step in our experiment is to obtain the region of interest. As the XML files in the LIDC record information about the lung nodules, our team designed a visual interface and parsed these XML files to obtain the relevant information, as shown

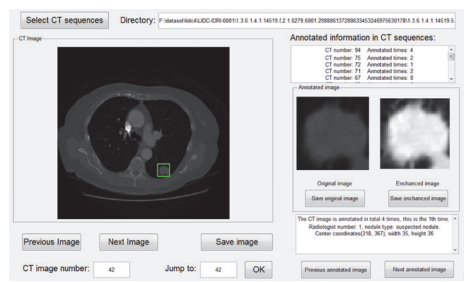


Figure 5: Visual interface for obtaining the relevant information from the lung nodule image.

in Figure 5. The rectangular regions containing lung nodules were extracted and viewed as regions of interest. Thus, the lung image database was constructed based on these regions.

In order to facilitate the research and analysis of lung lesions, we extracted the multiple features of lung nodules and stored them into our database in advance. Table 3 describes some feature values extracted from the lung nodule images.

4.2. Parameter Settings. There are three main parameters affecting the performance of the proposed retrieval framework. They are the length of the compact hash code *bit*, number of clusters k, and number of candidate clusters used for retrieval m.

The effect of a hashing method is to transform the high-dimensional image features of lung nodules into a low-dimensional hash code to represent each image. Additionally, based on the experience of a large number of studies, the length of hash codes in this paper is set to between 8 and 64 bits. Here, in order to determine the influence of the hash code length on the retrieval results, we first did our experiment using the KSH method only (without using the pruning algorithm). The retrieval precision for different hash code lengths is reported in Table 4, and the best retrieval results are obtained when $bit = 48$. Retrieval precision is one of

Table 3: Quantitative results of some features of lung nodule images.

Image	Gray		Morphology							Texture (0°)		
			7 invariant moments		Medical signs		Geometrical features					
	Mean	Entropy	M_1	M_2	Calcification degree	Cavitary ratio	Area	Roundness	Rectangle	ASM	CON	IDM
I_1	0.5327	4.9199	0.2789	0.0025	0.1720	0	343	0.3961	0.5489	0.0521	1.3518	0.7151
I_2	0.5403	4.8428	0.2731	0.0016	0.1954	0.2483	302	0.3503	0.5243	0.0451	1.9339	0.6360
I_3	0.4496	4.7236	0.3173	0.0032	0.2268	0	529	0.6274	0.3701	0.0854	2.1015	0.6853
I_4	0.3933	4.4967	0.3757	0.0037	0	0.1985	297	0.5632	0.4127	0.1415	1.5383	0.7105
I_5	0.5293	5.1302	0.2707	0.0017	0.2571	0.1684	412	0.5279	0.2953	0.0267	1.7759	0.6369

Table 4: Retrieval precision for the top 5 similar lung images for different hash code lengths.

Bits	8	16	24	32	40	48	56	64
Pre. for top 5	0.6997	0.7541	0.7731	0.8176	0.8351	0.8542	0.8207	0.8031

the criteria used to evaluate the performance of a retrieval method and defined as

$$\text{precision} = \frac{\text{relevant\_results}}{\text{results}}, \tag{16}$$

where results indicate the number of returned images and relevant_results denote the number of correct results in the returned images as judged by the label information.

Next, setting *bit* to 48, we discuss how to set parameters k and m appropriately to achieve a retrieval precision of 85.42% within the shortest retrieval time. Retrieval time refers to the period of time beginning with the encoding of the test images and ending when the similar lung nodule CT images have been obtained.

In order to reach the same retrieval precision without using a decision rule, the values of parameters m and k were acquired through experiment and are shown in Table 5. Figure 6 demonstrates how the retrieval time changes according to the number of clusters when the length of hash code is 48 bits. We can see that, as the number of clusters increases, the retrieval time decreases when $k < 35$, but when $k > 35$, the retrieval time increases. Considering Table 5 and Figure 6, the reason for this situation is that as the number of clusters increases, the reduced number of images in the candidate clusters is greater than the increased number of images in the newly produced candidate clusters. That is, the total number of images in the candidate clusters needed to reach the same retrieval precision is less than before. However, when $k > 35$, the situation is reversed. This may be similar to the phenomenon of overfitting in statistical learning. When the number of clusters increases, the retrieval result may be worse. Moreover, as the number of clusters increases, the time required to calculate the distance between the query image and each cluster center cannot be ignored. Hence, in order to obtain a better retrieval result, we set the length of the hash code to 48 bits and the number of clusters to 35. Additionally, as shown in Table 5, the number of candidate clusters is set to eight.

The retrieval results of lung nodule CT images based on the above parameter settings are shown in Figure 7, where the first two query images are malignant tumors, and the last two are benign tumors.

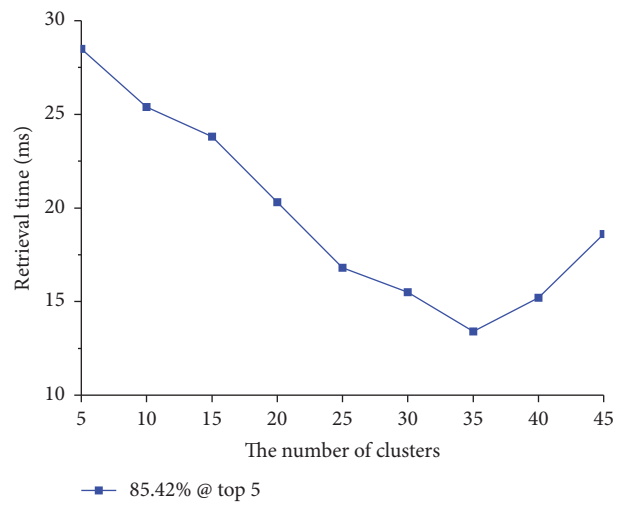


Figure 6: Retrieval time with respect to number of clusters for *bit* = 48.

4.3. Performance Comparison of the Pruning Algorithm. In these experiments, we applied the proposed pruning algorithm to different hash methods such as kernelized locality-sensitive hashing (KLSH) [11], spectral hashing (SH) [8], binary reconstructive embedding (BRE) [12], PCAH, and iterative quantization (ITQ) [9]. We compared the retrieval time and precision to evaluate the performance of the pruning algorithm. The experiment flow is shown in Figure 8.

First, the different hashing methods were utilized to retrieve the similar lung nodule CT images without using the pruning algorithm. Figure 9 shows the retrieval precision of these hashing methods, in which the retrieval result of the KSH method outperforms the other hashing methods.

We then applied the proposed pruning algorithm (without using the decision rule) to these hashing methods and

Table 5: Settings for parameters k and m to reach the retrieval precision at 48 bits.

Clusters k	5	10	15	20	25	30	35	40	45
Candidate clusters m	2	4	5	6	6	7	8	12	16

Table 6: Parameter settings of the different hash methods for the pruning algorithm evaluation.

Parameters	Different hash methods					
	KSH	KLSH	SH	BRE	PCAH	ITQ
Bit	48	40	24	48	16	40
m	8	9	11	7	9	8

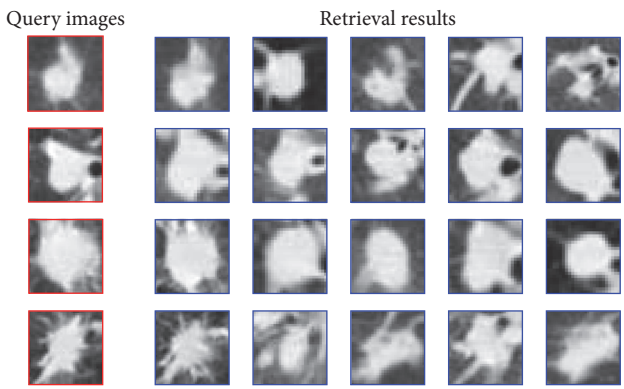


Figure 7: Retrieval results for lung nodule CT images using the proposed method.

validated its performance when the dataset is partitioned into 35 clusters. The parameter settings for the hash code length and number of candidate clusters are shown in Table 6, which differ depending on the highest precision that these hashing methods can reach. Figure 10 shows the retrieval time for all query images for the different hashing methods. We can see that the retrieval speed of these hashing methods using the pruning algorithm is about 2–4 times faster than when these methods do not use the pruning algorithm. Hence, the proposed pruning algorithm clearly reduces the retrieval time of lung nodule CT images.

Moreover, the decision rule designed in the pruning algorithm is also helpful for improving the retrieval precision to some extent. By comparing the distance between the query image and the clusters, the most similar lung image can be returned first. Figure 11 shows the influence of the decision rule on the highest retrieval precision of different hashing methods.

4.4. Evaluation of the Retrieval Framework. Figure 12 compares different classification methods, the support vector machine (SVM), back propagation (BP), and K nearest neighbors (KNN), with our hashing-based method, using the classification accuracy of benign and malignant lesions. In this evaluation, we compared them with respect to overall (all nodules in the test dataset), benign nodule, and malignant nodule accuracy in the test dataset. We judged whether the

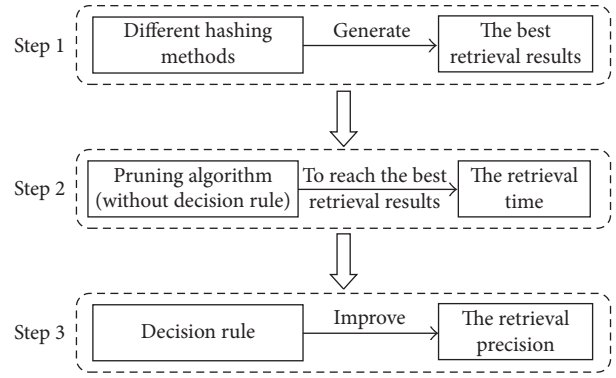


Figure 8: Experiment flow for evaluating the performance of the pruning algorithm.

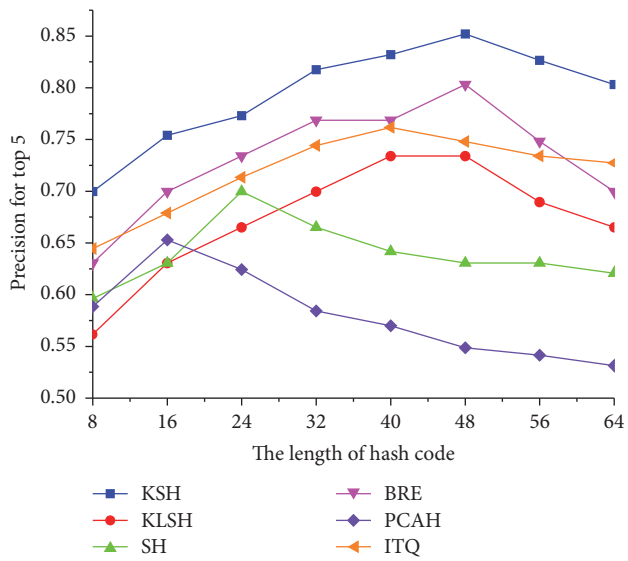


Figure 9: Retrieval precision of different hashing methods for different hash code lengths (without using the pruning algorithm).

query lesion was benign or malignant using the returned similar lung nodule CT images. The judging method is similar to the idea of the KNN algorithm, that is, if the number of benign lung nodules is greater than the number of malignant lung nodules in the returned similar images, the query image is diagnosed as a benign tumor. Otherwise, the query tumor is diagnosed as malignant.

The KNN algorithm is always used as a baseline classical classification method in machine learning. Here, we employ the Euclidean distance to obtain similar samples and set K to 5. However, the calculation is not efficient enough for the high-dimensional image features. Our hashing method leverages the compact hash code and code inner products

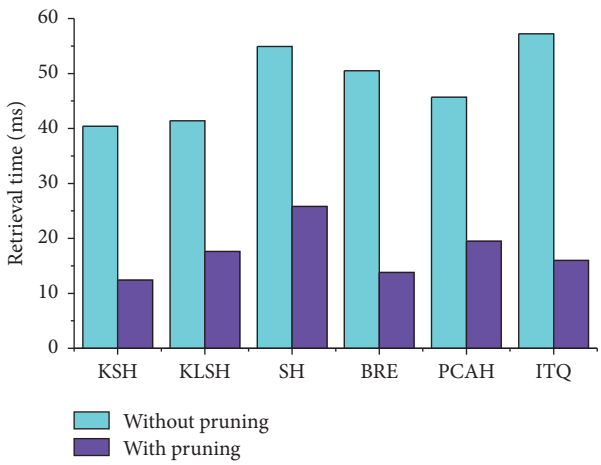


Figure 10: Comparison of different hashing methods with and without the pruning algorithm in terms of retrieval time.

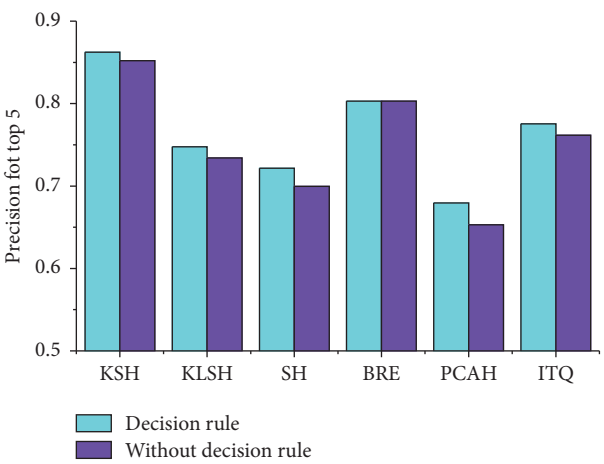


Figure 11: Influence of decision rule on retrieval precision.

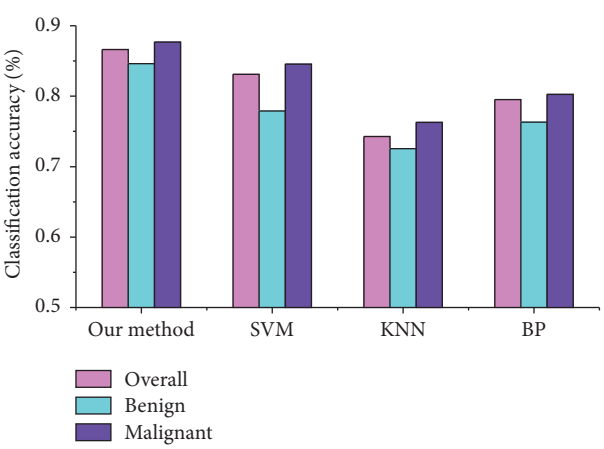


Figure 12: Comparison of the accuracy of different classification methods.

to measure the similarity, which is helpful for improving its efficiency. The BP neural network is one of the most widely used neural network models. The learning rate is set to 0.01, and the number of iterations is 500 in our experiments. A SVM is a supervised learning model that uses supervised information to bridge the semantic gap. Hence, the classification results of SVM with a radial basis function are better than KNN and the BP method.

Furthermore, we can see that our method significantly outperforms the other three classification methods. The overall classification accuracy can reach 86.62%, with accuracies of 84.61% for the benign lesions and 87.67% for the malignant lesions. This improvement illustrates that hash functions with supervised information actually preserve the similarity of the images in the original feature space and validate its retrieval performance.

5. Conclusion

In this paper, in order to improve the retrieval efficiency of lung nodule CT images, we presented a retrieval framework based on the KSH method and a pruning algorithm. Specifically, a clustering method is first used to partition the dataset into several clusters. The KSH method is then utilized to compress the high-dimensional feature vectors into compact hash codes. Finally, a pruning algorithm is employed to narrow the retrieval range and further shorten the retrieval time while improving the retrieval precision. Here, the hash functions are used to map a 104-dimensional image feature of lung nodules into a 48-bit binary code, which, to some extent, reduces the memory space. Low memory cost, fast query speed, and a higher precision demonstrate the suitability of the proposed retrieval framework for lung nodule image retrieval. However, in this paper, we only handle benign and malignant lung nodules, which is a relatively easy task. In future work, the method should be further refined to retrieve lung images at the level of medical signs (such as calcification, lobulation, and speculation) with a higher retrieval precision, helping physicians make reliable diagnostic decisions for clinical cases.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61540007 and 61373100). This work was also supported in part by the State Key Laboratory of Computer Science Open Foundation of China (BUAA-VR-15KF02) and the State Key Laboratory of Virtual Reality Open Foundation of China (BUAA-VR-16KF-13).

References

[1] J. Zhao, G. Ji, X. Han, Y. Qiang, and X. Liao, "An automated pulmonary parenchyma segmentation method based on an

improved region growing algorithmin PET-CT imaging," *Frontiers of Computer Science*, vol. 10, no. 1, pp. 189–200, 2016.

[2] M. C. Oliveira and J. R. Ferreira, "A Bag-of-Tasks approach to speed up the lung nodules retrieval in the BigData age," in *Proceedings of the 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom '13)*, pp. 632–636, Lisbon, Portugal, October 2013.

[3] G. Ng, Y. Song, W. Cai, Y. Zhou, S. Liu, and D. D. Feng, "Hierarchical and binary spatial descriptors for lung nodule image retrieval," in *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '14)*, pp. 6463–6466, August 2014.

[4] P. Aggarwal, R. Vig, and H. K. Sardana, "Semantic and content-based medical image retrieval for lung cancer diagnosis with the inclusion of expert knowledge and proven pathology," in *Proceedings of the IEEE 2nd International Conference on Image Information Processing (ICIIP '13)*, pp. 346–351, Shimla, India, December 2013.

[5] X. Yu, S. Zhang, B. Liu, L. Zhong, and D. N. Metaxas, "Large scale medical image search via unsupervised PCA hashing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '13)*, pp. 393–398, Portland, Ore, USA, June 2013.

[6] J. Wang, H. T. Shen, J. Song et al., "Hashing for similarity search: a survey," https://arxiv.org/abs/1408.2927.

[7] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," in *Proceedings of the 25th International Conference on Very Large Data Bases (VLDB '99)*, pp. 518–529, Morgan Kaufmann, 2000.

[8] Y. Weiss, A. Torralba, and R. Fergus, "Spectral hashing," in *Proceedings of the 22nd Annual Conference on Neural Information Processing Systems (NIPS '08)*, pp. 1753–1760, Vancouver, Canada, December 2008.

[9] Y. Gong, S. Lazebnik, A. Gordo et al., "Iterative quantization: a Procrustean approach to learning binary codes for large-scale image retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 2916–2929, June 2011.

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

[11] B. Kulis and K. Grauman, "Kernelized locality-sensitive hashing for scalable image search," in *Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV '09)*, pp. 2130–2137, IEEE, Kyoto, Japan, October 2009.

[12] B. Kulis and T. Darrell, "Learning to hash with binary reconstructive embeddings," in *Proceedings of the Advances in Neural Information Processing Systems 22 (NIPS '09)*, NIPS Proceedings, pp. 1042–1050, 2009.

[13] W. Liu, J. Wang, R. Ji, Y.-G. Jiang, and S.-F. Chang, "Supervised hashing with kernels," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2074–2081, Providence, RI, USA, June 2012.

[14] J. Liu, S. Zhang, W. Liu, X. Zhang, and D. N. Metaxas, "Scalable mammogram retrieval using anchor graph hashing," in *Proceedings of the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI '14)*, pp. 898–901, IEEE, Beijing, China, May 2014.

[15] M. Jiang, S. Zhang, J. Huang, L. Yang, and D. N. Metaxas, "Joint kernel-based supervised hashing for scalable histopathological image analysis," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., vol. 9351 of *Lecture Notes in Computer Science*, pp. 366–373, Springer, Berlin, Germany, 2015.

[16] X. Zhang, W. Liu, and S. Zhang, "Mining histopathological images via hashing-based scalable image retrieval," in *Proceedings of the IEEE 11th International Symposium on Biomedical Imaging (ISBI '14)*, pp. 1111–1114, Beijing, China, April 2014.

[17] T. Aggarwal, A. Furqan, and K. Kalra, "Feature extraction and LDA based classification of lung nodules in chest CT scan images," in *Proceedings of the IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI '15)*, pp. 1189–1193, Kochi, India, August 2015.

[18] H. M. Orozco, O. V. Villegas, G. C. Sánchez, J. O. Domínguez, and J. N. Alfaro, "Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine," *BioMedical Engineering Online*, vol. 14, no. 1, article 9, 2014.

[19] J. Zhao, G. Ji, Y. Qiang, X. Han, B. Pei, and Z. Shi, "A new method of detecting pulmonary nodules with PET/CT based on an improved watershed algorithm," *PLoS ONE*, vol. 10, no. 4, article e123694, 2015.

[20] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[21] D. Hamad and P. Biela, "Introduction to spectral clustering," in *Proceedings of the International Conference on Information and Communication Technologies: From Theory To Applications*, pp. 1–6, April 2008.

[22] S. G. Armato III, G. McLennan, L. Bidaut et al., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 3530251, 7 pages
http://dx.doi.org/10.1155/2016/3530251

Research Article

Automated Segmentation of Coronary Arteries Based on Statistical Region Growing and Heuristic Decision Method

Yun Tian,[1] Yutong Pan,[1] Fuqing Duan,[1] Shifeng Zhao,[1] Qingjun Wang,[2] and Wei Wang[3]

[1] *College of Information Science and Technology, Beijing Normal University, Beijing 100875, China*
[2] *Department of Radiology, Navy General Hospital, Beijing 100048, China*
[3] *Peking University Cancer Hospital & Institute, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Beijing 100142, China*

Correspondence should be addressed to Wei Wang; tyun0619@163.com

Received 17 July 2016; Accepted 26 September 2016

Academic Editor: Yong Xia

The segmentation of coronary arteries is a vital process that helps cardiovascular radiologists detect and quantify stenosis. In this paper, we propose a fully automated coronary artery segmentation from cardiac data volume. The method is built on a statistics region growing together with a heuristic decision. First, the heart region is extracted using a multi-atlas-based approach. Second, the vessel structures are enhanced via a 3D multiscale line filter. Next, seed points are detected automatically through a threshold preprocessing and a subsequent morphological operation. Based on the set of detected seed points, a statistics-based region growing is applied. Finally, results are obtained by setting conservative parameters. A heuristic decision method is then used to obtain the desired result automatically because parameters in region growing vary in different patients, and the segmentation requires full automation. The experiments are carried out on a dataset that includes eight-patient multivendor cardiac computed tomography angiography (CTA) volume data. The DICE similarity index, mean distance, and Hausdorff distance metrics are employed to compare the proposed algorithm with two state-of-the-art methods. Experimental results indicate that the proposed algorithm is capable of performing complete, robust, and accurate extraction of coronary arteries.

1. Introduction

Over the past decades, coronary artery disease (CAD) has been the main cause of human deaths in the world [1]. Many factors can lead to CAD, and, of these, stenosis caused by atherosclerosis is the most common. Coronary arteries are usually extracted first to diagnose stenosis. An inaccurate segmentation of coronary arteries can result in fatal false treatments because a missing segment or mixed extraction of other structures can lead to the oversight of existing stenosis or improper narrow lumen detections.

Many studies have been conducted on automated or semiautomated segmentation of coronary arteries on computed tomography angiography (CTA) images. Automated segmentation methods can automatically segment regions of interest of images, without any human intervention. However, the complexity of such methods is usually relatively high [2, 3]. Compared with automated segmentation methods, semiautomated segmentation methods [4, 5] require the interactions of therapists, making the methods less convenient than automated methods. However, the performances of semiautomated methods are sometimes better than those of the automated methods. Öksüz et al. [4] proposed a hybrid method composed of threshold preprocessing, vessel enhancement, and traditional 3D region growing. Threshold preprocessing in this method retains many uninterested regions. In addition, traditional 3D region growing requires human interactions and lacks accuracy. Thus, the precision and automation of this method must be improved.

Kitamura et al. [5] proposed a novel coronary segmentation method based on multilabel graph cuts; this method utilizes higher-order potentials to impose shape priors. However, the method derives a limitation from the Hessian-based features, which cannot distinguish a very wide

variety of structures. In addition, the method requires human interactions. Research on automated segmentation methods has been making great progress recently, and automated methods are gradually becoming popular in clinical diagnosis. However, automated methods have problems in both efficiency and accuracy. To solve this problem, Lugauer et al. [6] proposed a lumen segmentation method for coronary CTA. They utilized a Markov random field formulation with convex priors that rely on the training of a large dataset. Their method is sensitive to the dataset and its training and also proved to be inefficient because the dataset analysis is time-consuming. Zheng et al. [7] proposed a machine learning-based method that can gain better performance than empirically designed measurement (e.g., the widely used Hessian-based method), but the learning process requires the analysis of a large expert-annotated dataset, which is also very time-consuming. Zhou et al. [8] presented a method following the steps of heart region extraction, multiscale coronary artery response method for vascular structure enhancement, automated detection of seed points, and 3D dynamic balloon-tracking method for coronary arteries tracking. However, the EM algorithm used in the heart region extraction has low efficiency because of the huge amount of points in the CTA volume. Meanwhile, Bouraoui et al. [9] introduced an automated method based on advanced mathematical morphology techniques. They employed a blurry grey-level hit-or-miss transform method to detect seed points automatically. However, the 13 structure candidates employed in their method occasionally fail to detect seed points because they cannot cover all patient conditions. Many of the works mentioned above use region growing as a part of the segmentation process, and these region growing methods are not fully automated and lack robustness and accuracy. In addition, the precision of coronary artery extraction needs to be improved.

In this paper, we present a novel scheme for extracting coronary arteries from CTA images. First, the heart region is extracted using a multi-atlas-based approach. Second, the vessel structures are enhanced via a 3D multiscale line filter. Next, a set of seed points is detected automatically through threshold preprocessing and a subsequent morphological operation. Based on the set of seed points, statistics-based region growing is applied. Upon generating the results by setting conservative parameters, a heuristic decision method is used to obtain the desired result. Each step of the algorithm is fully automated for an efficient and automated pipeline. The proposed algorithm outperforms two state-of-the-art methods, which are used for comparison. The experiments are carried out on eight-patient multicenter multivendor cardiac CTA volume data, and the DICE similarity index, mean distance, and Hausdorff distance metrics are employed to compare the different methods.

The rest of the paper is organized as follows. We introduce two related works in Section 2 to be compared with our proposed method. Section 3 describes the proposed method in detail. We validate the proposed method in Section 4 through the well-known evaluation methodology mentioned above. Conclusions are provided in Section 5.

2. Related Works

2.1. Öksüz et al. The algorithm proposed by Öksüz et al. [4] consists of five stages, of which three are related to coronary artery segmentation. First, pulmonary vessels are removed by performing the thresholding and morphological procedures. Then, Frangi vesselness filter [10] is applied on the processed data. Finally, vessel segmentation is achieved by successively performing 3D region growing and fast marching.

The Frangi vesselness filter used in their algorithm is based on the eigenvalues of the Hessian matrix:

$$H = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix}, \tag{1}$$

where I_{xx}, I_{zy}, and so forth are the partial second derivatives of image $I(x, y, z)$. An ideal bright 3D line is defined as

$$I(x, y, z) = \exp\left(\frac{-\left(x^2 + y^2\right)}{2\sigma^2}\right), \tag{2}$$

where the direction of the maximum second derivative is identical to the direction of the z-axis (i.e., the direction of the line), and its value is zero. Any second derivative orthogonal to the z-axis has a negative value at each point in the central region of the line cross section. Ideally, the conditions of a bright line can be regarded as $\lambda_1 \cong 0$, $\lambda_2 \cong \lambda_3 \ll 0$.

In the method of Öksüz et al., the pulmonary vessels may not be completely removed by simply setting the thresholding and morphological operations. Furthermore, the region growing in their method is not fully automated.

2.2. Zhou et al. The algorithm proposed by Zhou et al. [8] consists of four stages. First, the heart region is extracted using the EM algorithm. Then, a 3D multiscale filter is applied to the heart region. Next, seed points are detected automatically on the processed data. Finally, a dynamic balloon-tracking algorithm is applied to track the coronary arteries.

The idea of this method is similar with our algorithm. However, the EM algorithm employed in the heart extraction is time-consuming, which is unacceptable in the clinical diagnosis process. Moreover, multi-atlas-based heart extraction is more precise than the EM algorithm. The multi-atlas-based method is based on the registration of the standard images to obtain the optimal parameters, which determine the deformation of the reference images to form the extracted goal. Although this method requires also a large amount of computation, the Gaussian pyramid model can be employed to improve the computational efficiency. The dynamic balloon tracking can also terminate in the coronary segments where high-grade stenosis exists. In comparison, statistical region growing, based on a group of seed points, can handle this situation. As a result, our method performs better in the diseased coronary segments.

3. Proposed Methodology

The proposed method, which can robustly and automatically extract complete coronary arteries, has four steps as follows: (1) heart segmentation, (2) vessel enhancement, (3) extraction of seed points, and (4) coronary artery extraction.

3.1. Segmentation of the Heart Region. We apply a two-stage registration method denoted as an optimization problem searching for an optimal transformation T, which minimizes the dissimilarity between a fixed image $I_F(x)$ and a moving image $I_M(x)$:

$$\widehat{T} = \arg\min C\left(I_F(x), I_M(T(x))\right). \quad (3)$$

In the equation above, C is a cost function that measures the dissimilarity between two images.

In the first stage, an affine registration is used to spatially align the fixed and moving images roughly. The cost function is defined as

$$C\left(C, I_F, I_M\right) = \frac{1}{|\sigma_F|} \sum_{x_i \in \sigma_F} \left[I_F(x_i) - I_M(T_C(x_i))\right]^2, \quad (4)$$

where σ_F is the fixed image domain. This equation is calculated using image samples that are randomly chosen in each iteration in the entire image domain. Then, 256 iterations of the gradient descent optimizer are performed.

In the second stage, a B-spline registration is utilized. The result of the affine registration is used as the initialization of B-spline registration. In each iteration, one voxel is selected randomly in the entire image domain. The remaining image samples are picked in a 50 mm square neighborhood around that voxel. The algorithm optimizes the localized similarity measure using the equation

$$L\left(C, I_F, I_M\right) = \sum_{m \in L_M} \sum_{f \in L_F} \left(p(f, m; C) \cdot \log_2 \left(\frac{p(f, m; C)}{p_F(f)\, p_M(m; C)} \right) \right), \quad (5)$$

where L_F and L_M are sets of regularly spaced intensity bin centers, p is the discrete joint probability, and p_F and p_M are the marginal discrete probabilities of the fixed and moving images obtained by summing p over f and m, respectively.

Furthermore, a Gaussian pyramid model is applied during the registration to improve the computational efficiency. The transformation parameters calculated in (1) are propagated to label the regions that must be segmented to the patient images. Finally, majority voting is used to combine the labeled atlas images and generate the final heart segmentation results. The atlas images employed in our method are similar to those used in [11].

3.2. Vessel Structures Enhancements. After the heart region is extracted, a 3D multiscale line filter, previously described by Sato et al. [12], is applied to segment the curvilinear structures in the heart region. The method used here is similar to the Frangi vesselness filter, which has been introduced in Section 2.1 (i.e., the method proposed by Öksüz et al.).

3.3. Automated Seed Points Detection. As coronary arteries are enhanced in CTA images, a threshold operator plays an important role in noncardiovascular removal. The threshold value used in this process should be conservative in consideration of the various pathologies of different patients. Thus, in this study, a threshold of 120 HU is used in our method, and a 3D erode filter is employed to suppress noise points. The kernel of the erosion process should be sufficiently large to ensure that the remaining points belong to coronary arteries. In our method, the erosion kernel is defined as a 4×4×3 cube. Therefore, a point set is attached to the coronary arteries, wherein all of the points are to be applied in the next step.

3.4. Coronary Artery Extraction. By using all the previously detected seed points, coronary arteries are extracted using a statistics-based growing method, which identifies voxels with similar statistics via connectivity. The method is based on the iterative computation of the statistical information of voxel intensities included in the current region. As for each seed point, the mean and variance across a 26-connected neighborhood are calculated to define a range:

$$I = [m - v * d,\ m + v * d], \quad (6)$$

where I represents the consistent interval; m and d stand for the mean and standard deviation of the seed point, respectively; and v is a bounds control parameter, which is automatically set through the heuristic decision manner. 26-neighborhood voxels with intensities within this range are included in the region. If the seed points were included in the growing region, they are removed from the seed point set. Thus, the growing efficiency can be improved. It is noteworthy that some parts of coronary arteries may not be included by the region growing of a seed point, but they may be included by the region growing of other seed points. Therefore, using a seed set to carry out the statistics-based region growing, not only can the robustness of the proposed algorithm be enhanced, but also the integrity of the segmented coronary arteries can be improved.

After this initial segmentation is calculated, the mean and variance, of all voxels from the current segmented coronary arteries, are computed again to define a new intensity range. The intensity range is used to detect whether intensities of the voxels in the neighborhood of the current segmentation fall within the range. If a voxel's intensity is within the interval, the voxel is included in the segmented arteries. This process is repeated until no new voxels are added or a specified number of iterations are reached. In this study, the number of iterations was set to 5.

Whenever a seed point is provided, the mean and standard deviation can be calculated automatically. However, v cannot be automatically obtained. Hence, a heuristic decision method is presented to acquire v without human interactions, and its flow chart is shown in Figure 1. The process is described below.

(1) Initialize v_0 to 1.0 and make i (set to 0 at first) refer to the number of times the method should be repeated.

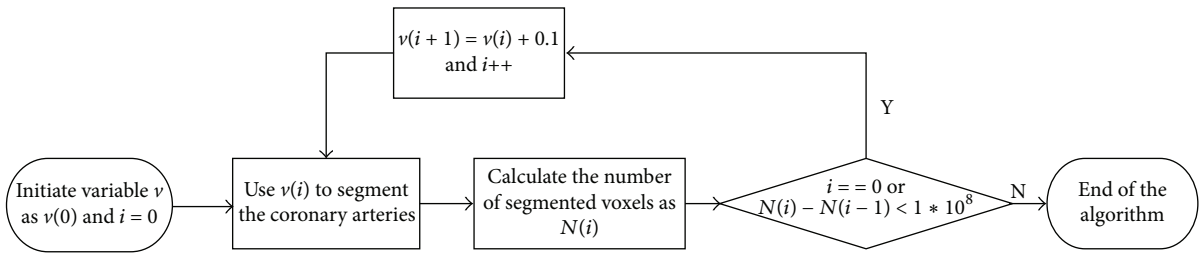


Figure 1: Flow chart of the statistics-based method.

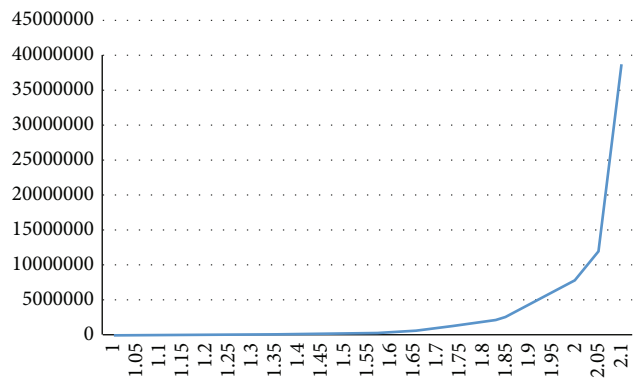


Figure 2: Sketch map of the mutation between the desired and redundant results (the horizontal axis represents the incremental variable v, and the vertical axis represents the number of the segmented voxels).

(2) Perform segmentation using v_i and calculate N_i (i.e., the number of segmented points). If N_i is calculated for the first time, go to step (4); else continue to step (3).

(3) Compare N_i with N_{i-1}. If the difference between them is over $1 * 10^8$, the algorithm is ended.

(4) Gradually increase v_i by a step of 0.1; go to step (2).

This heuristic decision method is based on the fact that a mutation between the desirable and redundant segmentations exists, as shown in Figure 2. Specifically, before the advent of oversegmentation, the segmented results of each increment of v, compared with that of the previous v, the change of the number of the segmented voxels is less than 1×10^7 orders of magnitude. If the change of the number of the segmented voxels is greater than 1×10^8, this means "mutation." The reason is that the redundant segmentation almost comprises the whole heart, whereas the coronary arteries are relatively small parts of the heart region.

4. Result and Discussion

4.1. Data. CTA provides a visualization of the whole chest, including vessel lumen, atherosclerotic, and stenosis, without the invasive catheterization procedure [13]. Thus, CTA is less harmful compared with the traditional 2D X-ray angiography because only the contrast medium is required to be injected before proceeding to CTA. Meanwhile, its 3D reconstruction capability is highly suitable for the treatment of CAD.

Table 1: CT acquisition parameters.

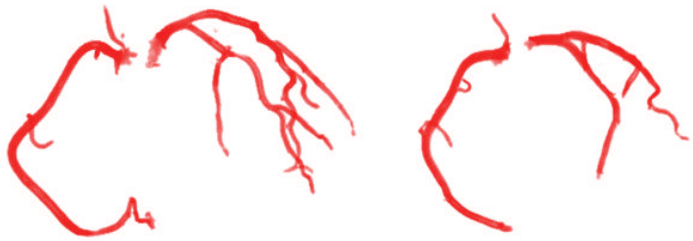
Parameter name	Value
Voxel spacing	$0.33 \times 0.33 \times 0.4\ \text{mm}^3$
Resolution	512×512 voxels/slice
Slice thickness	0.8 mm
Tube voltage	120 kV
Exposure time	1833 ms
Series description	75%
Table height	89 mm

We employ CTA images as materials in the testing and development of our method. The detailed acquisition parameters of CTA are shown in Table 1.

4.2. Result and Validation. The validation process of the proposed method was built based on the publicly accessible standardized coronary artery evaluation framework presented by Kirişli et al. [14]. The method presented in this paper was implemented using C++ within the VolView open-source platform (http://www.kitware.com/opensource/volview.html). Figure 3 shows images of the left and right coronary arteries extracted by our method from two-patient volume CTA data.

In the experiment, two state-of-the-art methods (those presented by Öksüz et al. [4] and Zhou et al. [8]) were compared to our method using the volumetric overlap (DICE) and max/mean surface distance (MAXSD/MSD) as metrics. DICE is the dice similarity coefficient, which is used to evaluate the volumetric cardinality of different algorithms. Table 2 lists the DICE, MAXSD, and MSD values calculated separately for healthy and diseased arteries over a dataset of eight patients with different grades of stenosis (D and H denote diseased and healthy vessels, resp.). As can be seen, the proposed method performs best on both healthy and diseased vessel segments for DICE and MAXSD metrics. Furthermore, the method achieved the highest rank compared to the two state-of-the-art methods.

Visual results and segmentation comparisons between the proposed method and one of the two state-of-the-art methods are shown in Figure 4. The first row of Figure 4 shows the comparison between our method and that proposed by Öksüz et al. [4] using four patients with different



(a) (b)

Figure 3: Extraction results by our method from different patient volume CTA data. (a) A patient who suffers from low-grade stenosis in both the right coronary artery (RCA) and the left coronary artery (LCA). (b) A patient who suffers from high-grade stenosis in the LCA and low-grade stenosis in the RCA.

Table 2: Comparison of the segmentation results (D and H denote diseased and healthy vessels, resp.).

Method	DICE D [%]	DICE H [%]	MSD D [mm]	MSD H [mm]	MAXSD D [mm]	MAXSD H [mm]	Rank avg.
Proposed method	**0.71**	**0.76**	0.34	0.41	**2.47**	**2.75**	**4.2**
Öksüz et al. [4]	0.60	0.68	0.45	0.55	3.94	6.48	6.9
Zhou et al. [8]	0.69	0.72	**0.32**	**0.39**	2.87	3.20	4.4

stenosis in both the LCA and RCA. The second row of Figure 4 shows the comparison between our method and that proposed by Zhou et al. [8] using the same dataset as that of the first row. Each dataset result from the different comparisons is shown in every column. The red vessels in Figure 4 are common segments extracted by both the proposed and state-of-the-art methods, and the green vessels are the segments extracted by the proposed method and missed by the traditional methods. It can be seen from the first row of Figure 4 that Öksüz's method is only capable of extracting the thick vessels, which is not connective. The reason is that Öksüz's method employs the thresholding and morphological operations to extract arteries based on the globe information of volume data, which fails to cope with the coronary arteries with different intensities. It can be seen from the second row of Figure 4 that Zhou's method obtains a better result than that of Öksüz's method and is capable of extracting most coronary arteries. However, some weak arteries are still missing, as shown in the first column of the second row in Figure 4. The reason is that the artery segmentation is terminated in advance by the dynamic tracking balloon where there is a high-grade stenosis. It is clearly seen from Figure 4 that our method is capable of accurately extracting the thick and thin arteries, and they are red and green. At the same time, the segmented coronary arteries appear with better connectivity. The excellent performance of our method is attributed to the automated seed points detection and the elaborately exploited statistics-based region growing.

Figure 5 shows the difference between the segmentation methods and gold standards derived from the manual segmentation of expert radiologists in Beijing's Navy General Hospital to ensure the validity and authenticity of our experiment. In Figure 5 the common parts are painted red, and the differences are painted green. Two different patients with low- and high-grade stenosis were employed as samples in the comparison of each method with the gold standard. Figure 5(a) shows our method and the gold standard using a sample patient with low-grade stenosis. Figure 5(b) shows our method and the gold standard using a sample patient with high-grade of stenosis. Figure 5(c) shows the method proposed by Öksüz et al. [4] and the gold standard using a sample patient with low-grade stenosis. Figure 5(d) shows the method proposed by Öksüz et al. [4] and the gold standard using a sample patient with low-grade stenosis. Figure 5(e) shows the method proposed by Zhou et al. [8] and the gold standard using a sample patient with low-grade stenosis. Figure 5(f) shows the method proposed by Zhou et al. [8] and the gold standard using a sample patient with high-grade stenosis. As can be seen from Figure 5, our method yielded a robust segmentation for each volume data. This can be attributed to the fact that a segment missing by one seed point in the proposed method can be supplemented by other seed points in the subsequent growing process because of the employment of a novel seed point detection method, and the heuristic decision that determines the desired results automatically ensures the absence of human interaction in the pipeline. The detailed information on the equipment used in our experiment is as follows: Intel® Core™ i7-3770 CPU, 8.0 GB RAM, 64-bit Ubuntu 14.04. The time consumed by the proposed method was less than 60 seconds for each set of volume data.

5. Conclusion

We proposed a fully automated segmentation of coronary arteries from 3D cardiac CTA (CCTA). This algorithm is a combination of various technologies, including multi-atlas, multiscale vascular enhancement, morphological, and

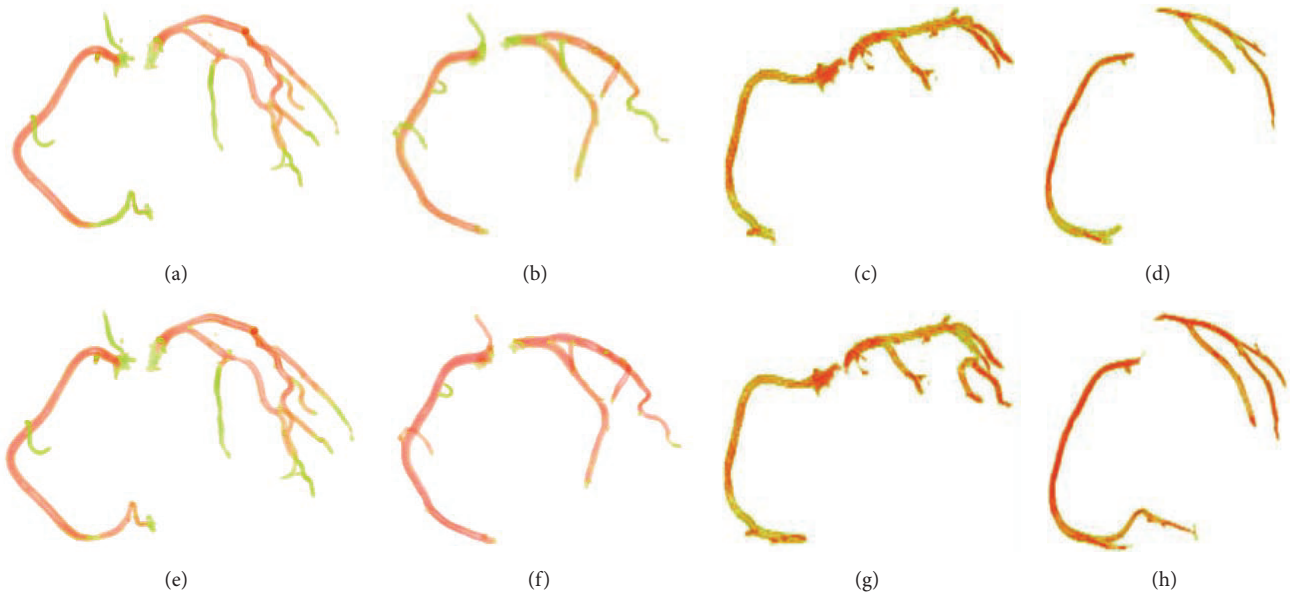


Figure 4: Comparison of our method with Öksüz's method [4] and Zhou's method [8] using four patients with different stenosis. Rows 1 and 2 show the comparisons between our method and Öksüz's method [4] and Zhou's method [8], respectively. Each dataset result from the different comparisons is shown in every column.

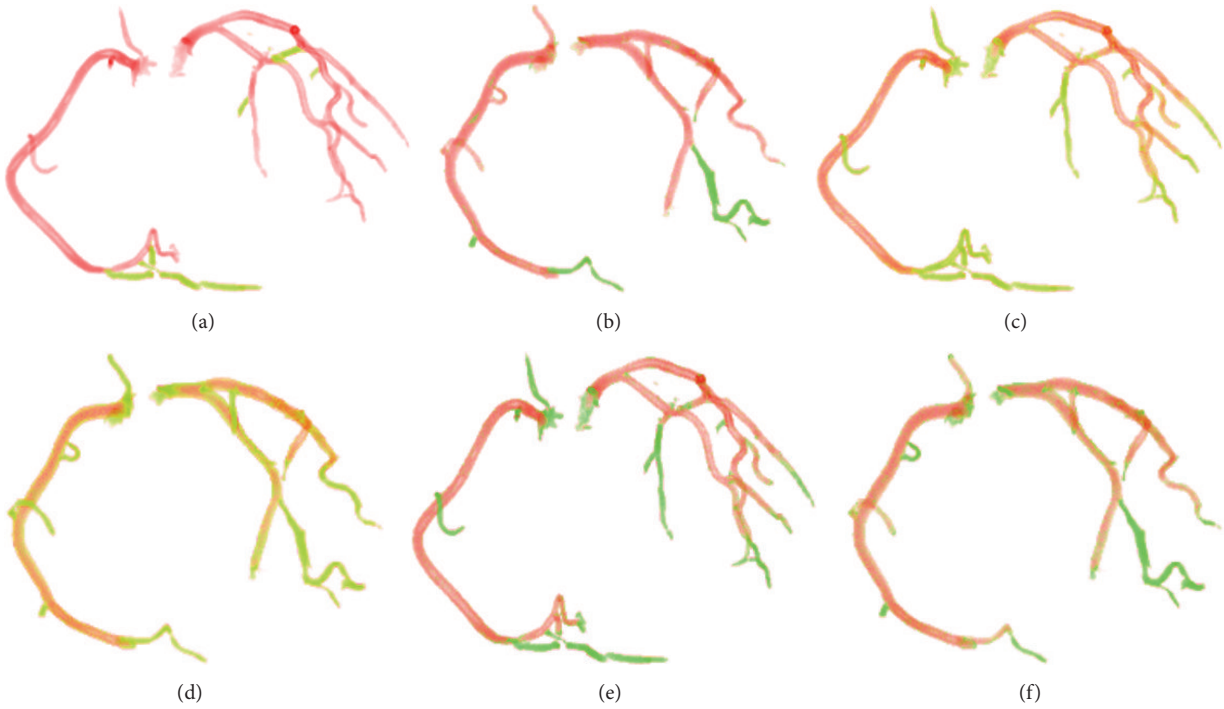


Figure 5: Comparison results between each method and the gold standard. (a) Our method and the gold standard using a sample patient with low-grade stenosis. (b) Our method and the gold standard using a sample patient with high-grade stenosis. (c) The method proposed by Öksüz et al. [4] and the gold standard using a sample patient with low-grade stenosis. (d) The method proposed by Öksüz et al. [4] and the gold standard using a sample patient with low-grade stenosis. (e) The method proposed by Zhou et al. [8] and the gold standard using a sample patient with low-grade stenosis. (f) The method proposed by Zhou et al. [8] and the gold standard using a sample patient with high-grade stenosis.

statistical decision technologies. First, the heart region is extracted by a multi-atlas-based approach, which is capable of segmenting the heart due to the employed prior knowledge from the reference heart atlas set. At the same time, the registration efficiency is improved by a Gaussian pyramid model employed. Then, a 3D multiscale vessel filter is applied to enhance the coronary artery structures, which effectively enhances the coronary artery contrast, because the shape information of the blood vessel is considered. Subsequently, seed points are detected automatically through threshold preprocessing and a morphological operation. Based on the set of seed points, statistics-based region growing is applied, which grows the coronary arteries in virtue of the local statistical information of the seek point. Thus, the connectivity of the arteries can be well guaranteed. Finally, results are obtained by setting conservative parameters, and then a heuristic decision is employed to automatically obtain the desired result. The coronary arteries are gradually segmented from the CCTA data in a coarse-to-fine manner. No manual interaction is involved in the entire segmentation process because each required parameter is searched using novel specific algorithms. Hence, the proposed algorithm is capable of performing complete, robust, and accurate extraction of coronary arteries.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work is supported by the National Science Foundation of China (Grant no. 61472042), the Beijing Natural Science Foundation (Grant no. 4152027), the Fundamental Research Funds for the Central Universities (no. 2015KJJCB25), and the Program for New Century Excellent Talents in University (NCET-13-0051).

References

[1] R. J. Bache, "Coronary artery disease: regulation of coronary blood flow," in *Coronary Artery Disease*, pp. 57–67, Springer, Berlin, Germany, 2015.

[2] C. Zhou, H.-P. Chan, A. Chughtai et al., "Automated coronary artery tree extraction in coronary CT angiography using a multiscale enhancement and dynamic balloon tracking (MSCAR-DBT) method," *Computerized Medical Imaging and Graphics*, vol. 36, no. 1, pp. 1–10, 2012.

[3] R. Shahzad, H. Kirişli, C. Metz et al., "Automatic segmentation, detection and quantification of coronary artery stenoses on CTA," *International Journal of Cardiovascular Imaging*, vol. 29, no. 8, pp. 1847–1859, 2013.

[4] İ. Öksüz, D. Ünay, and K. Kadıpaşaoğlu, "A hybrid method for coronary artery stenoses detection and quantification in CTA images," in *Proceedings of the MICCAI Workshop 3D Cardiovascular Imaging: A MICCAI Segmentation*, 2012.

[5] Y. Kitamura, Y. Li, W. Ito, and H. Ishikawa, "Coronary lumen and plaque segmentation from CTA using higher-order shape prior," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014*, P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, Eds., vol. 8673 of *Lecture Notes in Computer Science*, pp. 339–347, Springer, Berlin, Germany, 2014.

[6] F. Lugauer, Y. Zheng, J. Hornegger, and B. M. Kelm, "Precise lumen segmentation in coronary computed tomography angiography," *Medical Computer Vision: Algorithms for Big Data*, vol. 8848, pp. 137–147, 2014.

[7] Y. Zheng, M. Loziczonek, B. Georgescu, S. K. Zhou, F. Vega-Higuera, and D. Comaniciu, "Machine learning based vesselness measurement for coronary artery segmentation in cardiac CT volumes," in *Proceedings of the Medical Imaging 2011: Image Processing*, vol. 7962, 2011.

[8] C. Zhou, H.-P. Chan, A. Chughtai et al., "Automatic identification of origins of left and right coronary arteries in CT angiography for coronary arterial tree tracking and plaque detection," in *Proceedings of the Medical Imaging 2013: Computer-Aided Diagnosis*, vol. 8670 of *Proceedings of SPIE*, Medical Imaging 2013, Lake Buena Vista, Fla, USA, February 2013.

[9] B. Bouraoui, C. Ronse, J. Baruthio, N. Passat, and P. Germain, "3D segmentation of coronary arteries based on advanced mathematical morphology techniques," *Computerized Medical Imaging and Graphics*, vol. 34, no. 5, pp. 377–387, 2010.

[10] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Interventation—MICCAI '98*, W. M. Wells, A. Colchester, and S. L. Delp, Eds., vol. 1496 of *Lecture Notes in Computer Science*, pp. 130–137, Springer, Berlin, Germany, 1998.

[11] H. A. Kirisli, M. Schaap, S. Klein et al., "Fully automatic cardiac segmentation from 3D CTA data: a multi-atlas based approach," in *Proceedings of the Medical Imaging 2010: Image Processing*, B. M. Dawant and D. R. Haynor, Eds., Proceedings of SPIE, San Diego, Calif, USA, February 2010.

[12] Y. Sato, S. Nakajima, N. Shiraga et al., "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Medical Image Analysis*, vol. 2, no. 2, pp. 143–168, 1998.

[13] S. Motoyama, H. Ito, M. Sarai et al., "Plaque characterization by coronary computed tomography angiography and the likelihood of acute coronary events in mid-term follow-up," *Journal of the American College of Cardiology*, vol. 66, no. 4, pp. 337–346, 2015.

[14] H. A. Kirişli, M. Schaap, C. T. Metz et al., "Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography," *Medical Image Analysis*, vol. 17, no. 8, pp. 859–876, 2013.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 4674658, 12 pages
http://dx.doi.org/10.1155/2016/4674658

Research Article

DTI Image Registration under Probabilistic Fiber Bundles Tractography Learning

Zhe Guo,[1] Yi Wang,[1] Tao Lei,[2,3] Yangyu Fan,[1] and Xiuwei Zhang[2]

[1] *School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China*
[2] *School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China*
[3] *College of Electrical & Information Engineering, Shaanxi University of Science & Technology, Xi'an 710021, China*

Correspondence should be addressed to Zhe Guo; guozhe@nwpu.edu.cn

Received 7 July 2016; Accepted 30 August 2016

Academic Editor: Andrey Krylov

Diffusion Tensor Imaging (DTI) image registration is an essential step for diffusion tensor image analysis. Most of the fiber bundle based registration algorithms use deterministic fiber tracking technique to get the white matter fiber bundles, which will be affected by the noise and volume. In order to overcome the above problem, we proposed a Diffusion Tensor Imaging image registration method under probabilistic fiber bundles tractography learning. Probabilistic tractography technique can more reasonably trace to the structure of the nerve fibers. The residual error estimation step in active sample selection learning is improved by modifying the residual error model using finite sample set. The calculated deformation field is then registered on the DTI images. The results of our proposed registration method are compared with 6 state-of-the-art DTI image registration methods under visualization and 3 quantitative evaluation standards. The experimental results show that our proposed method has a good comprehensive performance.

1. Introduction

Diffusion Tensor Imaging (DTI) is a Magnetic Resonance Imaging (MRI) technique which measures diffusion properties of water molecules in tissue to gained neural bundle images, which cannot be obtained by other imaging modalities [1]. It captures vital information that import for vivo investigation of white matter and connectivity alterations, thus playing an increasingly significant role *in vivo* studies of anatomical structure and functional connectivity in the brain regions [2]. DTI image registration is an essential step for diffusion tensor image analysis. DTI registration is involved in many clinical diagnoses of disease diffusion tensor image analysis; all need image registration techniques [3]. For ordinary medical image registration, the corresponding points of two images will be transformed to have the exact consistency on the space position and the anatomical structure by space transformation; the registration process is essentially a multiparameter optimization problem [4]. Tensor image registration will encounter many ordinary medical image registration problems but also includes some special difficulties due to the particularity of the DTI data.

According to the object of the registration algorithm, the existing DTI image registration algorithms can be divided into three categories: the scalar image based registration algorithm, the tensor image based registration algorithm, and the fiber bundle based registration algorithm [5]. The scalar image based registration algorithm has low computational complexity; however, as this algorithm does not make full use of all the directions and structure information of the DTI images, it will lose some important data in the registration process. The tensor image based registration algorithm should ensure the consistency of the tensor direction and anatomical structure before and after the transformation. Fiber bundle based registration algorithm directly uses white matter fiber bundles for registration and can avoid the estimation error of DTI direction and hence improve accuracy and robustness of registration. Therefore, in recent years, this kind of registration algorithm becomes the mainstream of diffusion tensor magnetic resonance image registration method.

In the fiber bundle based registration method, fiber tracking is a very important step, and the correctness of fiber tracking directly affects the accuracy of the registration [6]. The neural fiber tracking technique based on DTI can

be roughly divided into two categories: deterministic fiber tracking technique and probabilistic fiber tracking technique. As the diffusion tensor of the voxel is sensitive to noise, the result of the deterministic fiber tracking will be affected by the noise also. Furthermore, due to the effect of volume, there is more than one fiber beam in the unit voxel. For the area including two or multiple nerves and fibers cross through, the accuracy of deterministic fiber tracking is not high [7]. Because of the introduction of probability statistics method, the probabilistic fiber bundle tracking technique can solve the problem of volume effect and noise interference [8].

Probabilistic tractography uses a deterministic streamline algorithm to generate thousands of trajectories by Monte Carlo methods. The directions of the line segments are repeatedly sampled from a Bayesian posterior distribution [8]. The probability of a trajectory to the selected sample voxel is then defined as the number of virtual fibers passing a voxel. Probability distribution based on *a priori* assumptions about the form of the uncertainty in the data is used in most probabilistic tractography methods. Nevertheless, since a parametric description of subject artifacts is generally unavailable, the uncertainty is modeled without considering the artifacts interference.

To resolve this disadvantage, a typical active sample selection learning method called bootstrap [9] has been incorporated in probabilistic tractography [10]. Bootstrap method is nonparametric procedure which assesses the measurement uncertainty of parameters without the assumption of a noise model and the acquirement of large amount of datasets [11]. Consequently, the local directions are derived by resampling from the acquired data itself instead of a probability distribution.

Based on the above analysis, in this paper, we proposed a DTI image registration method under probabilistic fiber bundles tractography learning. We improve the residual error estimation step in bootstrap method used in active sample selection learning for the probabilistic tractography. Our method assumes that, in the case of independent and identically distributed error, the residuals can be adjusted and the error model is then modified by using finite sample set, therefore, improving the study ability of samples. Subsequently, the tracked fiber bundles can be registered by using symmetric image standardization registration algorithm. The results of our proposed registration method are compared with 6 state-of-the-art DTI image registration methods under visualization and 3 quantitative evaluation standards [12] for the comprehensive analysis. The experimental results show that our proposed registration method under probabilistic fiber bundles tractography learning has a good comprehensive performance.

2. Related Works

DTI registration methods are mainly divided into three parts by the processing object: the scalar image based registration algorithm, the tensor image based registration algorithm, and the fiber bundle based registration algorithm.

2.1. Scalar Image Based Registration. Scalar image based registration methods convert tensor images into scalar images, for example, fractional anisotropy (FA) images, by rotationally invariant measures, and then perform registration on the scalar images. Studholme et al. [13] proposed the rigid registration method (denoted as Rigid), which was based on normalized mutual information. This method was commonly used in therapy planning, clinical diagnosis, and automatic clinical image registration as a rough registration. Multiresolution elastic matching algorithm (denoted as Elastic) [14] and multiresolution B-spline method [15] were proposed successively and applied to the registration of diffusion tensor images; the latter was proved to have the high geometric fidelity [16]. Consequently, Andersson et al. [17] developed B-spline registration based on sum-of-squared differences (denoted as FSL), the regularization that was based on membrane energy, and a multiscale Leven berg-Marquardt minimization avoided the local minimum value. Soon after, affine image coregistration technique (denoted as Affine) [18] was performed in some cases to align images before the application of higher order registration. In 2008, the literature [19] discussed normalized mutual information criterion, in which the symmetrized Kullback-Leibler divergence was used to improve fluid registration of diffusion tensor images. This algorithm was diffeomorphic and reversible consistency but performed badly on smoothness and was time-consuming. Recently, Hufnagel et al. [20] mentioned the block-matching algorithm in his article. In this method, the determined sparse displacement vector field was used for nonlinear transformation parameters estimation.

Diffeomorphic mapping is a smooth spatial transform, in which the topology of the images is preserved, as well as voxel correspondence based on the second-order tensor field of Riemannian manifold. It was combined with Lie group [21, 22] structure to perform relatively simple calculation. Without the Riemannian manifold measure, directly computing differences between tensors with Euclidean space would lead to the "tensor swelling effect" [23] and could not guarantee reversible consistency of the transformation. Diffeomorphic mapping can avoid the "tensor swelling effect," guarantee the reversible consistency of the transformation and the smoothness, and enhance the computational efficiency and registration precision.

Based on the above advantages, Cao et al. [24] developed a large diffeomorphic registration algorithm for vector fields. Due to computational difficulties, this algorithm was not applied widely. However, this method was a successful foundation for differential homeomorphism registration method. In 2008, Vercauteren et al. [25] proposed the symmetric log-domain diffeomorphic registration method. The parameterization of diffeomorphic transformations was done completely in log-domain, based on stationary velocity field and Lie group structure, which guaranteed the invertibility of deformation and had access to the true inverse transformation. Almost simultaneously, Avants et al. [26] developed a symmetric image normalization method (denoted as SyN). The Euler-Lagrange equation was used for the optimization. In 2009, Vercauteren et al. [27] proposed an efficient nonparametric diffeomorphic image registration algorithm. It

optimized the entire space of displacement fields based on Thirion's demons algorithm [28]. In 2010, the literature [29] compared symmetric log-domain diffeomorphic registration and asymmetric log-domain diffeomorphic registration. The results showed that the former has good reversible consistency by catching transformation information.

2.2. Tensor Image Based Registration. The registration of tensor image is more difficult than scalar image based registration. One reason is multidimensionality of the data. Another is that anatomical structure has changed after image transformations. We need to ensure the tensor orientations to keep consistence with the anatomy.

In 2003, Park et al. [30] proposed multichannel DTI registration method based on the Demons algorithm. The whole diffusion tensor and various features of tensor were used in this algorithm and this improved the quality of registration. But the tensor reorientation was not explicitly optimized and only applied tensor reorientation iteratively. In 2006, Zhang et al. [3] proposed diffeomorphic deformable tensor registration (named as DTI-TK). This method measured tensor similarity as a whole and enabled explicit optimization of tensor reorientation without additional correction to tensor orientation. In 2009, Yeo et al. [31] proposed exact finite-strain diffeomorphic registration, which combined exactly finite strain reorientation with the object function of Demons. This algorithm was reversible consistency and achieved significantly better registration with the exact gradient. However, the tensor reorientation was not optimized explicitly. In 2009, Yap et al. [32] proposed tensor image morphing for elastic registration. This algorithm leveraged tensor regional distributions and local boundaries directly and was improved by utilizing automatic detecting structure characteristics and thin-plate-spline (TPS) [33]. Recently, DTI-TK was improved in different ways, such as utilizing various tensor characteristics and orientation features with neighborhood interpolation [34], combining tract and tensor features [35], and also extending Statistical Parametric Mapping to reduce the computation complexity [36].

2.3. Fiber Bundle Based Registration. With direct registration of fiber, we can avoid the estimation error of DTI direction and improve the accuracy as well as robustness of the registration. In 2007, Mayer and Greenspan [37] proposed direct registration based on white matter (WM) fiber where fibers were represented as 3D points to be registered. This algorithm adopted an iterative closest fiber approach, and each iteration estimated the 12-parameter affine transformation. However this method was time-consuming. In 2010, Shadmi et al. [38] presented piecewise affine registration of fiber. The registration of fiber was considered as a problem of probability density estimation. The energy function was optimized by the gradient descent method and evaluated by residual mean square error. The algorithm made full use of fiber orientation, so it improved accuracy and robustness of the registration. In 2010, Zvitia et al. [39] proposed registration of WM fibers by Adaptive-Mean-Shift (AMS) and Gaussian Mixture Modeling (GMM). The fibers were projected into a high dimensional feature space based on 3D coordinates. The fiber modes were produced by the AMS, and the GMM of fibers was obtained by Gaussian distribution. The registration of WM fibers depended on the alignment of two GMMs.

Compared to the deterministic fiber tracking technology, probabilistic tractography technique can more reasonably trace to the structure of the nerve fibers and in a certain extent overcome the internal defect of the single tensor model. Since the probability statistics method is introduced, probabilistic tractography can effectively reduce uncertainty of tracking results by noise and other environmental factors and thus has better performance of antinoise interference. But there are few researches on the DTI image registration based on the probabilistic fiber bundles tractography.

To improve the efficiency of DTI image registration, we proposed a DTI image registration method under probabilistic fiber bundles tractography learning. We first get the distribution of the whole brain white matter fiber bundles based on probabilistic tractography. Then, the tracked fiber bundles can be registered by using symmetric image standardization registration algorithm, and the calculated deformation field acts on the DTI images, finally implementing the accurate DTI images registration. For the experiments, we compared our method with the state of the art methods under visualization and three quantitative evaluation standards and gave a comprehensive analysis.

Our method is innovative in the following two aspects:

(1) Using fiber bundles tracked by probabilistic tractography to calculate the deformation field of DTI image registration: Registration based on white matter fiber bundles can avoid the estimation error of DTI direction. Furthermore, probabilistic tractography technique can more reasonably trace to the structure of the nerve fibers and can effectively reduce uncertainty of tracking results by noise and other environmental factors.

(2) Improving the residual error estimation step in bootstrap method used in active sample selection learning for the probabilistic tractography: Our method assumes that, in the case of independent and identically distributed error, the residuals can be adjusted and the error model is then modified by using finite sample set, therefore, improving the study ability of samples.

3. Probabilistic Fiber Bundles Tractography Learning

3.1. Probabilistic Tractography. Given a brain diffusion MRI image, the DTI can be modeled as a simple diffusion with a Gaussian profile [40]:

$$G(x; D, t) = \left((4\pi t)^3 \det(D)\right)^{-1/2} \exp\left(\frac{-x^T D^{-1} x}{4t}\right), \quad (1)$$

where D is the diffusion tensor, t is the diffusion time, and x describes the element of the MRI image.

To sample the ellipsoid structure based on probability distribution, the 3×3 diffusion tensor D needs to be solved. The estimation of the diffusion coefficients of the tensor D can be implemented by six independent measurements along noncollinear gradient orientations. The solution to D, using singular value decomposition (SVD), identifies a new basis system describing the diffusion profile at each voxel using eigen values λ_1, λ_2, and λ_3 and the corresponding eigen vector $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ that indicate the preferred direction of water diffusion.

Probabilistic tracking algorithm devised by Friman et al. [8] is based on a Bayesian inference and estimation scheme. Due to noise or complex fiber architectures, uncertainties of probability are not disregarded but captured in the model itself, in form of the posterior distribution at each voxel. Given a source region A and a target region B, the probability of connectivity between A and B is given as

$$p(A \longrightarrow B \mid D) = \sum_{n=1}^{\infty} \int_{\Omega_{AB}^{n}} p(n)\, p(v_{1:n} \mid D), \qquad (2)$$

where $p(v_{1:n} \mid D)$ is the probability of the fiber path going from A to B, given the diffusion tensor D, and v represents a voxel. Ω_{AB}^{n} represents the sampling space of the connectivity between A and B of path length 1 through n.

In order to make (2) analytically solvable, a rejection sampling strategy can be employed. Specifically, a large number of sampled fiber paths starting from region A are drawn randomly, and the probability of the path between A and B is then evaluated. These random paths need to be found by working at each step of the path up until the predetermined length n. We assume these steps are unit length vectors and only depended on previous step direction. Under this assumption, the posterior distribution at each step is calculated based on the diffusion data D. This distribution can be described in terms of Bayes theorem as

$$p(\hat{v}_i, \theta \mid \hat{v}_{i-1}, D) = \frac{p(D \mid \hat{v}_i, \theta)\, p(\hat{v}_i \mid \hat{v}_{i-1})\, p(\theta)}{p(D)}, \qquad (3)$$

where $p(D \mid \hat{v}_i, \theta)$ represents the likelihood distribution using a constrained model based on a Gaussian diffusion profile at the current point. $p(\hat{v}_i \mid \hat{v}_{i-1})$ is the prior to indicate that the current point depends on the direction of the previous step. The nuisance priors, $p(\theta)$, are the parameters of the Gaussian profile modeled as dirac priors which can significantly save computation time. $p(D)$ is the normalizing constant. Those expressions combined together give the probability distribution at $\hat{v}_i$ over a unit sphere.

3.2. Active Sample Selection Learning. In the probabilistic tractography, sample selection is a very important step. Through the sample selection step, a large number of samples describing the fiber paths starting from region A can be obtained, and the probability density function of the path between A and B is then estimated according to a nonparametric procedure. After that, the probabilistic tractography can be completed based on a Bayesian inference and estimation scheme.

Sample selection learning is one kind of learning methods which learn from the environment to obtain a number of concept related examples and derive general concept after the induction. Bootstrap [9] is a typical active sample selection learning method which includes a nonparametric procedure to estimate the probability density function (PDF), by randomly selecting individual measurements, with replacement, from a set of repeated measurements, thus generating many bootstrap samples [10].

Having observed a random sample $V = (v_1, \ldots, v_n)$ with size n, from a distribution with cumulative density function F,

$$F \longrightarrow (v_1, \ldots, v_n), \qquad (4)$$

the empirical distribution function $\hat{F}$ is then defined to be the discrete distribution that puts probability $1/n$ on each v_i, $i = 1, \ldots, n$. The arrow notation ($\rightarrow$) indicates that the sample values are outcomes of random variables with independent and identically distributed attribute, each with distribution function F, that is, v_i i.i.d F [41]. A bootstrap sample is a random sample of size n drawn from $\hat{F}$, denoted as $V^* = (v_1^*, v_2^*, \ldots, v_n^*)$, obtained by

$$\hat{F} \longrightarrow (v_1^*, v_2^*, \ldots, v_n^*). \qquad (5)$$

The star notation in the upper right corner of V^* indicates that V^* is not the actual data set V, but a randomized version of V. These significant amounts of bootstrap samples enable us to estimate the sampling distribution statistics for making inferences about a population parameter φ. If estimate is denoted as $\hat{\varphi} = s(V)$, for each bootstrap sample, a bootstrap replication of $\hat{\varphi}$ can be computed by [41]

$$\hat{\varphi}^* = s(V^*). \qquad (6)$$

A collection of bootstrap replication will provide us with the information needed to obtain the sampling distribution estimation of $\hat{\varphi}$.

The wild bootstrap (WB) proposed originally by Wu [42] is suited when the residuals of regression model exhibit heteroskedasticity. The observations in this case, $V = [(x_1, y_1), \ldots, (x_n, y_n)]$, are assumed to be instances of bivariate random variable $(\mathbf{X}, \mathbf{Y})$. $\mathbf{X}$ is a R^{d_X}-valued predictor random variable and $\mathbf{Y}$ is a R^{d_Y}-valued response random variable. If $\hat{\ell}(x)$ is an estimate of the regression function $\ell(x) = E(\mathbf{Y} \mid \mathbf{X} = x)$ of $\mathbf{Y}$ on $\mathbf{X}$, WB resamples the residuals by assuming the "true" residual distribution is symmetric. For the least square regression, the finite sample is used to replace the residuals $r_i = y_i - \hat{\ell}(x_i)$ by the factor $(1 - h_i)^{-1/2}$, $i = 1, \ldots, n$, where h_i is the ith diagonal element from the hat matrix of the ordinary least squares solution.

Our method assumes that, in the case of independent and identically distributed error, the residuals can be adjusted and the error model is then modified by using finite sample set. Consequently, our method generates each bootstrap sample using

$$V^* = \left[\left(x_1, \hat{\ell}(x_1) + \tilde{r}_1^*\right), \ldots, \left(x_n, \hat{\ell}(x_n) + \tilde{r}_n^*\right)\right], \qquad (7)$$

where $\tilde{r}_i^* = \tilde{r}_i + \gamma^{-1} h_i \psi(\tilde{r}_i)$, $\gamma = \int \psi(x) f(x) dx$, f is the density function of r_i, ψ is the score function, and $h_i = x_i^T (\sum_k x_k x_k^T)^{-1} x_i$, $k \ll n$. The tildes denote the corrected residuals.

4. Materials and Registration Method

4.1. Ethical Standards

Ethical Approval. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent. Informed consent was obtained from all individual participants included in the study.

4.2. Materials

Diffusion MRI Data. The open accessed IXI dataset from Hammersmith Hospital of London was used (http://www.brain-development.org/). A 3 Tesla Philips MRI scanner was used to scan the healthy subjects. With spatial resolution 1.7409 × 1.7355 × 1.9806 mm, the volume data of head is 128 × 128 × 64 voxels. Diffusion weighted images are along unique gradient directions with $b = 1000\,\mathrm{s/mm}^2$ (repetition time = 11894.44 ms; echo time = 51 ms). More parameter information can be found at the website.

Template and Subject. In this paper, 10 data were chosen randomly as subjects (5 male, average age = 51.586 years, min age = 30.89 years, and max age = 63.68 years; 5 female, average age = 51.512 years, min age = 33.76 years, and max = 74.01 years) and also another data was chosen as the template (male, age = 37.83 years). The template is shown in Figure 1. The white matter area in FA is obviously highlighted in Figures 1(d), 1(e), and 1(f). TR means the trace of diffusion tensor.

4.3. Preprocessing. Brain Extraction Tool (BET) in FMRIB software Library was used to extract brain tissue for each subject and template. The mask used for skull stripping was generated from each subject or template individually and checked manually. Before tensor estimation, diffusion weighted images (DWIs) in 15 diffusion gradient directions were eddy-current corrected with FMRIB software Library.

4.4. Registration Method. In this paper, the symmetric image standardization algorithm (also called symmetric image normalization, SyN) proposed by Avants et al. [26] is used to register the tracked fiber bundles. In this method, the cross-correlation is made as similarity criterion, and the Euler-Lagrange equation is used for algorithm optimization. In this way, the diffeomorphism transformation can be decomposed into two parts and also ensures the reversible consistency of the spatial transform.

In the spatial domain of fiber bundles Ω, if the diffeomorphism transformation function is φ, affine transformation of fiber bundle can be noted as

$$\varphi(\partial\Omega) = \mathbf{A}(\mathbf{Id}), \tag{8}$$

where $\mathbf{A}(\mathbf{Id})$ is the affine transformation, and the symmetrical and time varying velocity field is

$$\frac{d\varphi(\mathbf{x}, t)}{dt} = v(\varphi(\mathbf{x}, t), t). \tag{9}$$

Through the integration of time and smooth velocity field, the diffeomorphism transformation φ can be obtained. φ can be decomposed into two parts: φ_1 and φ_2, transformed to the middle point along the geodesic line, for the fiber bundles to be registered and standard fiber bundle templates, respectively. The parameters are

$$\begin{aligned} v(\mathbf{x}, t) &= v_1(\mathbf{x}, t) \quad t \in [0, 0.5], \\ u(\mathbf{x}, t) &= u_2(\mathbf{x}, 1-t) \quad t \in [0.5, 1], \end{aligned} \tag{10}$$

the corresponding fiber bundle set can be obtained by integral transform, and the similarity measurement is

$$\left|\varphi_1(\mathbf{x}, t)\,\mathbf{I} - \varphi_2(\mathbf{x}, 1-t)\,\mathbf{J}\right|^2. \tag{11}$$

The Euler-Lagrange equation is then used for algorithm optimization. For computation from the fiber bundle to be registered to the standard template fiber bundle or from the standard template fiber bundle to the fiber bundle to be registered, the path is the same ($\mathbf{I} \Leftrightarrow \mathbf{J}$), which ensures the reversible consistency. The formula of reversible consistency can be described as

$$\begin{aligned} \varphi_1^{-1}(\varphi_1) &= \mathbf{Id}, \\ \varphi_2^{-1}(\varphi_2) &= \mathbf{Id}. \end{aligned} \tag{12}$$

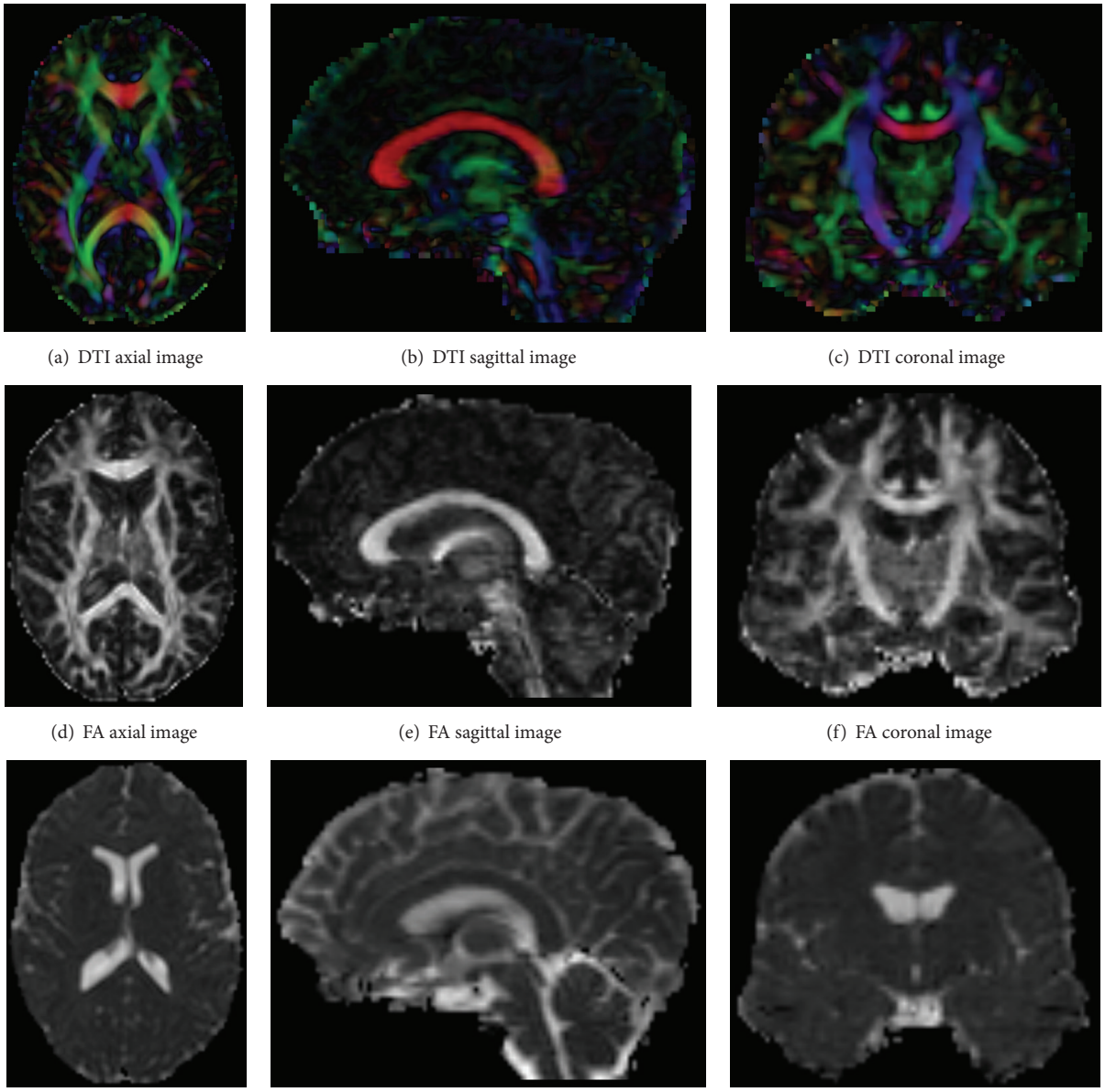
SyN algorithm can deal with both small and large deformation. The results will not change by the input data order, and the diffeomorphic mapping ensures the precision of the reversible consistency transform.

4.5. Evaluation Criteria

4.5.1. Dyadic Coherence κ [43]. After the eigen decomposition of the diffusion tensor, the eigenvalues could be denoted in descending order as $\lambda_1 > \lambda_2 > \lambda_3$, and the corresponding eigenvectors are denoted as $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$. Dyadic coherence describes the variability in the dominant diffusion direction. For each voxel, the dyadic coherence is defined as

$$\kappa = 1 - \sqrt{\frac{\beta_2 + \beta_3}{2\beta_1}}, \tag{13}$$

where κ measures the variability of eigenvectors, which ranges from 0 to 1 (0 for randomly and 1 for identically oriented directions). β_j ($j = 1, 2, 3$) is the eigenvalue of the mean dyadic tensor [38]. The higher value for the dyadic tensor will represent better eigenvector alignment and higher fiber alignment accuracy.



(a) DTI axial image (b) DTI sagittal image (c) DTI coronal image

(d) FA axial image (e) FA sagittal image (f) FA coronal image

(g) TR axial image (h) TR sagittal image (i) TR coronal image

Figure 1: 2D views of the template (the color images are encoded as follows: red for left-right, green for anterior-posterior, and blue for inferior-superior).

4.5.2. Overlap of Eigenvalue-Eigenvector (OVL) [44]. The overlap of eigenvalue-eigenvector pairs is defined as

$$\mathrm{OVL} = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{3}\left(\lambda_j^i\lambda_j'^i\left(\mathbf{e}_j^i\cdot\mathbf{e}_j'^i\right)^2\right)}{\sum_{j=1}^{3}\lambda_j^i\lambda_j'^i}, \tag{14}$$

where λ_j^i, $\mathbf{e}_j^i$, $\lambda_j'^i$, and $\mathbf{e}_j'^i$ are the jth eigenvalue-eigenvector pairs from the ith subject and the template tensors, respectively. The value of OVL is more closer to 1; the alignment of tensor orientation and fibers is better.

4.5.3. Cross-Correlations (CC_x) [45]. The cross-correlations of the WM voxels between subjects and template are computed by using the FA and TR:

$$\mathrm{CC}_x = \frac{\sum_v X_1(v)\,X_2(v)}{\sqrt{\sum_v X_1(v)\,X_1(v)\sum_v X_2(v)\,X_2(v)}}, \tag{15}$$

where v indexes over all the voxels. $X_1(v)$ and $X_2(v)$ are scalar images derived from DTI and could be replaced by FA or TR. The value ranges between 0 and 1. The higher cross-correlation describes the higher similarity between two maps.

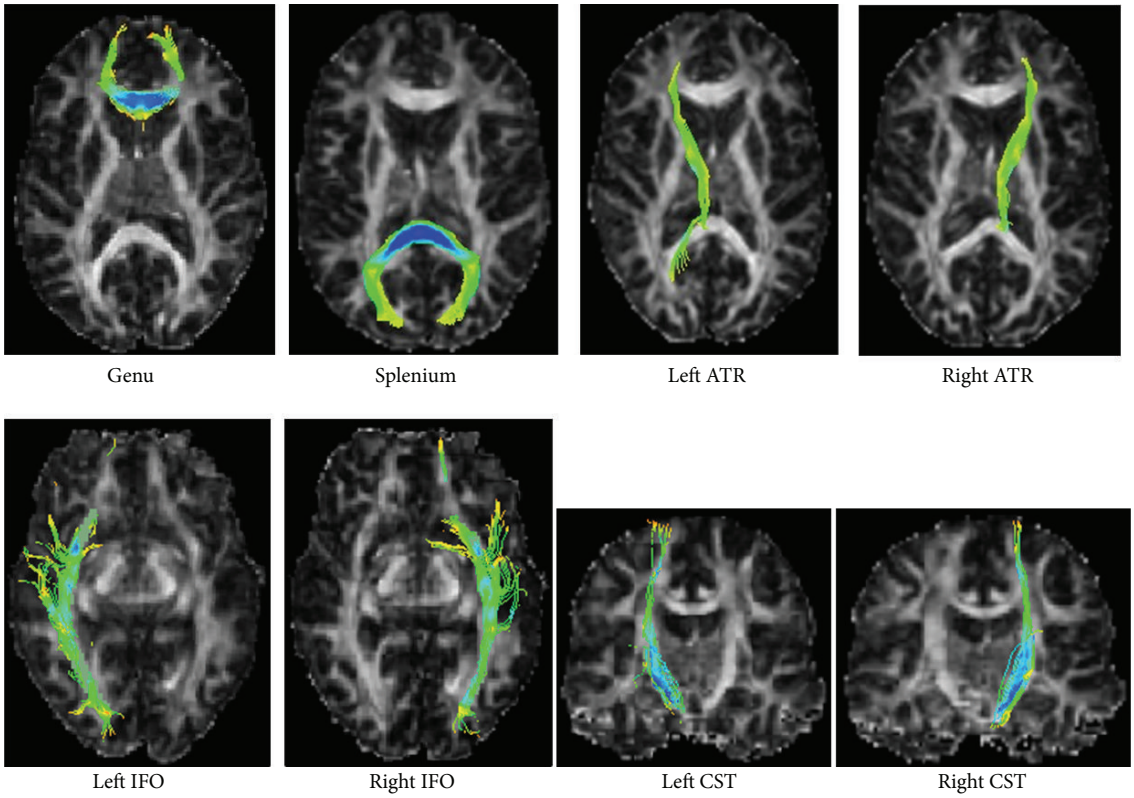


Figure 2: Fiber bundles tracking results (8 ROIs is Genu of the Corpus Callosum (Genu), Splenium of the Corpus Callosum (Splenium), Left Anterior Thalamic Radiations (ATR), Right ATR, Left Inferior Frontooccipital Fasciculi (IFO), Right IFO, Left Corticospinal/Corticobulbar Tracts (CST), and Right CST).

5. Experimental Results

In order to test the performance of the proposed registration method based on probabilistic fiber bundles tractography learning, in this paper, we compare our method with 6 state-of-the-art methods, which are 5 scalar based methods: Rigid [13], Affine [18], Elastic [14], SyN [26], FSL [17], and one tensor based method: DTI-TK [3]. Dyadic coherence κ, overlap of eigenvalue-eigenvector (OVL), and cross-correlations (CC_x) are used as three evaluation criteria. Maps and empirical cumulative distribution functions (CDFs) are used for illustrating 7 registration algorithms. CDF is probability of variable less than or equal to a certain number; that is, $F(x) = p(X \leq x)$, where P is probability.

5.1. Fiber Bundles Tracking Results. The result of fiber bundles tracking will affect the accuracy of the following registration method as tracking result is the input of registration step; as a result, fiber bundles tracking is an important step in the whole algorithm system. Figure 2 gives fiber bundles tracking results of 8 regions of interest (ROIs) by the probabilistic fiber tracking algorithm proposed in this paper. Figure 3 shows the global display of fiber bundles tracking results for 11 experimental data. From Figures 2 and 3, it can be seen that our proposed probabilistic fiber bundles tracking method has the ability to tracking white matter fiber bundles of diffusion MRI image accurately.

5.2. Comparison of Registration Results. In this section, we test the registration effectiveness by visualization. The results of 7 registration algorithms are shown in Figure 4. In Figure 4, images of all the registered datasets were visualized, which could give qualitative results. From the visual results, our proposed method keeps the distribution character of subject white matter fiber bundles and also gets the better matching results with the template.

5.3. Comparison by Dyadic Coherence κ. The higher dyadic coherence value indicates better eigenvectors alignment and anatomical structure consistency. The empirical CDFs of dyadic coherence are presented in Figure 5(a). From the empirical CDFs of dyadic coherence, κ of DTI-TK is the biggest, which indicates the highest anatomical structure consistency. Our proposed method gets the second ranking, only slightly worse than the DTI-TK method and much better than the existing Rigid, Affine, Elastic, SyN, and FSL methods, while, with the increase of κ value, the empirical CDFs of FSL increase rapidly and even exceed the DTI-TK and ours, but the overall empirical CDFs shock more seriously, which means the algorithm performance is not stable.

5.4. Comparison by Overlap of Eigenvalue-Eigenvector (OVL). A higher OVL values represents a greater correspondence in anatomical structure between subjects. The empirical CDFs

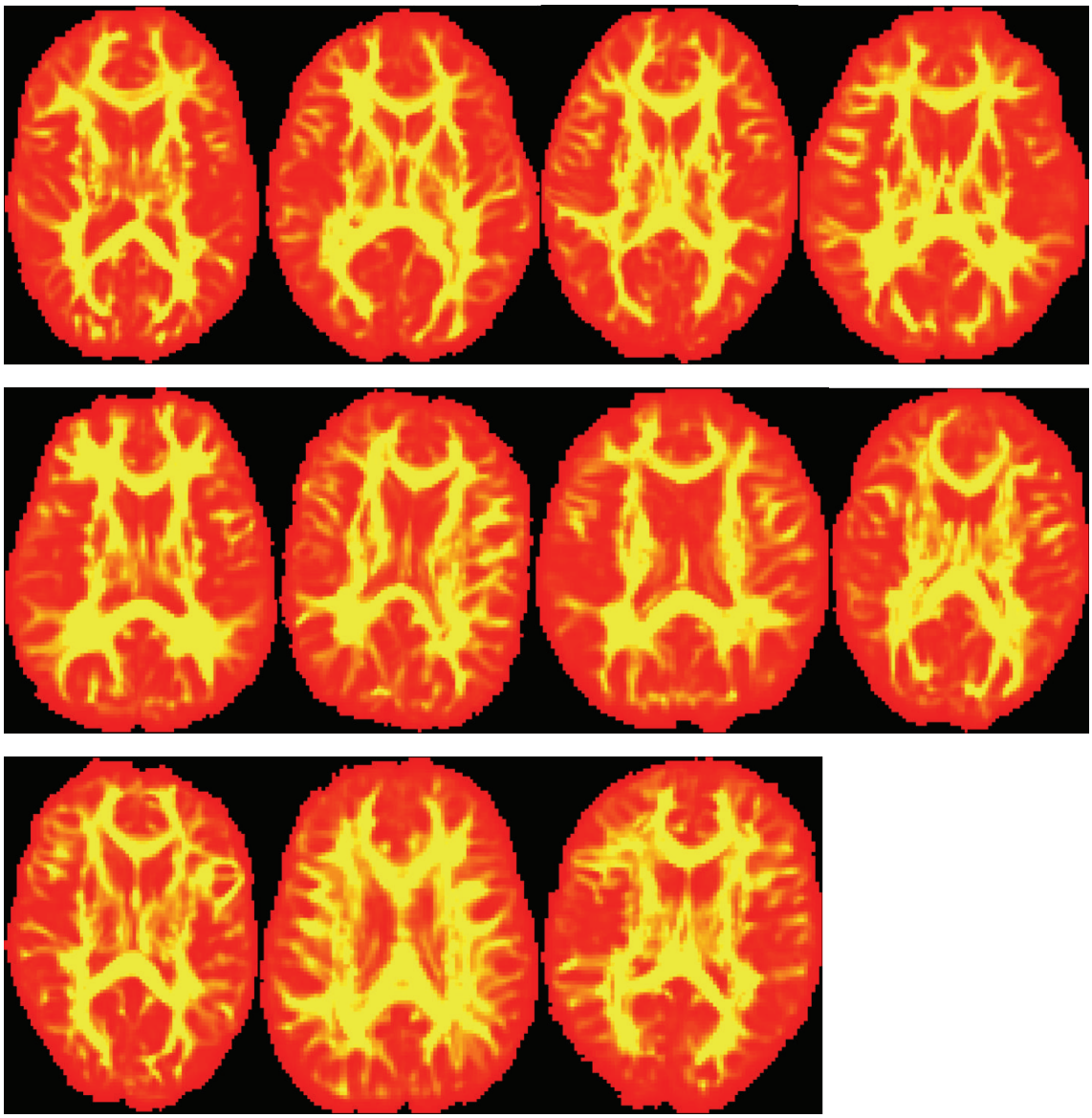
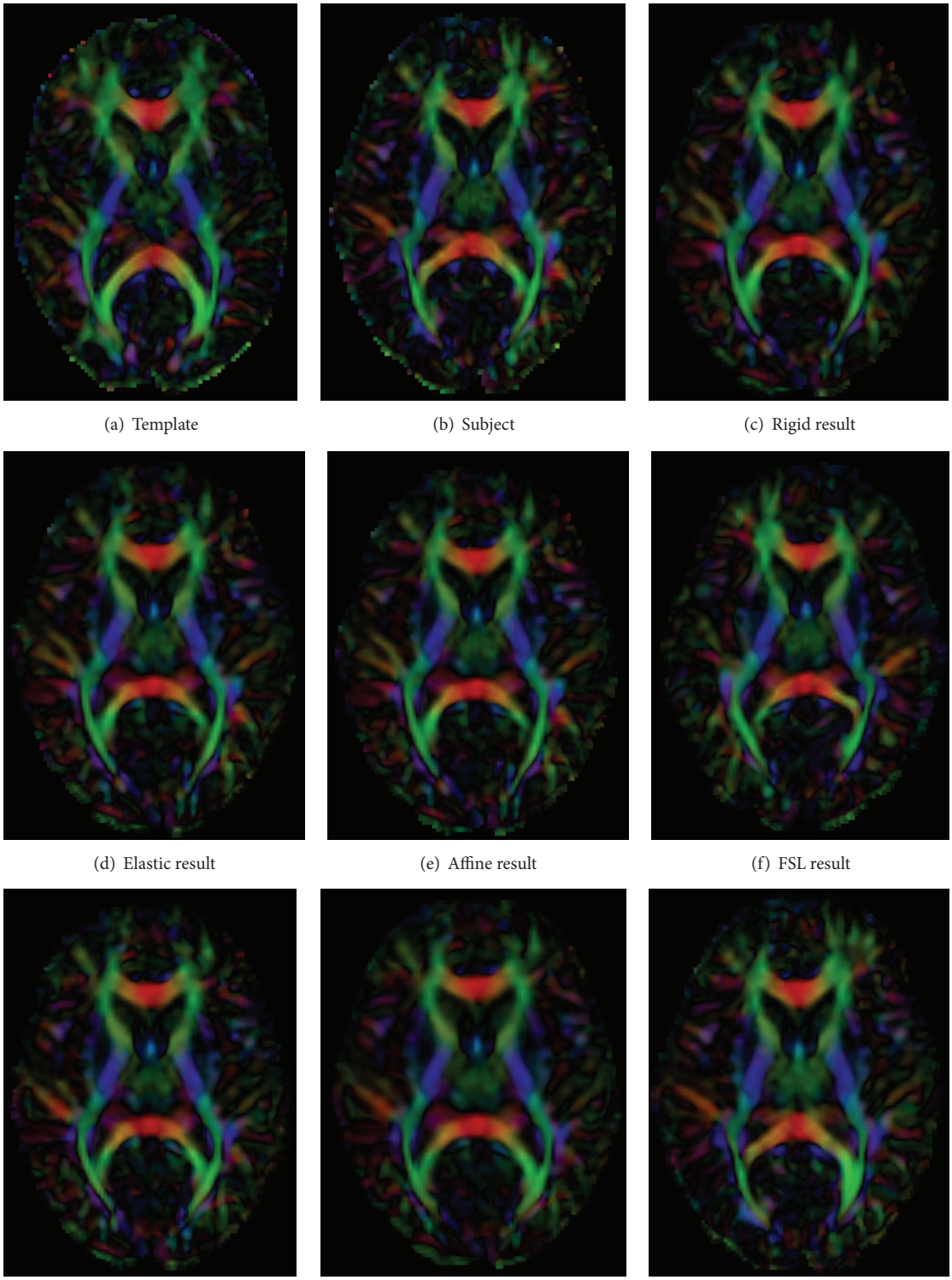


Figure 3: Global display of fiber bundles tracking results.

of OVL are presented in Figure 5(b). From Figure 5(b), it can be seen that when the OVL value is low, Rigid, FSL, and our proposed method show the better performance. As OVL increases, DTI-TK, Elastic, FSL, and our proposed method got the better performance. Considering the changing curve of OVL synthetically, DTI-TK, Elastic, and our proposed method are three stable methods. Our method is better than DTI-TK and quite equal to Elastic.

5.5. Comparison by Cross-Correlation of Diffusion (CC_x). For the cross-correlation, higher value represents the higher similarity between two maps. The empirical CDFs of cross-correlation for FA and TR are presented in Figures 6(a) and 6(b), respectively. From the empirical CDFs of CC_{FA}, DTI-TK and our method show the top two highest performance, our method only slightly worse than the DTI-TK method. From the empirical CDFs of CC_{TR}, our method is the best, which indicates highest image similarity.

From three evaluation criteria and visualization experimental results, our proposed method shows a high comprehensive performance. DTI-TK method is the currently recognized best registration method; our method shows the quite equal comprehensive performance. DTI-TK is a non-parametric, diffeomorphic deformable image registration, taking tensors as a whole and explicating the optimization of tensor reorientation. The disadvantage of the DTI-TK is that the image boundary is not smooth and the computing is complicated. Meanwhile, it only supports the affine transformation with the least parameters. Our method is based on the completely different algorithm theory; we completes the DTI Image registration method under probabilistic fiber bundles tractography learning. The distribution of the whole



(a) Template (b) Subject (c) Rigid result

(d) Elastic result (e) Affine result (f) FSL result

(g) SyN result (h) DTI-TK result (i) Proposed result

Figure 4: The results of 7 registration algorithms.

brain white matter fiber bundles is first obtained based on probabilistic tractography. Then, the tracked fiber bundles are registered by using symmetric image standardization registration algorithm, and the calculated deformation field acts on the DTI images. Those steps all have the advantages to improve the registration accuracy and robustness.

For the empirical CDFs of CC_{TR} and OVL, our method is better than DTI-TK. The experimental results show that

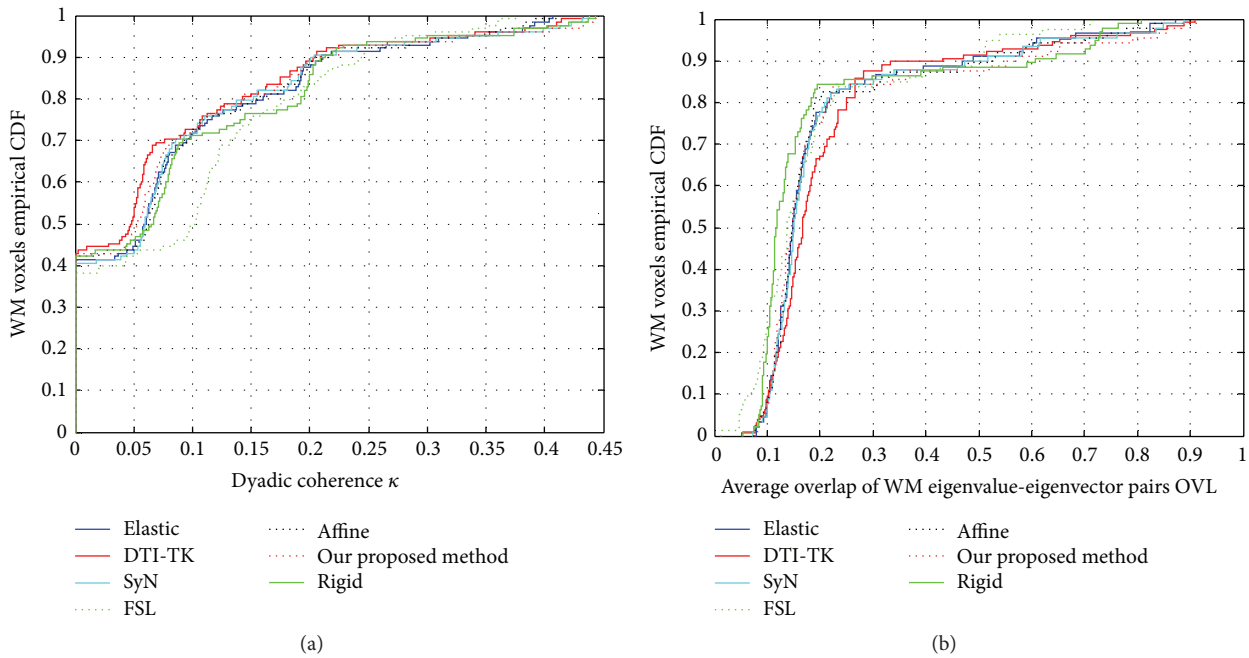


Figure 5: The empirical CDFs of dyadic coherence and OVL. (a) is the empirical CDFs of dyadic coherence; (b) is the empirical CDFs of OVL.

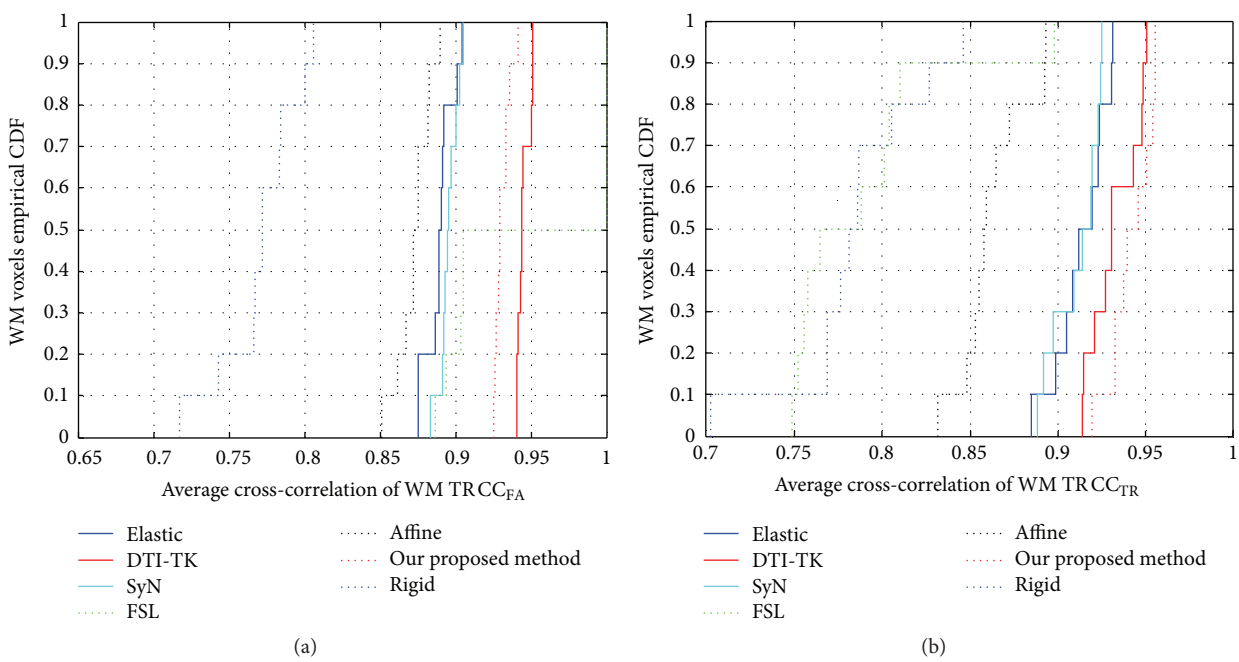


Figure 6: The empirical CDFs of cross-correlation. (a) is the empirical CDFs of the cross-correlations for FA; (b) is the empirical CDFs of the cross-correlations for TR.

the proposed method has a very good comprehensive performance and can be used for DT-MRI Image registration.

6. Conclusions

In this paper, we proposed a DTI Image registration method under probabilistic fiber bundles tractography learning, as the probabilistic tractography technique can more reasonably trace to the structure of the nerve fibers and in a certain extent overcome the internal defect of the single tensor model. We improved the residual error estimation step in bootstrap method used in active sample selection learning for the probabilistic tractography. The results of our proposed method were compared with 5 scalar based methods (Rigid,

Affine, Elastic, SyN, and FSL) and one tensor based method (DTI-TK). The visualization and 3 quantitative evaluation standards were used to give a comprehensive analysis. The experimental results show that our proposed probabilistic fiber bundles tracking method has the ability to track white matter fiber bundles of diffusion MRI image accurately. Our registration method gives a quite equal comprehensive performance with DTI-TK, much better than the others. Consequently, our method can be used for accurate and efficient DTI image registration.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61303123, 61402371, and 61461025, Natural Science Basic Research Plan in Shaanxi Province of China under Grants 2015JM6317 and 2015JQ6256, and the Fundamental Research Funds for the Central Universities under Grant 3102015JSJ0008.

References

[1] P. C. Sundgren, Q. Dong, D. Gómez-Hassan, S. K. Mukherji, P. Maly, and R. Welsh, "Diffusion tensor imaging of the brain: review of clinical applications," *Neuroradiology*, vol. 46, no. 5, pp. 339–350, 2004.

[2] Y. Xia, Z. Ji, and Y. Zhang, "Brain MRI image segmentation based on learning local variational Gaussian mixture models," *Neurocomputing*, vol. 204, pp. 189–197, 2016.

[3] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee, "Deformable registration of diffusion tensor MR images with explicit orientation optimization," *Medical Image Analysis*, vol. 10, no. 5, pp. 764–785, 2006.

[4] G. Saygili, M. Staring, and E. A. Hendriks, "Confidence estimation for medical image registration based on stereo confidences," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 539–549, 2016.

[5] Y. Wang, W. Zeng, L. Yu et al., "Advance in diffusion tensor image registration and its evaluation," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 2, pp. 562–570, 2016.

[6] R. S. Vorburger, C. Reischauer, and P. Boesiger, "BootGraph: probabilistic fiber tractography using bootstrap algorithms and graph theory," *NeuroImage*, vol. 66, pp. 426–435, 2013.

[7] S. Jbabdi, T. E. J. Behrens, and S. M. Smith, "Crossing fibres in tract-based spatial statistics," *NeuroImage*, vol. 49, no. 1, pp. 249–256, 2010.

[8] O. Friman, G. Farnebäck, and C.-F. Westin, "A Bayesian approach for stochastic white matter tractography," *IEEE Transactions on Medical Imaging*, vol. 25, no. 8, pp. 965–978, 2006.

[9] B. Efron, "Bootstrap methods: another look at the jackknife," in *Breakthroughs in Statistics*, pp. 569–593, Springer, New York, NY, USA, 1992.

[10] D. K. Jones, "Tractography gone wild: probabilistic fibre tracking using the wild bootstrap with diffusion tensor MRI," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1268–1274, 2008.

[11] B. Whitcher, D. S. Tuch, J. J. Wisco, A. G. Sorensen, and L. Wang, "Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging," *Human Brain Mapping*, vol. 29, no. 3, pp. 346–362, 2008.

[12] Y. Wang, Q. Yu, Z. Liu et al., "Evaluation on diffusion tensor image registration algorithms," *Multimedia Tools and Applications*, vol. 75, no. 13, pp. 8105–8122, 2016.

[13] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[14] D. C. Alexander and J. C. Gee, "Elastic matching of diffusion tensor images," *Computer Vision and Image Understanding*, vol. 77, no. 2, pp. 233–250, 2000.

[15] J. A. Schnabel, D. Rueckert, M. Quist et al., "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 573–581, Springer, Utrecht, Netherlands, October 2001.

[16] X.-F. Yao and Z.-J. Song, "Deformable registration for geometric distortion correction of diffusion tensor imaging," in *Computer Analysis of Images and Patterns. Part I*, vol. 6854, pp. 545–553, Springer, Heidelberg, Germany, 2011.

[17] J. L. R. Andersson, M. Jenkinson, and S. Smith, "Non-linear registration, aka spatial normalisation FMRIB," Tech. Rep. TR07JA2, FMRIB Analysis Group of the University of Oxford, 2007.

[18] A. Leemans, J. Sijbers, S. De Backer, E. Vandervliet, and P. M. Parizel, "Affine coregistration of diffusion tensor magnetic resonance images using mutual information," in *Advanced Concepts for Intelligent Vision Systems*, J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, Eds., vol. 3708 of *Lecture Notes in Computer Science*, pp. 523–530, Springer, Berlin, Germany, 2005.

[19] M.-C. Chiang, A. D. Leow, A. D. Klunder et al., "Fluid registration of diffusion tensor images using information theory," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 442–456, 2008.

[20] H. Hufnagel, X. Pennec, G. Malandain, H. Handels, and N. Ayache, "Non-linear 2D and 3D registration using block-matching and B-splines," in *Bildverarbeitung für die Medizin 2005*, pp. 325–329, Springe, Berlin, Germany, 2005.

[21] V. Arsigny, O. Commowick, X. Pennec et al., "A log-euclidean framework for statistics on diffeomorphisms," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 924–931, Springer, Copenhagen, Denmark, October 2006.

[22] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean metrics for fast and simple calculus on diffusion tensors," *Magnetic Resonance in Medicine*, vol. 56, no. 2, pp. 411–421, 2006.

[23] C. Feddern, J. Weickert, B. Burgeth, and M. Welk, "Curvature-driven PDE methods for matrix-valued images," *International Journal of Computer Vision*, vol. 69, no. 1, pp. 93–107, 2006.

[24] Y. Cao, M. I. Miller, R. L. Winslow, and L. Younes, "Large deformation diffeomorphic metric mapping of vector fields," *IEEE Transactions on Medical Imaging*, vol. 24, no. 9, pp. 1216–1230, 2005.

[25] T. Vercauteren, X. Pennec, A. Perchant et al., "Symmetric log-domain diffeomorphic registration: a demons-based

approach," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008: 11th International Conference, New York, NY, USA, September 6–10, 2008, Proceedings, Part I*, vol. 5241 of *Lecture Notes in Computer Science*, pp. 754–761, Springer, Berlin, Germany, 2008.

[26] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[27] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.

[28] J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.

[29] A. Sweet and X. Pennec, "Log-domain diffeomorphic registration of diffusion tensor images," in *Biomedical Image Registration*, B. Fischer, B. M. Dawant, and C. Lorenz, Eds., vol. 6204 of *Lecture Notes in Computer Science*, pp. 198–209, Springer, Berlin, Germany, 2010.

[30] H.-J. Park, M. Kubicki, M. E. Shenton et al., "Spatial normalization of diffusion tensor MRI using multiple channels," *NeuroImage*, vol. 20, no. 4, pp. 1995–2009, 2003.

[31] B. T. T. Yeo, T. Vercauteren, P. Fillard et al., "DT-REFinD: diffusion tensor registration with exact finite-strain differential," *IEEE Transactions on Medical Imaging*, vol. 28, no. 12, pp. 1914–1928, 2009.

[32] P.-T. Yap, G. Wu, H. Zhu, W. Lin, and D. Shen, "TIMER: tensor image morphing for elastic registration," *NeuroImage*, vol. 47, no. 2, pp. 549–563, 2009.

[33] P.-T. Yap, G. Wu, H. Zhu, W. Lin, and D. Shen, "F-TIMER: fast tensor image morphing for elastic registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 5, pp. 1192–1203, 2010.

[34] M. Ingalhalikar, J. Yang, C. Davatzikos, and R. Verma, "DTI-DROID: diffusion tensor imaging-deformable registration using orientation and intensity descriptors," *International Journal of Imaging Systems and Technology*, vol. 20, no. 2, pp. 99–107, 2010.

[35] Q. Wang, T. P. Yap, G. Wu et al., "Diffusion tensor image registration with combined tract and tensor," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 200–208, Springer, Toronto, Canada, 2011.

[36] L. Gan and G. Agam, "Efficient nonlinear DTI registration using DCT basis functions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '11)*, pp. 17–22, IEEE, Colorado Springs, Colo, USA, June 2011.

[37] A. Mayer and H. Greenspan, "Direct registration of white matter tractographies and application to atlas construction," in *Proceedings of the Workshop on Statistical Registration: Pair-Wise and Group-Wise Alignment and Atlas Formation*, Brisbane, Australia, November 2007.

[38] R. Shadmi, A. Mayer, N. Sochen, and H. Greenspan, "Piecewise smooth affine registration of point-sets with application to DT-MRI brain fiber-data," in *Proceedings of the 7th IEEE International Symposium on Biomedical Imaging: from Nano to Macro (ISBI '10)*, pp. 528–531, Rotterdam, Netherlands, April 2010.

[39] O. Zvitia, A. Mayer, R. Shadmi, S. Miron, and H. K. Greenspan, "Co-registration of white matter tractographies by adaptive-mean-shift and Gaussian mixture modeling," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 132–145, 2010.

[40] D. Pai, O. Muzik, and J. Huca, "Quantitative analysis of diffusion tensor images across subjects using probabilistic tractography," in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP '08)*, pp. 1448–1451, IEEE, San Diego, Calif, USA, October 2008.

[41] P.-T. Yap, H. An, Y. Chen, and D. Shen, "Uncertainty estimation in diffusion MRI using the nonlocal bootstrap," *IEEE Transactions on Medical Imaging*, vol. 33, no. 8, pp. 1627–1640, 2014.

[42] C. F. J. Wu, "Jackknife, bootstrap and other resampling methods in regression analysis," *The Annals of Statistics*, vol. 14, no. 4, pp. 1261–1295, 1986.

[43] N. Adluru, H. Zhang, A. S. Fox et al., "A diffusion tensor brain template for Rhesus Macaques," *NeuroImage*, vol. 59, no. 1, pp. 306–318, 2012.

[44] W. Van Hecke, A. Leemans, E. D'Agostino et al., "Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information," *IEEE Transactions on Medical Imaging*, vol. 26, no. 11, pp. 1598–1612, 2007.

[45] S. Zhang, H. Peng, R. J. Dawe, and K. Arfanakis, "Enhanced ICBM diffusion tensor template of the human brain," *NeuroImage*, vol. 54, no. 2, pp. 974–984, 2011.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 2860643, 7 pages
http://dx.doi.org/10.1155/2016/2860643

Research Article

Two-Layer Tight Frame Sparsifying Model for Compressed Sensing Magnetic Resonance Imaging

Shanshan Wang,[1,2] Jianbo Liu,[1,2] Xi Peng,[1,2] Pei Dong,[3] Qiegen Liu,[4] and Dong Liang[1,2]

[1] *Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Shenzhen, Guangdong 518055, China*
[2] *The Beijing Center for Mathematics and Information Interdisciplinary Sciences, Beijing 100048, China*
[3] *Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, The University of Sydney, Sydney, NSW 2006, Australia*
[4] *Nanchang University, Nanchang, Jiangxi, China*

Correspondence should be addressed to Dong Liang; dong.liang@siat.ac.cn

Received 20 April 2016; Revised 5 August 2016; Accepted 18 August 2016

Academic Editor: Andrey Krylov

Compressed sensing magnetic resonance imaging (CSMRI) employs image sparsity to reconstruct MR images from incoherently undersampled K-space data. Existing CSMRI approaches have exploited analysis transform, synthesis dictionary, and their variants to trigger image sparsity. Nevertheless, the accuracy, efficiency, or acceleration rate of existing CSMRI methods can still be improved due to either lack of adaptability, high complexity of the training, or insufficient sparsity promotion. To properly balance the three factors, this paper proposes a two-layer tight frame sparsifying (TRIMS) model for CSMRI by sparsifying the image with a product of a fixed tight frame and an adaptively learned tight frame. The two-layer sparsifying and adaptive learning nature of TRIMS has enabled accurate MR reconstruction from highly undersampled data with efficiency. To solve the reconstruction problem, a three-level Bregman numerical algorithm is developed. The proposed approach has been compared to three state-of-the-art methods over scanned physical phantom and in vivo MR datasets and encouraging performances have been achieved.

1. Introduction

Compressed sensing magnetic resonance imaging (CSMRI) is a very popular signal processing based technique for accelerating MRI scan. Different from the classical fixed-rate sampling dogma Shannon-Nyquist sampling theorem, CS exploits the sparsity of an MR image and allows CSMRI to recover MR images from less incoherently sampled K-space data [1]. The classical formulation of CSMRI can be written as

$$\begin{aligned} \min_{u} \quad & \|Wu\|_1 \\ \text{s.t.} \quad & F_p u = f, \end{aligned} \tag{1}$$

where $u \in \mathscr{C}^{Q\times 1}$ and $f \in \mathscr{C}^{P\times 1}$, respectively, denote the MR image and its corresponding undersampled raw K-space data, $F_p \in \mathscr{C}^{P\times Q}$ represents the undersampled Fourier encoding matrix with $P \ll Q$, and $\|Wu\|_1$ is an analysis model which sparsifies the image with transform $W \in \mathscr{C}^{Q\times Q}$ under the ℓ_1 norm constraint. P and Q are the number of image pixels and measured data. The classical formulation is typically equipped with total variation and wavelet and it can be solved very efficiently [1]. However, the efficiency comes at the expense of accuracy, especially with highly undersampled noisy measurements, due to lack of adaptability or insufficient sparsity promotion. To address this issue, there have been diverse methods proposed [2, 3] and we focus on the following three representative directions.

One main endeavor is employing nonlocal operations or redundant transforms to analytically sparsify the MR image [4]. Typical examples include nonlocal total variation regularization [5], patch-based directional wavelet [6], and wavelet tree sparsity based CSMRI techniques [7]. These methods generally have straightforward models; nevertheless, the reconstruction accuracy is not that perfectly satisfying due

to lack of adaptability. We proposed one-layer data-driven tight frame DDTF for undersampled image reconstruction [8]. It is generally very efficient. But its performance is still limited due to its insufficient sparsity promotion and reliance on the Bregman iteration technique for bringing back the image details.

The other effort is training adaptive dictionary to sparsely represent the MR image in the synthesis manner. For example, DLMRI [9], BPFA triggered MR reconstruction [10], and our proposed TBMDU [3] employ dictionary learning to adaptively capture image structures while promoting sparsity. These methods can generally achieve accurate MR image reconstruction with strong noise suppression capability. Unfortunately, the complexity of these approaches is very high and the sparsity is still directly limited to one-layer representation of the target image.

The third group endeavors could be regarded as the variants of the above two efforts, which target employing the advantages of both the analysis and synthesis sparse models. For example, the balanced tight frame model [11] introduces a penalty term to bridge the gap between the analysis and synthesis model. Unfortunately, although it possesses a fascinating mathematical explanation, the sparsity promotion is still limited to a single layer and therefore its performance is only comparable to the analysis one. To further promote sparsity, a wavelet driven dictionary learning (named WaveDLMRI) [12] technique and our proposed total variation driven dictionary learning approach (named GradDLRec) [13] adaptively represent the sparse coefficients derived from the analysis transform rather than directly encode the underlying image. Nevertheless, despite achieving encouraging performances, they still rely on the computationally expensive dictionary learning technique.

Recently, there are double sparsity model and doubly sparse transforms proposed in general image/signal processing community [14, 15]. The double sparsity model tries to train a sparse dictionary over a fixed base, while the doubly sparse transform is devoted to learning an adaptive sparse matrix over an analytic transform. There is no doubt that their application to image denoising has presented promising results, albeit the two-layer sparsifying model is more concerned to assist efficient learning, storage, and implementation by constraining the dictionary sparse rather than focus on further triggering of the sparsity of the image.

Motivated by the above observations, we try to develop a two-layer tight frame sparsifying (TRIMS) model for CSMRI by sparsifying the image with a product of a fixed tight frame and an adaptive learned tight frame. The proposed TRIMS has several merits: (1) the tight frame satisfies the perfect reconstruction property which ensures the given signal can be perfectly represented by its canonical expansion [16]; (2) a tight frame can be implemented very efficiently since it satisfies $W^H W = I$; (3) the adaptability has been kept by the second-layer tight frame tailored for the target reconstruction task; (4) the two-layer tight frame has enabled the image sparsity to be explored more sufficiently compared to the one-layer one. Furthermore, the two-layer tight frame also has a convolutional explanation, which extracts appropriate image characteristics to constrain MR image reconstruction [17]. We have compared our method with three state-of-the-art approaches of the above three directions, namely, DDTF-MRI, DLMRI, and GradDLRec on an in vivo complex valued MR dataset. The results have advised the proposed method could properly balance the efficiency, accuracy, and acceleration factors.

2. Theory

2.1. TRIMS Model. To reconstruct MR images from undersampled data, we propose a TRIMS model which can be implicitly described as

$$\min_{u, W_b \in \wedge} \left\| F_p u - f \right\|_2^2 + \alpha \left\| W_b \left(W_a u \right) \right\|_1, \tag{2}$$

where W_a is the fixed tight frame and W_b denotes the data-driven tight frame. $\wedge$ means the tight frame system, since a tight frame can be formulated with a set of filters under the unitary extension principle (UEP) condition [16]. The proposed model also has another approximately equivalent convolutional expression, which we name the explicit model

$$\min_{u, b_m} \left\| F_p u - f \right\|_2^2 + \alpha \sum_m \sum_n \left\| b_m * \left(a_n * u \right) \right\|_1, \tag{3}$$

where a_n are the fixed kernels and b_m denote the to-be-learned adaptive kernels.

2.2. TRIMS Algorithm. To solve the proposed model, we develop a three-level Bregman iteration numerical algorithm. Introducing a Bregman parameter c, we have the first-level Bregman iteration

$$\begin{aligned} \left\{ u^{k+1}, W_b^{k+1} \right\} &= \operatorname*{argmin}_{u, W_b \in \wedge} \left\| F_p u - f + c^k \right\|_2^2 \\ &\quad + \alpha \left\| W_b \left(W_a u \right) \right\|_1, \\ c^{k+1} &= c^k + F_p u^{k+1} - f. \end{aligned} \tag{4}$$

To attack the first subproblem in (4), we introduce an assistant variable $u_a = W_a u$ and obtain the second-level iteration

$$\begin{aligned} \left\{ u_a^{k+1}, W_b^{k+1} \right\} &= \operatorname*{argmin}_{u_a, W_b} \mu \left\| W_a u - u_a + d^k \right\|_2^2 \\ &\quad + \alpha \left\| W_b u_a \right\|_1, \\ u^{k+1} &= \operatorname*{argmin}_{u} \left\| F_p u - f + c^k \right\|_2^2 \\ &\quad + \mu \left\| W_a u - u_a^{k+1} + d^k \right\|_2^2, \\ d^{k+1} &= d^k + W_a u^{k+1} - u_a^{k+1}. \end{aligned} \tag{5}$$

The subproblem regarding the update of u is a simple least squares problem admitting an analytical solution. Its solution satisfies the following normal equation:

$$F_p^H \left(F_p u - f + c^k \right) + \mu W_a^H \left(W_a u - u_a^k + d^k \right) = 0. \tag{6}$$

Initialization: $k = 0$, W_a, W_b^0, $c^0 = 0$, $d^0 = 0$, $e^0 = 0$.
while stop condition is not met **do**
  $t \leftarrow 0.01$, $w \leftarrow 0$
  for $m = 1$ to M **do**
    $v^{k+1} = \text{shrink}(W_b u_a^k + e^k, 1/\alpha)$
    $e^{k+1} = e^k + W_b^k u_a^k - v^{k+1}$
    $u_a^{k+1} = (\mu(W_a u^k + d) + W_b^H(v^{k+1} - e^k))/(1 + \mu)$
    update W_b^{k+1} with the SVD technique of [16]
    $d^{k+1} = d^k + W_a u^{k+1} - u_a^{k+1}$
    update u^{k+1} as the inverse Fourier transform of the data acquired in (7)
  end for
  update $c^{k+1} = c^k + F_p u^{k+1} - f$
  $k \leftarrow k + 1$
end while

Algorithm 1: Reconstructing MR images from undersampled K-space data with TRIMS.

Since W_a is a tight frame satisfying $W_a^H W_a = I$, letting F denote the full Fourier encoding matrix normalized such that $F^H F = I$, we have

$$Fu\left(k_x, k_y\right) = \begin{cases} S\left(k_x, k_y\right), & \left(k_x, k_y\right) \notin \Omega, \\ \dfrac{S\left(k_x, k_y\right) + \mu S_0\left(k_x, k_y\right)}{1 + \mu}, & \left(k_x, k_y\right) \in \Omega, \end{cases} \tag{7}$$

where $S_0(k_x, k_y) = FF_p^H(f - c^k)$, $S(k_x, k_y) = FW_a^H(u_a^k - d^k)$, and Ω denotes the sampled K-space subset. In order to update u_a and W_b, we introduce another assistant variable $v = W_b u_a$ to decompose the coupling between W_b and u_a and therefore obtain the third-level Bregman iteration

$$\begin{aligned} \left\{v^{k+1}, W_b^{k+1}\right\} &= \underset{v, W_b}{\operatorname{argmin}} \left\|W_b u_a^k - v + e^k\right\|_2^2 + \alpha \|v\|_1, \\ u_a^{k+1} &= \underset{u_a}{\operatorname{argmin}}\, \mu \left\|W_a u^k - u_a + d^k\right\|_2^2 \\ &\quad + \left\|W_b^k u_a - v^{k+1} + e^k\right\|_2^2, \\ e^{k+1} &= e^k + W_b^k u_a^k - v^{k+1}. \end{aligned} \tag{8}$$

Similar to the update of u, we can easily get the least squares solution for u_a

$$u_a^{k+1} = \frac{\mu\left(W_a u^k + d^k\right) + W_b^H\left(v^{k+1} - e^k\right)}{1 + \mu}. \tag{9}$$

As for the update of v, we temporarily fix the value of W_b and can easily obtain its update rule with the iterative shrinkage/thresholding algorithm (ISTA)

$$v^{k+1} = \text{shrink}\left(W_b u_a^{k+1} + e^k, \frac{1}{\alpha}\right), \tag{10}$$

where $\text{shrink}(x, a) = \text{sign}(x)\max(0, |x| - a)$. Now fix v, we update W_b by minimizing

$$\underset{W_b \in \wedge}{\operatorname{argmin}} \left\|W_b u_a^k - v + e^k\right\|_2^2. \tag{11}$$

Instead of directly optimizing W_b, we sequentially partition the coefficient vectors $v - e$ into vectors and apply the technique of [16] to solve this subproblem using singular value decomposition (SVD), with the aim of learning its corresponding filter b_m. To facilitate the readers to grasp the overall picture, we summarize the proposed TRIMS in Algorithm 1.

3. Experiments and Results

We evaluated the proposed method on three datasets, namely, a T1-weighted brain image obtained from GE 3T commercial scanner with an eight-channel head coil (TE = 11 ms, TR = 700 ms, FOV = 22 cm, and matrix = 256 × 256), a PD-weighted brain image scanned from 3T SIEMENS with an eight-channel head coil and MPRAGE (3D flash with IR prep, TE = 3.45 ms, TR = 2530 ms, TI = 1100 ms, flip angle = 7 deg., slice = 1, matrix = 256 × 256, slice thickness = 1.33 mm, FOV = 256 mm, and measurement = 1), and a physical phantom scanned from a 3T commercial scanner (SIEMENS MAGNETOM TrioTim syngo) with a four-channel head coil (TE = 12 ms, TR = 800 ms, FOV = 24.2 cm, and matrix = 256 × 256). Informed consent was obtained from the imaging subject in compliance with the Institutional Review Board policy. The Walsh adaptive combination method is applied to combine the multichannel data to a single-channel one corresponding to a complex-valued image. We have compared the proposed method to three state-of-the-art methods, namely, the representative analysis transform based DDTF-MRI, the synthesis dictionary based DLMRI, and the analysis-synthesis mixture based GradDLRec approach. TRIMS was implemented with shift invariant Haar wavelet filters for the fixed tight frame (the size of each filter is 2 × 2) and for initializing the second-level tight frame (the size of each filter is 4 × 4). The other three algorithms were implemented with

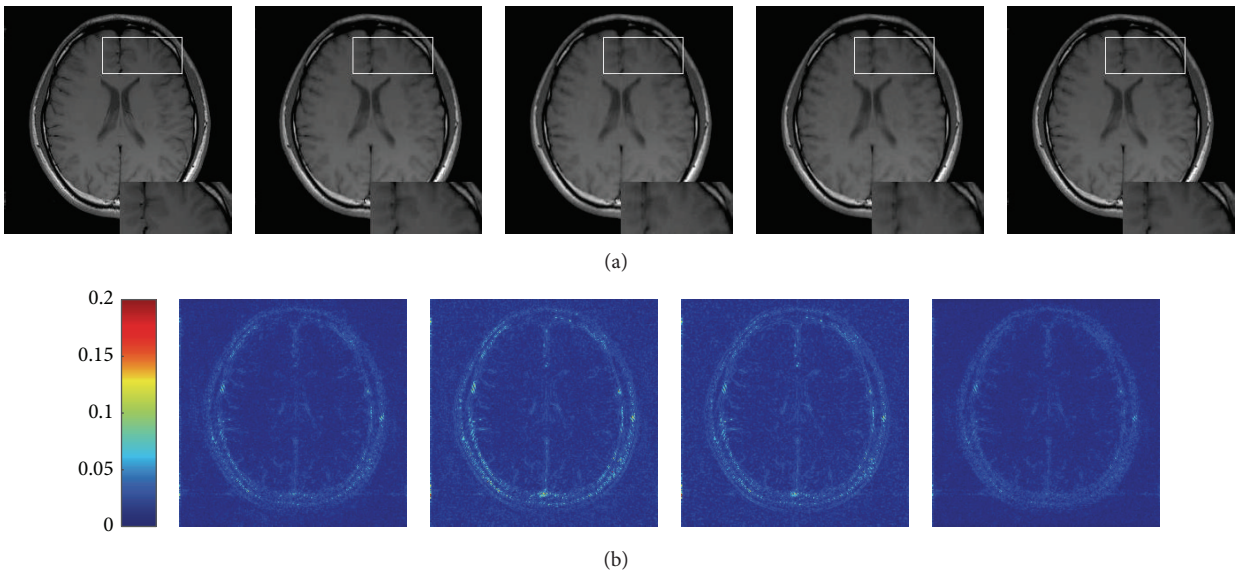


Figure 1: Visual quality comparison on GE MR images reconstructed by the four approaches from radially undersampled K-space data (25.16%). (a) From left to right: ground truth image and images reconstructed by the DDTF, DLMRI, GradDLRec, and proposed TRIMS; each one has an enlarged region for a closer comparison. (b) From left to right: color axis and difference images of the DDTF, DLMRI, GradDLRec, and TRIMS.

their recommended parameter settings. To quantitatively evaluate the reconstruction accuracy of each method, we have employed peak signal-to-noise ratio (PSNR), relative error, and structural similarity (SSIM) index [18] which are defined as follows:

$$\begin{aligned} \mathrm{PSNR} &= 20\log_{10}\frac{\max(u_0)\sqrt{Q}}{\|u_0-\hat{u}\|_2}, \\ \mathrm{err} &= \frac{\|u_0-\hat{u}\|_2}{\|u_0\|_2}, \\ \mathrm{SSIM} &= [l(\hat{u},u_0)]^{\alpha}\cdot[c(\hat{u},u_0)]^{\beta}\cdot[s(\hat{u},u_0)]^{\gamma}, \end{aligned} \tag{12}$$

where SSIM is multiplicative combination of the three terms, namely, the luminance term $l(\hat{u},u_0)$, the contrast term $c(\hat{u},u_0)$, and the structural term $s(\hat{u},u_0)$.

We firstly applied the four approaches to reconstruct T1-weighted MR image under the radial sampling scheme with the acceleration factor $R = 4$ (sampling ratio 25.16%). The reconstructed image obtained by each algorithm and the absolute difference between the reconstructed image and the ground truth image were displayed in Figure 1. We also present an enlargement area to reveal the fine details and structures each method has preserved. We can see that there exist somewhat blurring artifacts on the edges in the results reconstructed by the four methods. However, TRIMS can reconstruct an image closer to the one reconstructed from the full data. The absolute difference maps also indicate that TRIMS incurs less errors while reconstructing the MR image compared to the other three approaches.

We further utilized the four approaches to reconstruct the PD-weighted brain image from 9.13% of 2D randomly sampled K-space data. Figure 2(a) displays the original image and the images reconstructed by the four approaches. For a close-up look, the white box enclosed part has been zoomed and presented at the right corner of the image. It can be observed that our method has produced an image closer to the original image. The four approaches were also evaluated on a scanned physical phantom which consists of quite a few regular structures with fine details. Figure 2(b) provided the visual comparison results of the phantoms reconstructed from 12.79% of 2D randomly sampled K-space data. An area with different scales of lines was enlarged in each image to visualize the reconstruction accuracy of each method. It can be observed that the enlarged parts in the reconstruction results suffer from blur. Nevertheless, the proposed method can still produce an image with less blurry artifacts.

To test the sensitivity of the four methods to acceleration factors, we retrospectively undersampled the full K-space data with the 2D variable density scheme at 2.5-, 4-, 6-, 8-, and 10-time acceleration and employed the four methods to reconstruct MR images from the undersampled data. Figure 3 has presented the average PSNR, relative error values, and SSIM over all the three images reconstructed by the four methods versus different acceleration factors. The two PSNR and relative error plots have demonstrated that the proposed method could achieve better reconstruction results at all acceleration rates. Nevertheless, we should admit that the plot of SSIM indicates that the proposed method does not produce the best results at all undersampling factors on average since the current tight frame size is relatively small based on the concern of the computational complexity. Better results can be produced if the size of the tight frame is set a little bigger.

We also have provided a comparison of the convergence property of the four methods over acceleration rates 2.5 and 6 on the T1-weighted image in Figure 4. As can be seen, the four methods all have approximately converged.

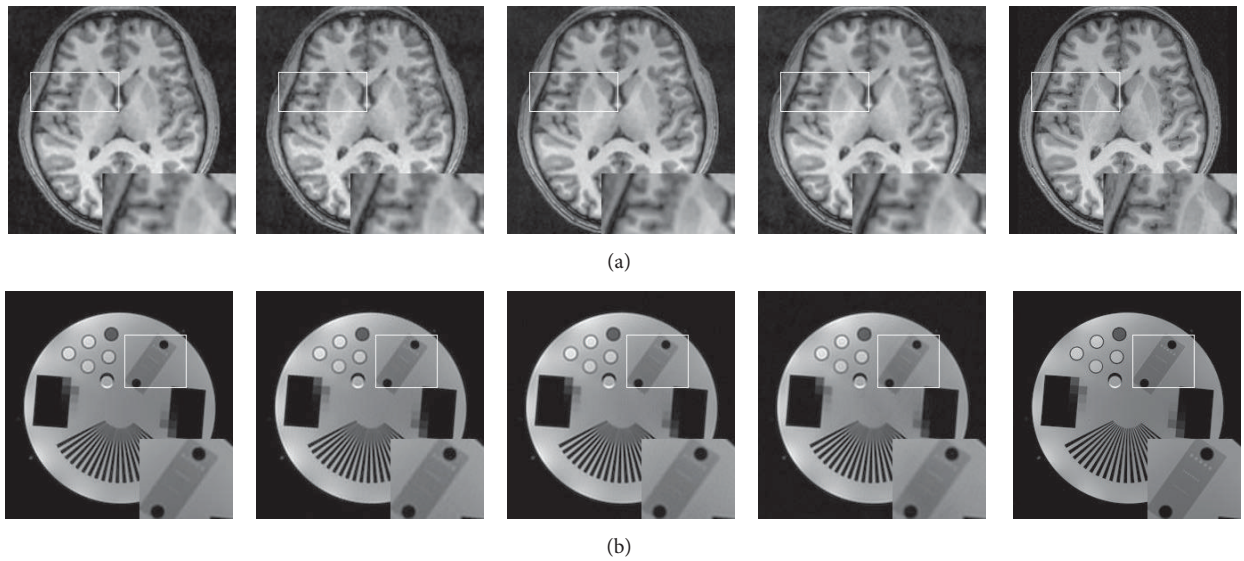


Figure 2: Visual quality comparison on PD-weighted and physical phantom MR images reconstructed by the four approaches from 2D randomly undersampled K-space data (9.312%). From left to right: ground truth image and images reconstructed by the DDTF, DLMRI, GradDLRec, and proposed TRIMS; each one has an enlarged region for a closer comparison.

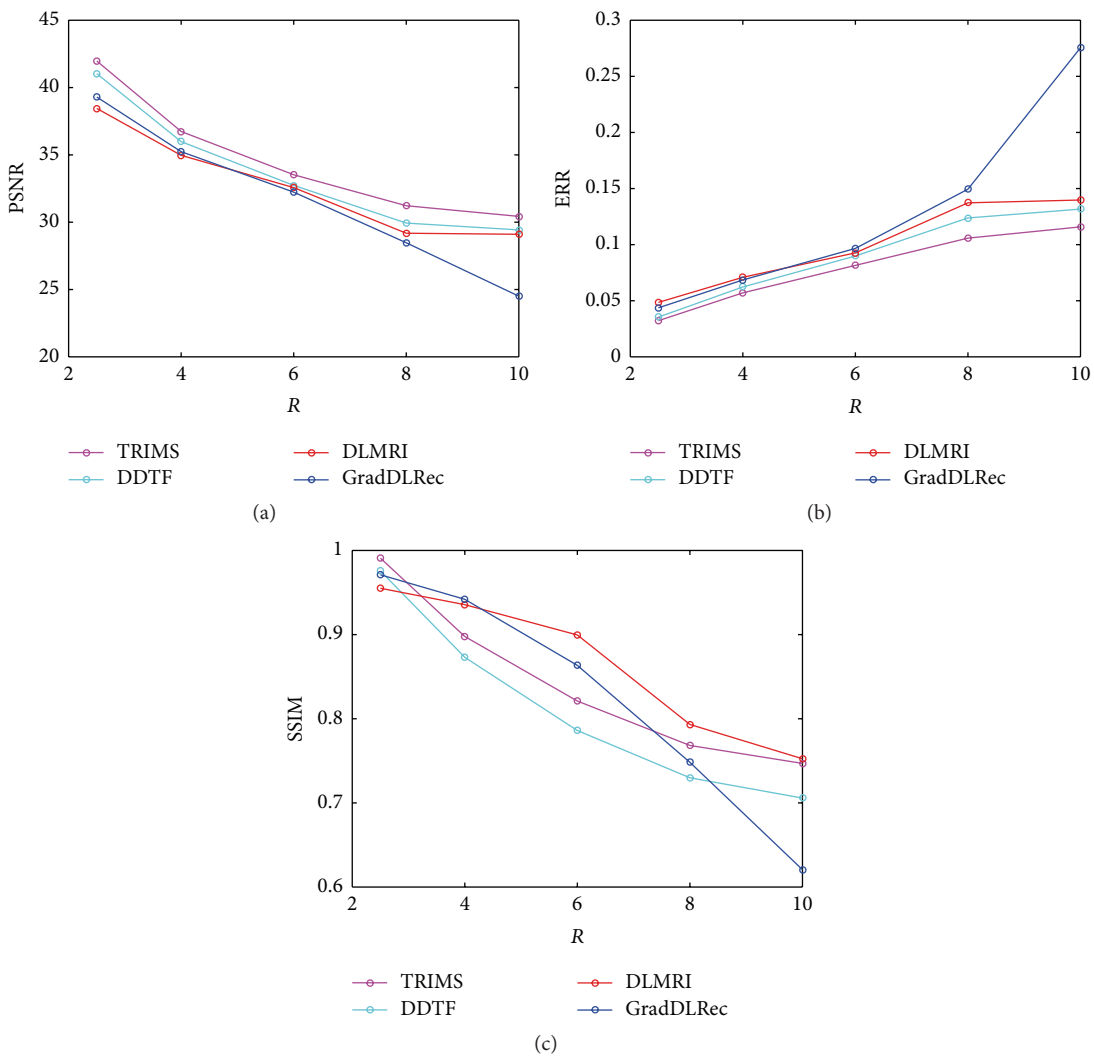


Figure 3: The average reconstruction errors in PSNR, relative error, and SSIM over all images with respect to different acceleration rates.

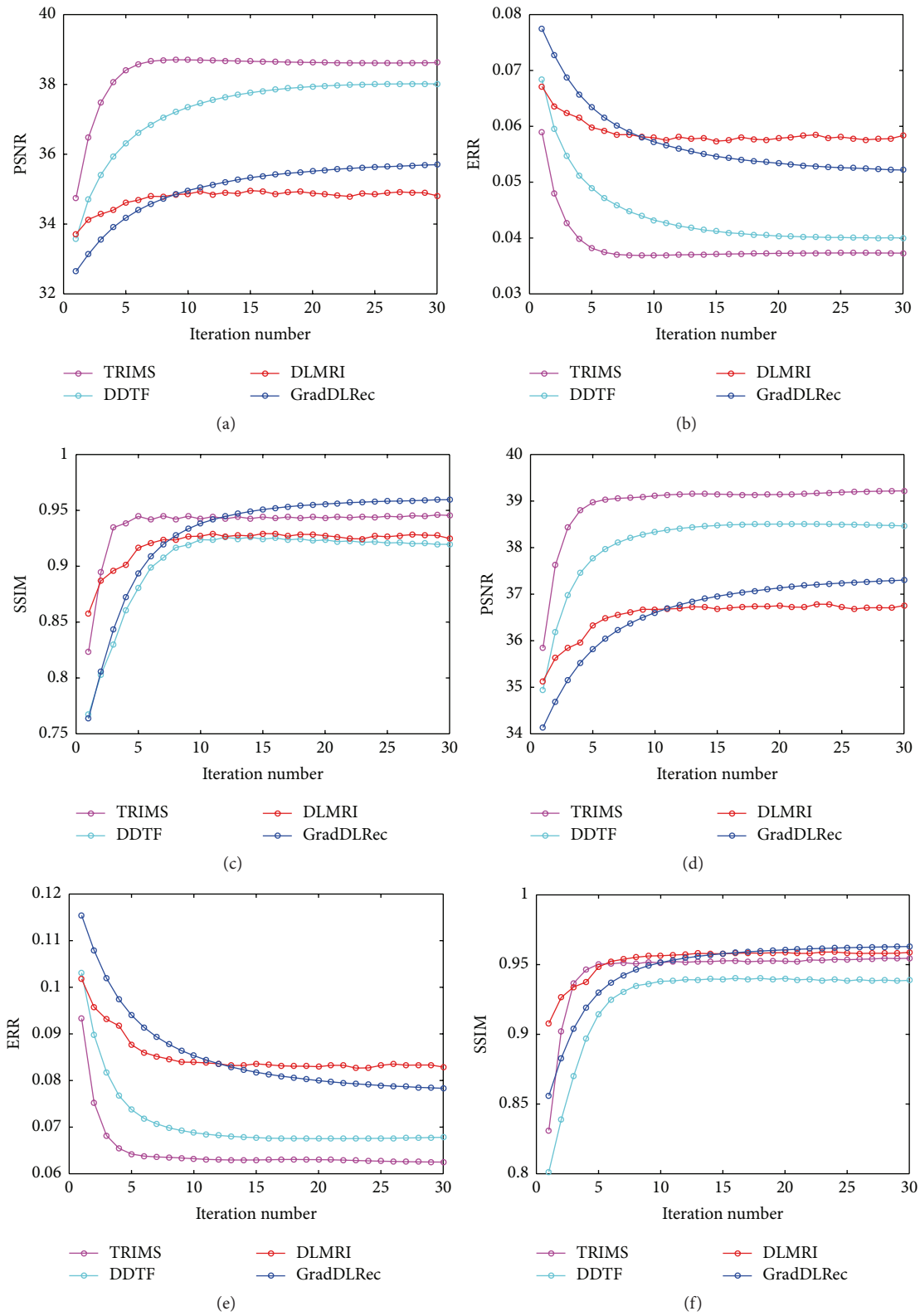


Figure 4: The convergence development of the four methods over acceleration rates 2.5 and 6 in PSNR, relative error, and SSIM while reconstructing the T1-weighted image.

Table 1: The computational time (in second) comparison over different acceleration rates.

Image	PD-weighted image					
Acceleration rate	2.5	4	6	8	10	20
TRIMS	137	139	140	139	137	137
DDTF	148	149	148	148	148	148
DLMRI	1294	1234	1215	1205	1188	1161
GradDLRec	2644	2475	2386	2352	2338	2298

Finally, we compare the computational time of the four methods, which were implemented on a Windows 7 (64-bit) operating system equipped with 8 GB RAM and Intel® Core™ i7-4770 CPU @ 3.40 GHz in MATLAB 2015a. Table 1 lists the computational time for each method over the six acceleration rates. We can observe that TRIMS is more efficient compared to DLMRI and GradDLRec. It is even more efficient than DDTF since DDTF needs to train 64 filters, each size of which is 8×8, while TRIMS only needs to train 16 filters whose size is 4×4. Furthermore, it is worth mentioning that although the size of the to-be-learned tight frame of TRIMS is smaller than that of DDTF, the two-layer sparsifying nature has facilitated TRIMS to achieve better reconstruction results in shorter time compared to DDTF.

4. Conclusions

This paper proposes a two-layer tight frame sparsifying model, namely, TRIMS, for compressed sensing magnetic resonance imaging. This approach explores the strength of adaptive learning technique and tight frames for accurate reconstruction of MR images from undersampled K-space data. The experimental results demonstrated that the proposed TRIMS could accurately reconstruct MR images from a variety of undersampled data with proper efficiency.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research was partly supported by the National Natural Science Foundation of China (61601450, 61471350, 11301508, 61671441, and 61401449), the Natural Science Foundation of Guangdong (2015A020214019, 2015A030310314, and 2015A030313740), the Basic Research Program of Shenzhen (JCYJ20160531183834938, JCYJ20140610152828678, JCYJ20150630114942318, and JCYJ20140610151856736), and the SIAT Innovation Program for Excellent Young Researchers (201403 and 201313).

References

[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: the application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[2] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic l0-minimization," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 106–121, 2009.

[3] Q. Liu, S. Wang, K. Yang, J. Luo, Y. Zhu, and D. Liang, "Highly undersampled magnetic resonance image reconstruction using two-level Bregman method with dictionary updating," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1290–1301, 2013.

[4] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen, "Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator," *Medical Image Analysis*, vol. 18, no. 6, pp. 843–856, 2014.

[5] D. Liang, H. Wang, Y. Chang, and L. Ying, "Sensitivity encoding reconstruction with nonlocal total variation regularization," *Magnetic Resonance in Medicine*, vol. 65, no. 5, pp. 1384–1392, 2011.

[6] X. Qu, D. Guo, B. Ning et al., "Undersampled MRI reconstruction with patch-based directional wavelets," *Magnetic Resonance Imaging*, vol. 30, no. 7, pp. 964–977, 2012.

[7] C. Chen and J. Huang, "The benefit of tree sparsity in accelerated MRI," *Medical Image Analysis*, vol. 18, no. 6, pp. 834–842, 2014.

[8] J. Liu, S. Wang, X. Peng, and D. Liang, "Undersampled MR image reconstruction with data-driven tight frame," *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 424087, 10 pages, 2015.

[9] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.

[10] Y. Huang, J. Paisley, Q. Lin, X. Ding, X. Fu, and X.-P. Zhang, "Bayesian nonparametric dictionary learning for compressed sensing MRI," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5007–5019, 2014.

[11] Y. Liu, J.-F. Cai, Z. Zhan et al., "Balanced sparse model for tight frames in compressed sensing magnetic resonance imaging," *PLoS ONE*, vol. 10, no. 4, Article ID e0119584, 2015.

[12] B. Ophir, M. Lustig, and M. Elad, "Multi-scale dictionary learning using wavelets," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1014–1024, 2011.

[13] Q. Liu, S. Wang, L. Ying, X. Peng, Y. Zhu, and D. Liang, "Adaptive dictionary learning in sparse gradient domain for image recovery," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4652–4663, 2013.

[14] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: learning sparse dictionaries for sparse signal approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, part 2, pp. 1553–1564, 2010.

[15] S. Ravishankar and Y. Bresler, "Learning doubly sparse transforms for images," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4598–4612, 2013.

[16] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye, "Data-driven tight frame construction and image denoising," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 89–105, 2014.

[17] X. Peng and D. Liang, "MR image reconstruction with convolutional characteristic constraint (CoCCo)," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1184–1188, 2015.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 8052436, 13 pages
http://dx.doi.org/10.1155/2016/8052436

Research Article

Pulmonary Nodule Detection Model Based on SVM and CT Image Feature-Level Fusion with Rough Sets

Tao Zhou,[1] Huiling Lu,[1] Junjie Zhang,[1] and Hongbin Shi[2]

[1] *School of Science, Ningxia Medical University, Ningxia, Yinchuan 750004, China*
[2] *Department of Urology, The Generel Hospital of Ningxia Medical University, Ningxia, Yinchuan 750004, China*

Correspondence should be addressed to Tao Zhou; zhoutaonxmu@126.com

Received 4 May 2016; Revised 23 July 2016; Accepted 2 August 2016

Academic Editor: Zexuan Ji

In order to improve the detection accuracy of pulmonary nodules in CT image, considering two problems of pulmonary nodules detection model, including unreasonable feature structure and nontightness of feature representation, a pulmonary nodules detection algorithm is proposed based on SVM and CT image feature-level fusion with rough sets. Firstly, CT images of pulmonary nodule are analyzed, and 42-dimensional feature components are extracted, including six new 3-dimensional features proposed by this paper and others 2-dimensional and 3-dimensional features. Secondly, these features are reduced for five times with rough set based on feature-level fusion. Thirdly, a grid optimization model is used to optimize the kernel function of support vector machine (SVM), which is used as a classifier to identify pulmonary nodules. Finally, lung CT images of 70 patients with pulmonary nodules are collected as the original samples, which are used to verify the effectiveness and stability of the proposed model by four groups' comparative experiments. The experimental results show that the effectiveness and stability of the proposed model based on rough set feature-level fusion are improved in some degrees.

1. Introduction

Lung cancer is a malignant tumor with the highest morbidity and mortality rate in the world, posing a serious threat to human life and health [1, 2]. The ability to estimate the risk of lung cancer is important in two common clinical models [3]: pulmonary nodules management and risk prediction model. Identification of early symptomatic in lung cancer is very important to improve early survival and reduce emergency presentations. Early detection is the most popular method to improve the effectiveness of the treatment of patients with lung cancer. Since pulmonary nodules are the early form of lung cancer [4], the detection of pulmonary nodules plays a critical role in the early diagnosis and treatment of lung cancer. Recent advances in computed tomography (CT) have a progressively increased spatial resolution and decreased acquisition times, making it possible for high resolution, multiangle, 3-dimensional, isotropic image of the whole lung to be acquired in less than 10 seconds. This has expanded capabilities for the early detection of small pulmonary nodules [4]. It is believed that early detection of lung cancer will result in earlier treatment at lower stages of the disease, thereby improving the 5-year survival rate, which has remained relatively constant at 15% for the last 30 years. However, with the wide application of CT in the lung imaging, the issues of CT data overloading and subjective interpretation of images result in a high clinical misdiagnosis rate [5].

Computer-Aided Diagnosis (CAD) systems provided a beneficial support and enhance the diagnostic accuracy. CAD is capable of performing the preliminary screen of the vast amounts of CT image and marking suspicious lesions, thereby helping radiologists to carry out the quadratic discrimination to reduce the workload and improve the accuracy rate of cancer diagnosis [6, 7].

Pulmonary nodule detection technology is one of the hot topics in the field of CAD in recent years. For example, ROI segment is a key problem, Xia et al. [8] using local variational Gaussian mixture models to segment brain MRI image Based on Learning Local Variational Gaussian Mixture Models, segmentation of breast ultrasound images are discussed by Xian

[9, 10] and Santos et al. [11] segment the lung parenchyma based on region growing algorithm. Magalhães Barros Netto et al. [12] use growing neural gas (GNG) to segment the lung parenchyma, the obtained pulmonary nodules are then separated from tissues containing blood vessels and bronchi according to the 3D distance transform, and finally SVM is used to carry out the effective identification of pulmonary nodules with shape and texture features. Ye et al. [13] firstly segment and extract region of interest (ROI) with fuzzy threshold in combination with Gaussian matrix, mean curvature, and Hessian matrix, then choose the local shape information and local intensity dispersion as the feature expression of ROI, and finally use the weighted SVM for recognition of pulmonary nodules. Tan et al. [14] segment pulmonary nodules based on the blood vessels and nodule enhancement filter proposed by Li et al. [15], then locate the clustering center of pulmonary nodules based on the divergence calculated by Gaussian template and achieve ROI extraction, and finally use the classifier based on genetic model, artificial neural network (ANN), and SVM for comparative analysis of the detection effectiveness of pulmonary nodules; Cascio et al. [16] use regional growth model and morphological operation to extract the ROI firstly, then reconstruct B-spline surface based on 3D spring model in order to extract the related 3D gray features and shape features, and detect the pulmonary nodules using ANN. Although the above literature explores the methods of detecting pulmonary nodules, overall, these are still two disadvantages of these methods in feature structure design and feature set expression as follows.

(1) When extracting and quantifying feature for ROI, the feature structure design is irrational, reflected by the fact that the combination of global features and local features and the combination of two-dimensional and three-dimensional features are not fully considered.

(2) When fusing feature data, the compactness of feature expression is a difficult problem. Therefore, feature redundancy is usually not eliminated. Moreover, the feature-level fusion method without prior knowledge is rarely used.

Rough set theory was developed by Zdzislaw Pawlak in the early 1980s and can be regarded as a new mathematical tool for feature selection, feature extraction, and decision rule generation without prior knowledge. Rough sets provide the mechanism to find the minimal set of attributes required to classify the training samples. This minimal set of attributes is called reduct and contains the same knowledge as the original set of attributes in a given information system. Therefore, reducts can be used to obtain different classifiers. Wang et al. [17] present a framework for a systematic study of the rough set theory. Various views and interpretations of the theory and different approaches to study the theory are discussed. The relationships between the rough sets and other theories, such as fuzzy sets, evidence theory, granular computing, formal concept analysis, and knowledge spaces, are examined. Cost of disease prediction and diagnosis can be reduced by applying machine learning and data mining methods. Disease prediction and decision-making play a significant role in medical diagnosis. Udhaya Kumar and Hannah Inbarani [18] put forward a novel neighborhood rough set classification approach to deal with medical datasets. Experimental result of the proposed classification algorithm is compared with other existing approaches such as rough set, Kth-nearest neighbor, support vector machine, BP NN, and multilayer perceptron to conclude that the proposed approach is a cheaper way for disease prediction and decision-making. Feature Selection (FS) is a solution that involves finding a subset of prominent features to improve predictive accuracy and to remove the redundant features. Thus, the learning model receives a concise structure without forfeiting the predictive accuracy built by using only the selected prominent features. Therefore, nowadays, FS is an essential part of knowledge discovery. Inbarani et al. [19] proposed new supervised feature selection methods based on hybridization of Particle Swarm Optimization (PSO), PSO based Relative Reduct (PSO-RR), and PSO based Quick Reduct (PSO-QR) presented for the diseases diagnosis, in order to seek to investigate the utility of a computer-aided diagnosis in the task of differentiating malignant nodules from benign nodules based on single thin-section CT image data. In Shah et al. [20], CT images of solitary pulmonary nodules were contoured manually on a single representative slice by a thoracic radiologist. Two separate contours were created for each nodule, one including only the solid portion of the nodule and one including any ground-glass components. For each contour, 75 features were calculated that measured the attenuation, shape, and texture of the nodule. These features were then input into a feature selection step and four different classifiers to determine if the diagnosis could be predicted from the feature vector. Hassanien [21] discuss a hybrid scheme that combines the advantages of fuzzy sets and rough sets in conjunction with statistical feature extraction techniques. An application of breast cancer imaging has been chosen and hybridization scheme have been applied to see their ability and accuracy to classify the breast cancer images into two outcomes: cancer or noncancer.

Based on the above reasons, a pulmonary nodule detection model based on rough set (RS) feature-level fusion and SVM is proposed in this paper. To overcome the first aforementioned disadvantage, the shape feature, intensity feature and texture feature are extracted. For shape feature, three new 3-dimensional features, namely, External Spherical Volume (ESV), Surface-Center Distance Standard Deviation (SCD-STD), and External Rectangle Cross Line Distance (ERCLD) are proposed. For intensity feature, three new 3-dimensional features, namely intensity gradient (from inside to outside), Laplace Divergence Mean (LDM), and Laplace Divergence Distance (LDD) are proposed. Regarding feature description, two-dimensional texture feature, three-dimensional shape feature, and intensity feature are used for quantification. With regard to the second aforementioned disadvantage, rough set feature-level fusion is adopted since it can fully retain the properties of the features without prior knowledge. Finally, a grid optimization model is employed to optimize the kernel function of support vector machine (SVM), which is used to conduct the recognition and detection of pulmonary nodules. In order to verify the validity and stability, advantages of

the model, four groups of comparative experiments are performed in this paper, that is, model validation experiments before and after rough set reduction, model stability experiments before and after rough set reduction, validation experiments of the superiority of the rough set feature-level fusion model, and comparative experiments with other pulmonary nodule detection models to compare the performance. The experimental results show that the method proposed in this paper can improve, to a certain extent, the rationality of feature structure and compactness of feature expression, thereby improving the detection accuracy of pulmonary nodules.

2. Related Theory

The description of ROI features is determined by both its comprehensiveness (features cannot be "observed" with "multiperspective" approach if the features amount is too little) and the accuracy of characterization (more quantized values diverged from the real information will cause a low feature discrimination). A large number of noise information sets will reduce the ROI feature extraction accuracy and affect the final results of detection. Therefore, for comprehensive and accurate expression of the morphological structure of ROI and local features, six new 3-dimensional features are proposed based on the analysis of ROI for lung CT image. These new 3-dimensional features are used to qualitatively analyze and quantitatively characterize the lesions from 2-dimensional and 3-dimensional perspectives in combination with other shape features, intensity features, texture features.

2.1. Pulmonary Nodules Features in CT Image

2.1.1. Shape Characteristics. Shape characteristics analyze the spatial distribution of gray values, by computing local features at each point in the image. Shape feature is the most intuitive visual feature, which can be used to describe the main medical signs of CT image of pulmonary nodule ROI, such as nodule sign, lobulation sign, spinous process sign, vacuole sign, and spicule sign, from the perspectives of geometric shape, edge roughness, and topology structure. In this paper the extracted components of the shape features mainly include perimeter, area, volume, roundness, rectangularity, elongation, Euler number, Harris, Hu moment, ESV, SCDSTD, and ERCLD. Here some features are given [22]:

(1) Area

$$S = \sum_{x=1}^{N}\sum_{y=1}^{M} f(x, y), \tag{1}$$

where $f(x, y)$ is the pixels of the target and M and N are the length and width, respectively.

(2) Perimeter

$$C = \sum_{i=1}^{M}\sum_{j=1}^{N} p(i, j), \tag{2}$$

where $p(i, j)$ is the pixels of the target edge and M and N are the length and width, respectively.

(3) Circularity

$$R_0 = \frac{C^2}{4\pi S}. \tag{3}$$

Circularity describes object shape that is close to the degree of circular, where S is the area of the target region and C is circumference of the target region. $0 < R_0 < 1$ and R_0 value reflects the complexity of the measurement boundary; the shape is more complex and the R_0 value is more smaller.

(4) Rectangularity

$$R = \frac{S}{(H * W)}, \tag{4}$$

where S is the area of the target region and H and, W are the length and width, respectively.

(5) Elongation

$$E = \frac{\min(H, W)}{\max(H, W)}. \tag{5}$$

Elongation can distinguish different shapes of the images (such as circle, square, ellipse, thin and long, and short and wide), where H and W are the length and width, respectively.

(6) Euler Number

$$E = C - H, \tag{6}$$

where C is the number of connection parts and H is the number of holes.

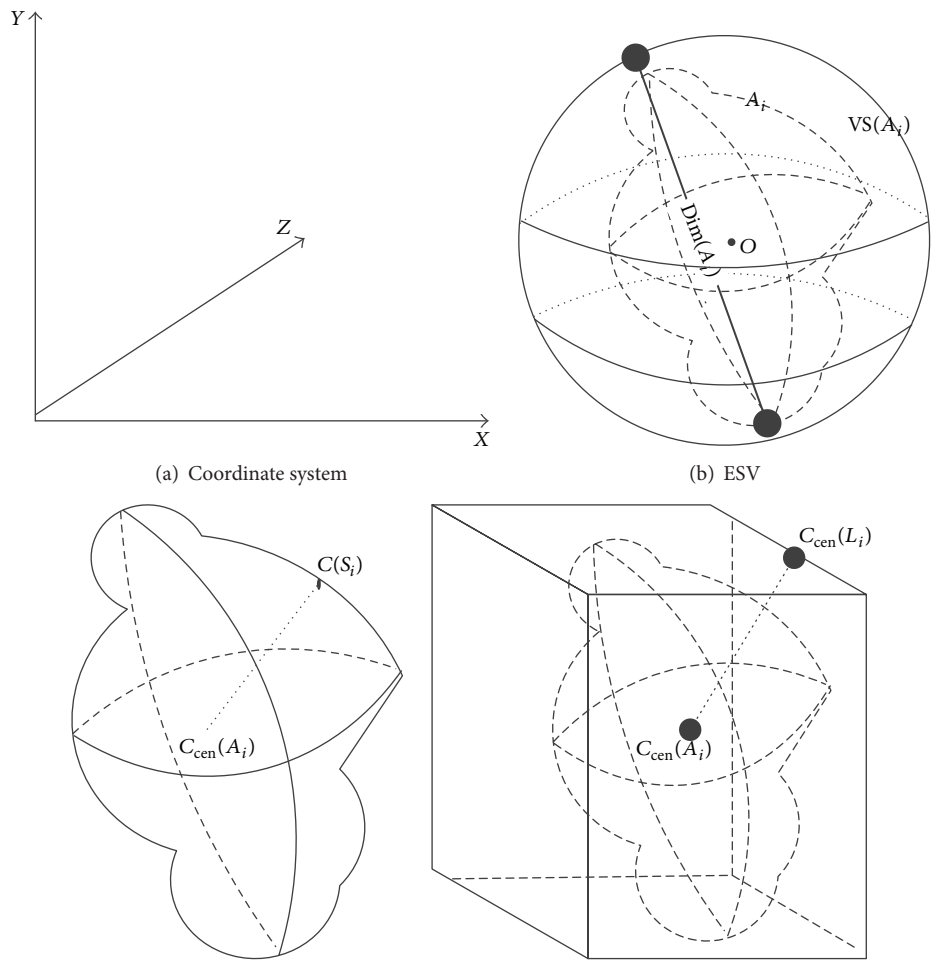
(7) External Spherical Volume (ESV). ESV is the ratio of each ROI A_i (maximum diameter is $\dim(A_i)$) to the External Spherical Volume $\mathrm{VS}(A_i)$ extracted from three-dimensional CT image, which reflects the similarity between the region and the sphere, as shown in Figure 1(b).

$$\begin{aligned} \mathrm{Volume}(\mathrm{VS}(A_i)) &= \frac{4}{3} \times \pi \times \left(\frac{\dim}{2}\right)^3 \\ E_1(A_i) &= \frac{\mathrm{Volume}(A_i)}{\mathrm{Volume}(\mathrm{VS}(A_i))}. \end{aligned} \tag{7}$$

(8) Surface-Center Distance STandard Deviation (SCDSTD). SCDSTD is the coordinate distance standard deviation of each individual element $C(S_i)$ and regional center $C_{\mathrm{cen}}(A_i)$ from the surface of each ROI; its value also describes the similarity with sphere of ROI. If the value is 0, $E_2(A_i)$ is a standard sphere. With the increase in $E_2(A_i)$ value, the magnitude of the deviation from the sphere in the region increases, as shown in Figure 1(c).

$$E_2(A_i) = \mathrm{std}\left(\frac{\|C(S_i) - C_{\mathrm{cen}}(A_i)\|}{\mathrm{mean}(\|C(S_i) - C_{\mathrm{cen}}(A_i)\|)}\right). \tag{8}$$

(9) External Rectangle Cross Line Distance (ERCLD). ERCLD is the distance from center voxel $C_{\mathrm{cen}}(A_i)$ of ROI to the center



(a) Coordinate system (b) ESV

(c) SCDSTD (d) ERCLD

Figure 1: Three-dimensional character sketch.

$\dim(L_i)$ $(i = 1, 2, \ldots, 12)$ of its 12 intersecting lines, which may indicate that the regional voxel is evenly distributed in the rectangular body, as shown in Figure 1(d).

$$E_3(A) = \frac{\left\| \text{mean}\left(C_{\text{cen}}(A_i) - C_{\text{cen}}(\dim(L_i))\right) \right\|}{\text{mean}\left(C_{\text{cen}}(\dim(L_i))\right)}. \quad (9)$$

2.1.2. Hu Moment Characteristics. Moments and the related invariants have been extensively analyzed to characterize the patterns in images. The moment invariants are independent of position, size, and orientation but also independent of parallel projection. Hu [23] was the first person to prove the central moment invariants. The central geometric moment invariants are derived based upon algebraic invariants, including six absolute orthogonal invariants and one skew orthogonal invariant. The moment invariants have been proved to be the adequate measures for tracing image patterns about the images translation, scaling, and rotation.

Hu moment invariants define seven values, computed by normalizing central moments through order three, which are invariant to object scale, position, and orientation, and a large number of papers that have significant contribution to the application of Hu moment. Two-dimensional moments of a digitally sampled $M * N$ image that has gray function $f(x, y)$ $(x = 1, 2, \ldots, M,\ y = 1, 2, \ldots, N)$ are given as

$$M_{p,q} = \sum_{x=1}^{M}\sum_{y=1}^{N} x^p y^q f(x, y) \quad p, q = 1, 2, 3, \ldots. \quad (10)$$

The moments $f(x, y)$ translated by an amount (a, b) are defined as

$$\begin{aligned} \overline{x} &= \frac{m_{10}}{m_{00}}, \\ \overline{y} &= \frac{m_{01}}{m_{00}} \\ \mu_{p,q} &= \sum_{x=1}^{M}\sum_{y=1}^{N} (x - \overline{x})^p (y - \overline{y})^q f(x, y) \\ &\qquad p, q = 1, 2, 3, \ldots. \end{aligned} \quad (11)$$

When a scaling normalization is applied, the central moments change as

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{00}^{\gamma}}, \quad \gamma = \left(\frac{(p+q)}{2}\right) + 1. \tag{12}$$

In terms of the central moments, the seven moments are given as

$$\begin{aligned}
C_1 &= \eta_{20} + \eta_{02} \\
C_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
C_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
C_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2 \\
C_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \\
&\quad \cdot \left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2\right] + (3\eta_{21} - \eta_{03}) \\
&\quad \cdot (\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2\right] \\
C_6 &= (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2\right] \\
&\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
C_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \\
&\quad \cdot \left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2\right] + (3\eta_{12} - \eta_{30}) \\
&\quad \cdot (\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2\right].
\end{aligned} \tag{13}$$

Hu 7-moment invariants vary widely, in order to compare, using logarithmic function to compress data, and hence the actual invariants moment features are C'_K:

$$C_K = \left|\log_{10}\left|C'_K\right|\right| \quad K = 1, 2, \ldots, 7. \tag{14}$$

The amended moment invariant features possess translation invariance, rotational invariance, and scale invariance.

2.1.3. Texture Characteristics. Tamura texture features, Tamura texture based on human visual perception in psychological research, are proposed by Tamura in 1978. Six components of Tamura texture feature correspond with 6 properties in psychology, three of them are coarseness, contrast, and directionality, which have the good application value in the texture synthesis, image recognition, and so on.

Texture is the gray distribution which appears repeatedly in the space position, so there are some relationships between two pixels at some distance from each other in image space, called gray spatial correlation properties in gray image. GLCM is a common method by studying the relevant relationship of gray image.

2.1.4. Intensity Features. Gray statistical feature is a quantitative method to describe the basic features of two-dimensional image region; it is called intensity feature from three-dimensional perspective [16]. In this paper, the extracted components of intensity features include the mean intensity, intensity variance, maximum and minimum intensity difference, skewness, kurtosis, intensity gradient (from inside to outside), Laplace Divergence Mean (LDM), and Laplace Divergence Distance (LDD).

(1) Intensity Gradient (from Inside to Outside). For ROI A_i with the voxel S_i volume greater than 0, morphological erosion processing is performed continuously and the ratio of the mean of the excluded area of each erosion processing to the mean of the last operation (initial value is 0) is calculated until the ratio is zero. Consider the following equation where n is the number of operations.

$$E_4(A_i) = \frac{K}{n}. \tag{15}$$

(2) Laplace Divergence Mean (LDM). According to the Laplacian convolution results with the original CT image, it is found that the nodule surrounding area with smaller gray value difference has a significant different divergence. Therefore, calculation of Laplace divergence is helpful to distinguish pulmonary nodules from interfering impurities.

$$E_5(A_i) = \text{mean}(A_i \times \text{La}). \tag{16}$$

(3) Laplace Divergence Distance (LDD). The difference between the maximum and minimum values of the Laplace divergence values is used to describe the range of regional divergence.

$$E_6(A_i) = \max(A_i \times \text{La}) - \min(A_i \times \text{La}). \tag{17}$$

Table 1 shows the feature set of 42 features based on the above feature description of ROI. To facilitate subsequent tests, features are numbered in the order as showed in Table 1; that is, the shape features are numbered fs1–fs18, the intensity features are numbered fi1–fi8, and texture features are numbered ft1–ft16, respectively.

2.2. Rough Set and Attribute Reduction. Rough set theory (RST), proposed by Pawlak in 1982, is one of the effective mathematical tools for processing fuzzy and uncertainty knowledge. Nowadays, RST has been applied to a variety of fields such as artificial intelligence, data mining, pattern recognition, and knowledge discovery. Rough set is founded on the assumption that with every object of the universe of discourse some knowledge is associated. Objects characterized by the same information are similar in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of rough set theory. Any set of all indiscernible objects are called an elementary set and form a basic granule of knowledge about the universe. Any union of some elementary sets is referred to as a crisp set, otherwise the set is rough set.

Definition 1. An information system S is a quadruple $S = (U, A, V, f)$, where U is a nonempty and finite set of objects, A is a nonempty and finite set of attributes, $V := \bigcup V_a$ with V_a being the domain of attribute a, and f is an information

Table 1: ROI feature set.

Feature type	Feature vectors	Dimensionality
Shape features (fs)	Perimeter, area, volume, roundness, rectangularity, length, Euler's number, ESV, SCDSTD, ERCLD, Hu moment	18
Intensity features (fi)	Mean intensity, intensity standard variance, maximum-minimum intensity difference value of variance, skewness, kurtosis, intensity gradient (from inside to outside), LDM, LDD	8
Texture features (ft)	Tamura texture features (contrast, direction, roughness), GLCM (angular second moment, moment of inertia, torque deficit, sum mean, variance, sum variance, difference variance, entropy, sum entropy, differential entropy, information measure, correlation coefficient, maximum correlation coefficient)	16

function such that $f(x,a) \in V_a$ for every $x \in U$ and every $a \in A$. A decision system is an information system $(U, C \cup D, V, f)$ with $C \cap D = \Phi$, where C and D are called the conditional and decision attribute sets, respectively.

For a subset P of A, let us define the corresponding equivalence relation as

$$\begin{aligned} \mathrm{IND}(P) = \{&(x,y) \in U \times U \mid f(x,a) \\ &= f(y,a) \text{ for any } a \in P\} \end{aligned} \tag{18}$$

and denote the equivalence class of IND(P) which contains the object $x \in U$ by $[x]_P$; that is,

$$[x]_P = \{y \in U \mid (x,y) \in \mathrm{IND}(P)\}. \tag{19}$$

The factor set of all equivalence classes of IND(P) is denoted by U/P; that is, $U/P = \{[x]_P \mid x \in U\}$.

As well known, attribute reduction is one of the key issues in RST. It is performed in information systems by means of the notion of a reduct based on a specialization of the notion of independence due to Marczewski. Up to now, much attention has been paid to this issue and many different methods of attribute reduction have been proposed for decision systems. For example, the reduction approaches are, respectively, based on partition, discernibility matrix, conditional information entropy, positive region, and ant colony optimization approach.

Definition 2. Let $S = (U, A, V, f)$ be an information system and $P \subseteq A$. For a subset X of U, $R_P(X) = \{x \in U \mid [x]_P \subseteq X\}$ and $R^P(X) = \{x \in U \mid [x]_P \cap X \neq \Phi\}$ are called P-lower and P-upper approximations of X, respectively.

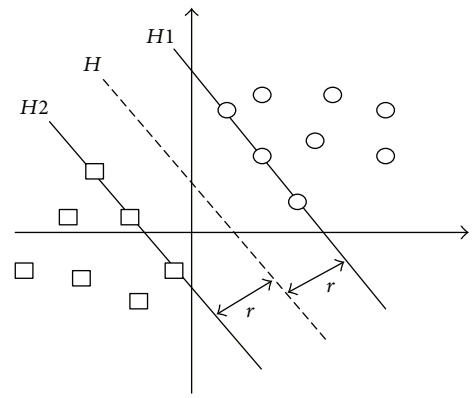


Figure 2: Optimal hyper plane.

Definition 3. Let $S = (U, A, V, f)$ be an information system and let P and Q be two subsets of A. Then, $\mathrm{POS}_P(Q) = \bigcup_{X \in U/Q} R_P(X)$ is called P-positive region of Q, where $R_P(X)$ is the P-lower approximation of X.

Definition 4. Let $S = (U, A, V, f)$ be a decision system, $a \in C$, and $P \subseteq C$. If $\mathrm{POS}_C(Q) = \mathrm{POS}_{C \setminus \{a\}}(Q)$, a is said to be D-dispensable in C; otherwise, a is said to be D-indispensable in C. The set of all the D-indispensable attributes is called the core of S and denoted by Core(S). Furthermore, if $\mathrm{POS}_P(Q) = \mathrm{POS}_C(Q)$ and each of the attributes of P is D-indispensable, then P is called a reduct of S.

2.3. SVM and Its Optimization. SVM is a pattern recognition method developed from statistical learning theory based on the idea of structural risk minimization principle. In the case of ensuring classification accuracy, SVM can improve the generalization ability of the learning machine by maximizing the classification interval. The biggest advantage of SVM is that it overcomes the overlearning and high dimension both of which lead to computational complexity and local extremum problems. A reliable classification model based on SVM is urgently needed for the study of hospitalization expenses of patients with gastric cancer.

SVM deals with linearly separable data (Figure 2); the assumption is that there are data sets $S = \{x_1, \ldots, x_n\}$ and data marker $G = \{y_1, \ldots, y_n\}$, where x_i is the input space vector of the data sample and y_i records the category of the sample.

The aim of SVM is to find an optimal hyper plane H to separate these two samples and make the largest interval. The optimal hyper plane H is expressed as

$$w^T x + b = 0, \tag{20}$$

where w is the weight vector and b is the threshold.

This problem is transformed into the optimal problem of w and b:

$$\begin{aligned} &\min_{w,b} \quad r(w) = \frac{1}{2}\|w\| \\ &y_i((w \cdot x) + b) \geq 1, \quad i = 1, \ldots, n. \end{aligned} \tag{21}$$

In order to simplify the formula, the Lagrange dual is introduced to meet the requirements of KKT (Karush-Kuhn-Tucker). The objective function is transformed into

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j \left(x_i \cdot x_j\right) - \sum_{j=1}^{n} \alpha_j \\ \text{s.t.} \quad & \sum_{i=1}^{n} y_i \alpha_i = 0, \quad \alpha_i \geq 0, \ i = 1, 2, \ldots, n. \end{aligned} \tag{22}$$

As for the linearly inseparable data, the penalty parameter C and relaxation variable ξ are introduced in the constraint condition, thus the generalization ability of SVM is increased, and the function is transformed into

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j \left(x_i \cdot x_j\right) - \sum_{j=1}^{n} \alpha_j \\ \text{s.t.} \quad & \sum_{i=1}^{n} y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \end{aligned} \tag{23}$$

where C is the artificial setting parameter. According to the practical experience, the bigger C, the greater separation interval. At the same time, it will increase the risk of generalization.

The final classification function is

$$f(x) = \operatorname{sgn}\left\{\left(\sum_{i=1}^{n} a_i^* y_i \left(x_i \cdot x\right)\right) + b^*\right\}. \tag{24}$$

For nonlinear classification data, SVM transforms them into linearly separable data in a high-dimensional space via nonlinear mapping of kernel function, and the optimal hyper plane is found in high-dimensional space. The kernel function which meets the mercer kernel condition corresponding to the transvection of a spatial transformation is used to realize the nonlinear transformation of linear classification.

The corresponding kernel function is defined as

$$K\left(x_i, x\right) = \left(\varphi\left(x_i\right), \varphi\left(x\right)\right). \tag{25}$$

At this point the final classification function is

$$f(x) = \operatorname{sgn}\left\{\left(\sum_{i=1}^{n} a_i y_i K\left(x_i \cdot x\right)\right) + b\right\}. \tag{26}$$

Penalty factor C and parameter g of the kernel function play an extremely important role in the performance of SVM classification. In order to obtain the optimal classification results, grid optimization model is used for optimization in this paper. In grid optimization model, the parameters to be searched are expressed in the form of grids in a certain space, and the optimal parameters are selected by traversing all the grids. Therefore, grid optimization model has the advantages of simplicity, convenience, good stability, and easiness to get the global optimal solution [24]. In the learning process of SVM, 10-fold cross-validation is used to calculate the kernel function parameters and penalty coefficient with the optimal classification performance, which are then applied to the SVM classifier for recognition and detection of pulmonary nodules. Finally, sensitivity, specificity, accuracy, and processing time are used as indexes to evaluate the detection of relevant experiments.

3. Pulmonary Nodule Detection Model

In this paper, CT images of 70 cases of patients with pulmonary nodules are used. The images are firstly segmented [7] to three different types of pulmonary nodules (solitary pulmonary nodules or SPN, vascular adhesion pulmonary nodules or VAPN, and pleural adhesion pulmonary nodules or PAPN), which are marked by radiologists, as well as a large number of nonnodular areas, including blood vessels, bones, and alveoli. Forty-two feature components characterizing ROI are extracted from the 2-dimensional and 3-dimensional perspectives, including six new 3-dimensional features proposed in this paper. They are composed of 18 shape features, 8 intensity features, and 16 texture features. The extracted feature set (identified as the FS) is discretized and normalized. Feature-level fusion of the improved feature data is performed for five times using rough set model (since the reduction of rough set feature subset is not unique, in this paper, the extracted feature sets are reduced for five times and are identified as RS1, RS2, RS3, RS4, and RS5). Feature subset RS1 is used for comparative experiment. Finally, SVM parameters are optimized using grid optimization model, and the improved SVM is used in the following four sets of comparative experiments: comparative analysis of the effectiveness and stability of classification before and after rough set reduction of features; comparative analysis of the recognition performance before and after feature-level fusion based on rough set or PCA; comparative analysis of the recognition performance of our proposed method and other methods. Based on the above views, we present a flow chart of pulmonary nodule detection model as shown in Figure 3.

4. Results and Discussion

4.1. Experimental Environments. In this paper, the hardware and software environments are as follows.

Software Environments. Windows 7 OS, the Matlab R2014b, ImageJ 1.48 u, and LibSVM.

Hardware Environments. Intel Core i5 4670-3.4 GHz, 8.0 GB of memory, and 500 GB hard disk.

Experimental Data. CT images of 70 cases of patients with pulmonary nodules are collected as experimental samples, which are marked by radiologists, with a size of 512 × 512 and a thickness of 2 mm. They are composed of 2232 CT images from 38 cases of patients with solitary pulmonary nodules (SPN), 17 cases of patients with vascular adhesion pulmonary nodules (VAPN), and 15 cases of patients with pleural adhesion pulmonary nodules (PAPN), respectively.

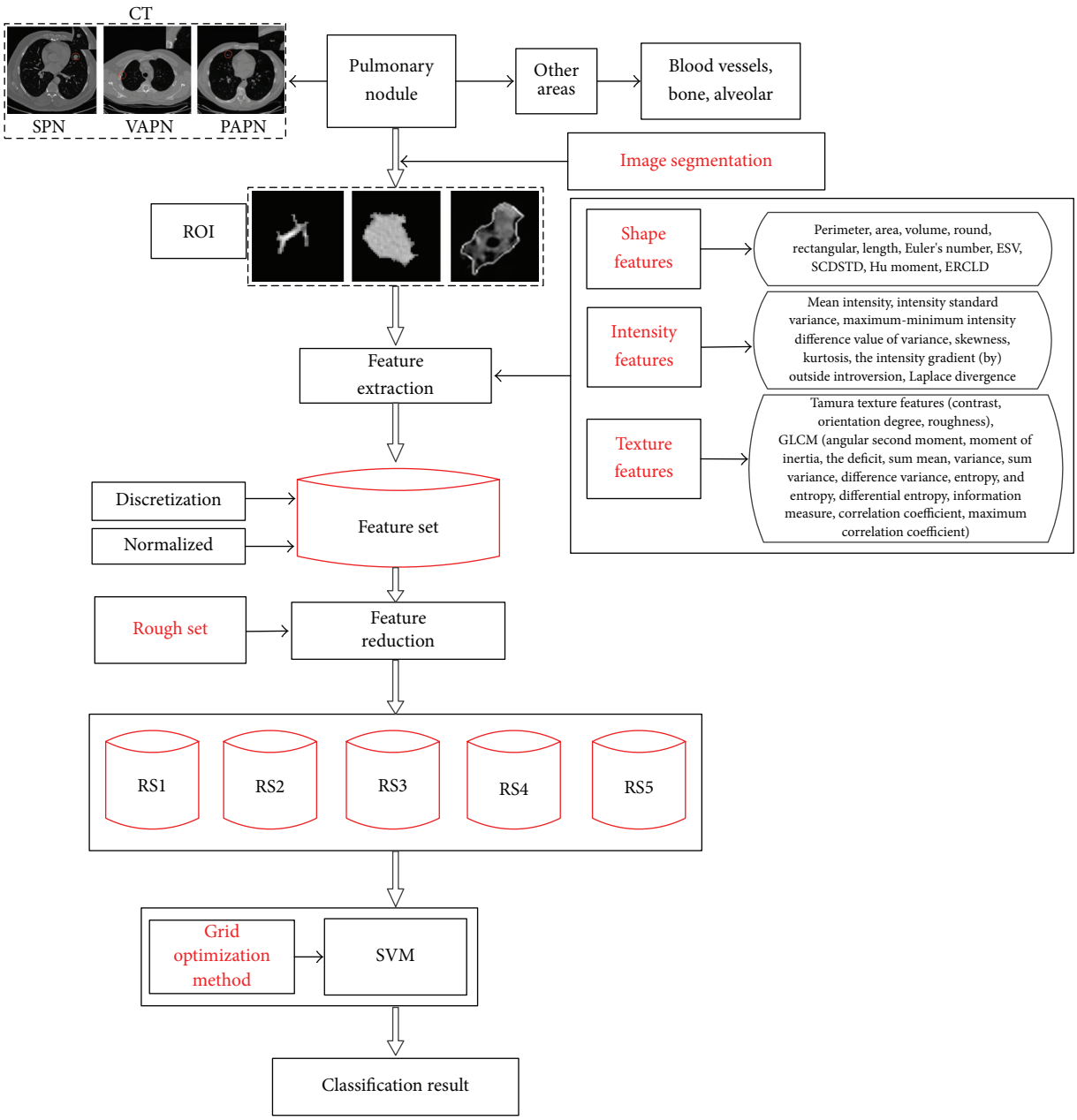
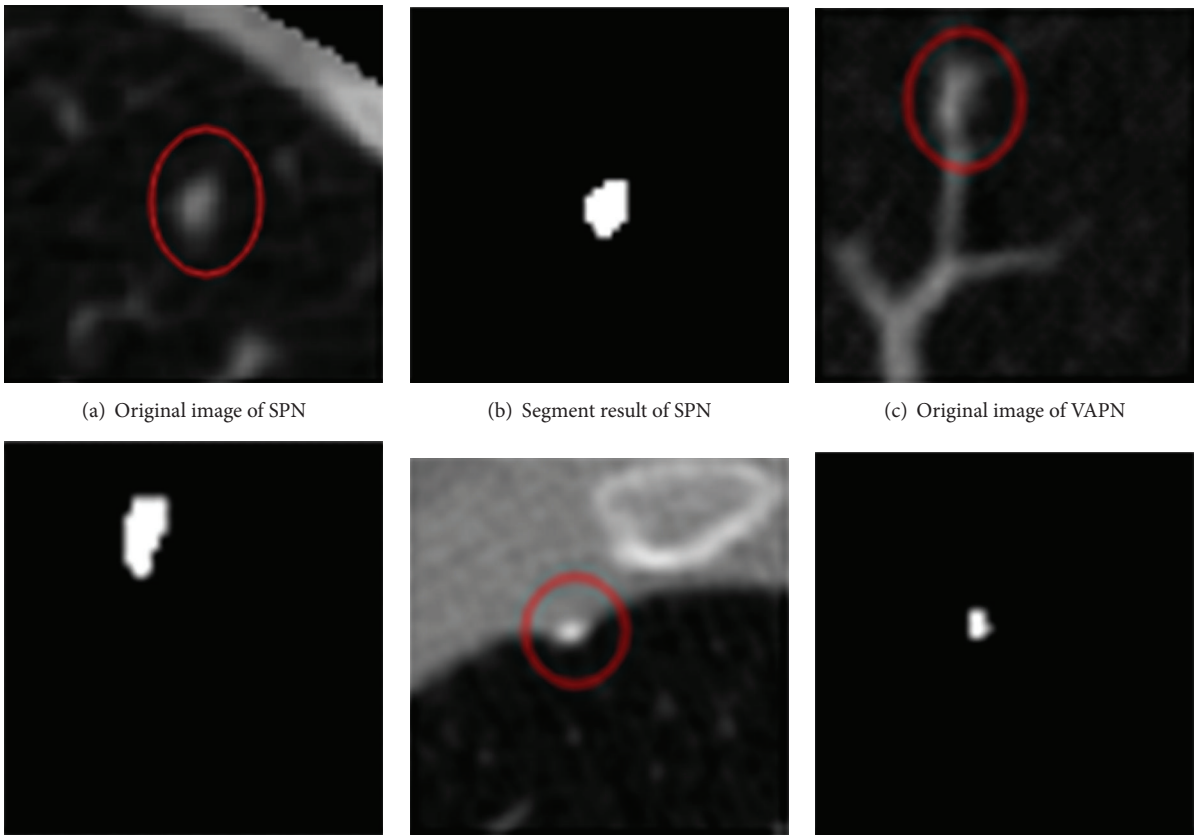


Figure 3: Flow chart of pulmonary nodule detection model.

Figure 4 shows the representatives of each type of pulmonary nodules and the corresponding segmentation results.

In this paper, 42-dimensional features of 70 marked pulmonary nodular areas and 70 randomly selected nonnodular areas are extracted. Table 2 shows the 42-dimensional feature values of the lung nodular and nonnodular areas. shape features are identified as the fs, intensity features are identified as the fi, and texture features are identified as the ft. In order to intuitively understand the distribution of different feature values and the discrimination comparison, external sphere volume (ESV) ratio and the standard deviation of surface-center distance (SCD) are calculated and plotted as box diagram as shown in Figure 5.

4.2. Feature-Level Fusion Based on Rough Set. In order to avoid the attribute value of small range of values dominated by that of large range of values and reduce the complexity of the statistical computation process, the extracted feature sets are firstly preprocessed by normalizing data with bigger difference and linearly mapping the data to $[0, 1]$. The preprocessed feature data are then fused for five times using rough set model. The fusion results are shown in Table 3.



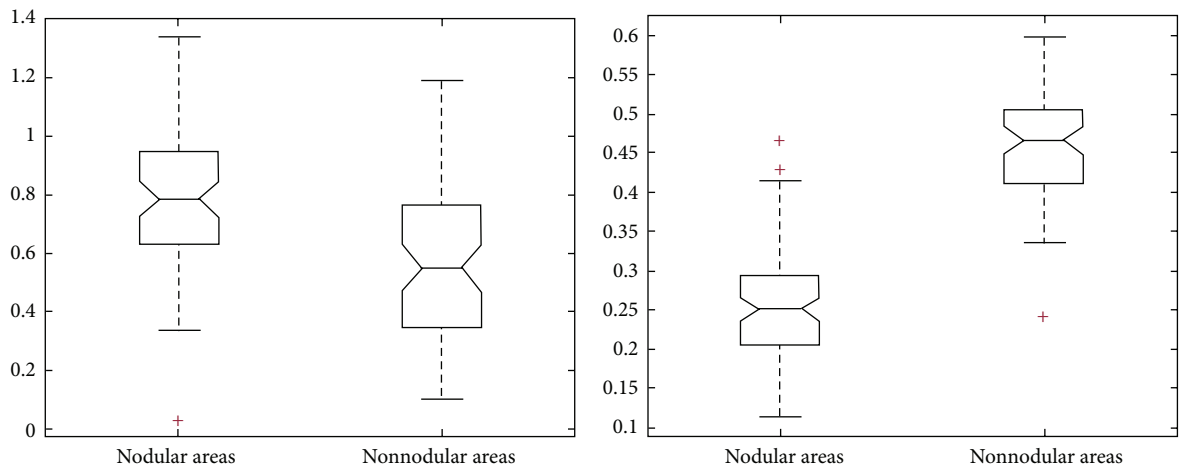
(a) Original image of SPN (b) Segment result of SPN (c) Original image of VAPN

(d) Segment result of VAPN (e) Original image of PAPN (f) Segment result of PAPN

Figure 4: Pulmonary nodule segmentation results.

Table 2: Feature values of pulmonary nodular areas and nonnodular areas.

Shape features (fs)		Intensity features (fi)		Texture features (ft)	
Nodular areas	Nonnodular areas	Nodular areas	Nonnodular areas	Nodular areas	Nonnodular areas
95	78	59.06	91.0987	8.3104	5.4016
159	128	14.06	4.4872	12.041	12.5216
284	178	0.5956	−0.39568	0.4303	0.0067
0.6517	0.211	2.7348	1.8669	0.7709	0.7275
0.6961	2.1587	55.1865	14.3481	0.7169	0.9865
0.3529	0.7778	0.5	1	0.8059	5.3894
0	1	13.9598	20.6044	0.1942	0.0487
0.3186	1.0295	729.905	354.6389	0.7708	0.7273
0.0686	1.0197			0.8059	5.3498
0.0042	0.0458			3.5042	5.0971
0.0021	0.0295			0.6514	0.8453
0.0013	0.0268			0.0971	0.6143
0.0005	0.0011			4.4033	82.1862
0	1			0.0691	5.0061
14	9			−0.5785	−0.4245
0.5356	0.5571			2.307	3.2239
0.3072	0.501788				
0.1738	0.207122				



(a) Boxplot of ESV (b) Boxplot of SCDSTD

Figure 5: Pulmonary nodule area and the pulmonary nodules boxplot. "+" refers to upper and lower bounders of ESV value and SCDSTD value.

Table 3: Feature reduction based on rough sets.

Feature subset	Reduction results	Dimensionality
RS1	fs4, fs16, fs17, fs18, fi2, fi4, fi6, fi7, fi8, ft2, ft4, ft5, ft6, ft7, ft8, ft9, ft10, ft11, ft13, ft14, ft15, ft16	21
RS2	fs4, fs9, fs16, fs18, fi1, fi2, fi5, ft2, ft5, ft6, ft8, ft9, ft10, ft11, ft12, ft13, ft15	17
RS3	fs9, fs17, fs18, fi1, fi2, fi5, fi7, fi8, ft2, ft6, ft7, ft8, ft9, ft10, ft11, ft12, ft14, ft15, ft16	19
RS4	fs9, fs16, fs18, fi1, fi2, fi5, fi7, fi8, ft5, ft6, ft7, ft8, ft9, ft10, ft11, ft12, ft14, ft15, ft16	19
RS5	fs9, fs16, fs17, fs18, fi1, fi2, fi4, fi5, fi7, fi8, ft2, ft5, ft6, ft7, ft8, ft9, ft10, ft12, ft15, ft16	20

4.3. Pulmonary Nodule Detection with SVM Based on Grid Optimization

4.3.1. The Model Effectiveness Experiment. Tenfold cross-validation is used to calculate the accuracy, sensitivity, specificity, and processing time of classification before and after rough set reduction (RS1(70×21) obtained from experiment one is used as the data set after reduction), and the recognition performance of classifier is compared before and after reduction. The results are shown in Table 4.

Experimental results show that pulmonary nodule detection accuracy is increased significantly after feature-level fusion, with a decrease in the missed diagnosis rate, reflected by the increased sensitivity, and the misdiagnosis rate, reflected by the increased specificity. The processing time is also shorter after reduction. These results indicate that the feature-level fusion of the extracted feature set with 42 dimensionalities based on rough set model is effective, which not only improves the compactness of the feature set (to eliminate redundancy and low degree of differentiation features component), but also corrects the abnormal data of the feature set, thereby further improving the performance of pulmonary nodule detection. Table 5 shows the effectiveness of the five rough set reduction subsets.

4.3.2. The Model Stability Experiment. The feature data of pulmonary nodules are tested with RS1($70 * 21$) as the dataset for classification for five rounds with a different ratio of training set over testing set of 50/20, 40/30, 35/35, 35/35, or 20/50. Each round of test is carried out with a randomly selected ratio of training set over testing set and the mean of 10 test results is used as the corresponding accuracy, sensitivity, specificity, and running time of the model. The results are shown in Table 6.

The experimental results show that, with the decrease in the ratio of training set over testing set, the decrease in the classification accuracy of feature subset after rough set reduction is not obvious, whereas that of feature set before rough set reduction is fluctuating to certain extent (Figure 6 is more intuitive). These results indicate that the classification stability of the feature level fusion model based on rough set is higher and is less susceptible to the interference of sample data. Table 7 shows the stability of 5 groups feature subset after rough set reduction.

4.3.3. The Superiority of Feature-Level Fusion Model Based on Rough Set. Since PCA is a well-developed model, characterized by simple calculation and easy programming, it has become the preferred dimension reduction method for most of the feature-level fusion model in order to analyze comparatively two types of feature-level fusions. In this paper, PCA-based feature-level fusion of the extracted feature sets is performed at the same time, and the tenfold cross-validation results are shown in Table 8. Figure 7 shows the classification performance of the two types of feature-level fusion methods

Table 4: Statistics of effectiveness before and after rough set reduction.

	Serial number	Accuracy (%)	Sensibility (%)	Specificity (%)	Processing time (s)
Before reduction	1	96.42	92.86	100	1.0610
	2	91.96	83.93	100	0.6170
	3	95.54	100	91.07	0.5490
	4	89.28	100	78.57	0.5630
	5	95.54	91.07	100	0.5470
	6	98.21	96.43	100	0.5460
	7	94.64	89.29	100	0.5460
	8	95.53	91.07	100	0.5460
	9	91.96	83.93	100	0.5460
	10	97.32	100	96.64	0.5300
	Mean	94.64	92.86	96.43	0.6051
After reduction (Rs1)	1	100	100	100	0.9370
	2	100	100	100	0.4360
	3	100	100	100	0.3870
	4	100	100	100	0.4210
	5	100	100	100	0.4210
	6	100	100	100	0.3900
	7	100	100	100	0.4060
	8	91.67	100	83.33	0.4060
	9	100	100	100	0.3740
	10	100	100	100	0.3930
	Mean	99.17	100	98.33	0.4571
Increase after reduction		4.53	7.14	1.9	0.148

Table 5: Effectiveness of rough set reduction subsets.

Subset	Average accuracy (%)	Average sensitivity (%)	Average specificity (%)	Processing time (s)
RS1	99.17	100	98.33	0.4571
RS2	97.5	96.67	98.33	0.4650
RS3	99.17	100	98.33	0.4656
RS4	100	100	100	0.4731
RS5	98.33	98.33	98.33	0.4850
Mean	98.83	99	98.66	0.4672

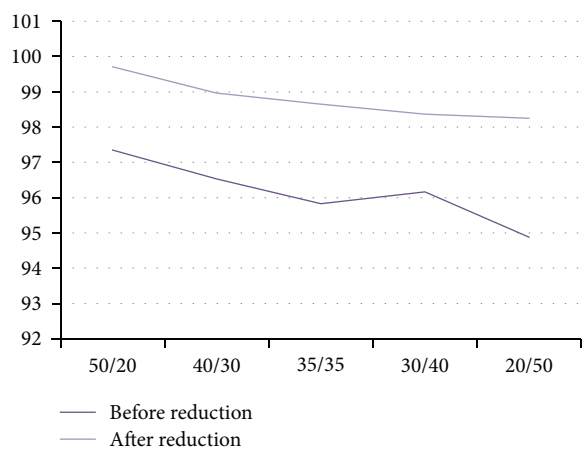


Figure 6: Comparative results of feature subsets before and after rough set reduction.

(feature subset RS1 from Table 3 is used, and the running time is 100 × actual time).

Experimental results show that various performance indicators of the feature-level fusion model based on rough set are better than those based on PCA, indicating that the rough set is more suitable than PCA to eliminate redundant information.

4.3.4. Comparison with Other Pulmonary Nodule Detection Methods. Pulmonary nodule detection accuracy and False Positives per scan (FP/s) are used as the evaluation indexes of pulmonary nodule detection methods to compare and analyze the method proposed in this paper and other five detection methods of pulmonary nodules (the optimal detection accuracy is used for all detection methods). The results are shown in Table 9 (Pr: private database; L: LIDC).

Experimental results show that the proposed method is superior to the other pulmonary nodule detection methods to a certain extent, indicating that this method not only improves the comprehensiveness and accuracy of the feature description of ROI by supplementing and improving the

Table 6: Stability statistics of rough set reduction subsets.

	Training set/testing set	Accuracy (%)	Sensitivity (%)	Specificity (%)	Running time (s)
Before fusion	50/20	97.35	94.71	100	0.4873
	40/30	96.53	93.08	98.32	0.3846
	35/35	95.83	92.39	97.79	0.4254
	30/40	96.16	95.58	96.74	0.3560
	20/50	94.88	94.63	95.86	0.4236
	Mean	96.15	94.08	97.742	0.4154
After fusion (Rs1)	50/20	99.71	99.41	100	0.2684
	40/30	98.96	99.58	98.46	0.2568
	35/35	98.65	99.23	98.08	0.2382
	30/40	98.37	98.60	98.14	0.2646
	20/50	98.25	97.67	98.84	0.2636
	Mean	98.79	98.84	98.70	0.2583

Table 7: Classification performance of rough set reduction subset.

Subset	Average accuracy (%)	Average sensitivity (%)	Average specificity (%)	Running time (s)
RS1	99.17	100	98.33	0.2583
RS2	97.5	96.67	98.33	0.2870
RS3	99.17	100	98.33	0.2560
RS4	100	100	100	0.2531
RS5	98.33	98.33	98.33	0.2656
Mean	98.834	99	98.66	0.2620

Table 8: Classification performance of feature reduction based on PCA.

Serial number	Accuracy (%)	Sensitivity (%)	Specificity (%)	10 × running time (s)
1	91.67	83.33	100	0.9970
2	96.74	93.48	100	0.4830
3	96.74	93.48	100	0.4880
4	98.91	100	97.83	0.4950
5	93.48	86.96	100	0.4950
6	96.74	100	93.48	0.5140
7	96.74	100	93.48	0.5120
8	94.57	89.13	100	0.4890
9	97.83	95.65	100	0.4990
10	95.65	93.48	97.83	0.5180
Mean	95.91	93.55	98.26	0.5490

feature components, but also improves the firmness of the feature set by integrating the concept of feature-level fusion based on rough set to exclude the redundant features and data with irregular information, thereby improving the overall pulmonary nodule detection performance.

Table 9: Comparison of the performance of different lung nodule detection methods.

Author	Database	Nodule numbers	Accuracy (%)	FP/s
Santos et al. [11]	L	260	88.4	1.17
Magalhães Barros Netto et al. [12]	L	48	90.65	0.138
Ye et al. [13]	Pr	220	90.2	8.2
Tan et al. [14]	L	172	87.5	4
Cascio et al. [16]	L	148	97	6.1
Our method	Pr	70	99.17	0.47

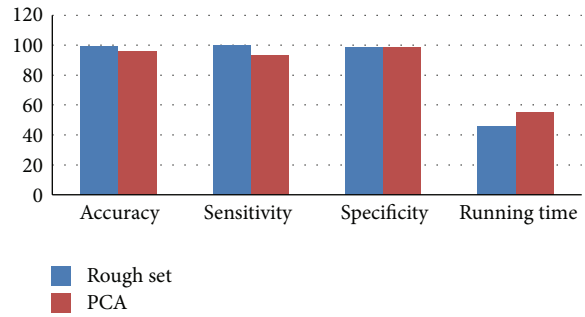


Figure 7: Comparison of two feature-level fusion models.

5. Conclusions

In this paper the research status quo of pulmonary nodule detection methods is analyzed and a pulmonary nodule detection model is proposed based on rough set based feature-level fusion. To address the issues that the feature description is insufficient and the characterization is inaccurate in the process of feature extraction, six new 3D features, in combination with other 2D and 3D features, are

proposed to extract and quantify the feature information of ROI in this model. A rough set based feature-level fusion is employed to reduce the dimensionality of the feature sets since there is redundant information in the extracted high-dimensional features. In addition, a grid optimization model is adopted to optimize the SVM kernel function, which is used as the classifier for detection and recognition of pulmonary nodule. Finally, the pulmonary nodule detection performance of the proposed method is verified with four groups of comparative experiments. The experimental results show that the proposed pulmonary nodule detection method based on rough set based feature-level fusion is effective, with the classification accuracy that can basically meet the requirements of medical imaging for the detection of pulmonary nodules and therefore is of great value for the detection of pulmonary nodules and auxiliary diagnosis of lung cancer.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The work is partially supported by National Natural Science Foundation of China (Grant No. 61561040), Natural Science Foundation of Ningxia (Grant No. NZ16067), and Scientific Research Fund of Ningxia Education Department (Grant No. NGY2016084).

References

[1] S. P. Riaz, M. Lüchtenborg, V. H. Coupland, J. Spicer, M. D. Peake, and H. Møller, "Trends in incidence of small cell lung cancer and all lung cancer," *Lung Cancer*, vol. 75, no. 3, pp. 280–284, 2012.

[2] J. Ferlay, I. Soerjomataram, R. Dikshit et al., "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012," *International Journal of Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.

[3] D. R. Baldwin, "Prediction of risk of lung cancer in populations and in pulmonary nodules: significant progress to drive changes in paradigms," *Lung Cancer*, vol. 89, no. 1, pp. 1–3, 2015.

[4] S. Diciotti, G. Picozzi, M. Falchini, M. Mascalchi, N. Villari, and G. Valli, "3-D segmentation algorithm of small lung nodules in spiral CT images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 7–19, 2008.

[5] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, The Associated Press, 4th edition, 2010.

[6] B. Sahiner, H.-P. Chan, L. M. Hadjiiski et al., "Effect of CAD on radiologists' detection of lung nodules on thoracic CT scans: analysis of an observer performance study by nodule size," *Academic Radiology*, vol. 16, no. 12, pp. 1518–1530, 2009.

[7] Y. Xia, S. Lu, L. Wen, S. Eberl, M. Fulham, and D. D. Feng, "Automated identification of dementia using FDG-PET imaging," *BioMed Research International*, vol. 2014, Article ID 421743, 8 pages, 2014.

[8] Y. Xia, Z. Ji, and Y. Zhang, "Brain MRI image segmentation based on learning local variational Gaussian mixture models," *Neurocomputing*, vol. 204, pp. 189–197, 2016.

[9] M. Xian, Y. Zhang, and H. D. Cheng, "Fully automatic segmentation of breast ultrasound images based on breast characteristics in space and frequency domains," *Pattern Recognition*, vol. 48, no. 2, pp. 485–497, 2015.

[10] M. Xian, Y. Zhang, H.-D. Cheng, F. Xu, and J. Ding, "Neutro-connectedness cut," *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4691–4703, 2016.

[11] A. M. Santos, A. O. de Carvalho Filho, A. C. Silva et al., "Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 27–39, 2014.

[12] S. Magalhães Barros Netto, A. Corrêa Silva, R. Acatauassú Nunes, and M. Gattass, "Automatic segmentation of lung nodules with growing neural gas and support vector machine," *Computers in Biology and Medicine*, vol. 42, no. 11, pp. 1110–1121, 2012.

[13] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, and G. Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic CT images," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 7, pp. 1810–1820, 2009.

[14] M. Tan, R. Deklerck, B. Jansen, M. Bister, and J. Cornelis, "A novel computer-aided lung nodule detection system for CT images," *Medical Physics*, vol. 38, no. 10, pp. 5630–5645, 2011.

[15] Q. Li, F. Li, and K. Doi, "Computerized Detection of Lung Nodules in Thin-Section CT Images by Use of Selective Enhancement Filters and an Automated Rule-Based Classifier," *Academic Radiology*, vol. 15, no. 2, pp. 165–175, 2008.

[16] D. Cascio, R. Magro, F. Fauci, M. Iacomi, and G. Raso, "Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models," *Computers in Biology and Medicine*, vol. 42, no. 11, pp. 1098–1109, 2012.

[17] G. Y. Wang, Y. Y. Yao, and H. Yu, "A survey on rough set theory and applications," *Chinese Journal of Computers*, vol. 32, no. 7, pp. 1229–1246, 2009.

[18] S. Udhaya Kumar and H. Hannah Inbarani, "A novel neighborhood rough set based classification approach for medical diagnosis," *Procedia Computer Science*, vol. 47, pp. 351–359, 2015.

[19] H. H. Inbarani, A. T. Azar, and G. Jothi, "Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 175–185, 2014.

[20] S. K. Shah, M. F. McNitt-Gray, S. R. Rogers et al., "Computer-aided diagnosis of the solitary pulmonary nodule," *Academic Radiology*, vol. 12, no. 5, pp. 570–575, 2005.

[21] A. Hassanien, "Fuzzy rough sets hybrid scheme for breast cancer detection," *Image and Vision Computing*, vol. 25, no. 2, pp. 172–183, 2007.

[22] Z. Guo, Y. Li, Y. Wang, S. Liu, T. Lei, and Y. Fan, "A method of effective text extraction for complex video scene," *Mathematical Problems in Engineering*, vol. 2016, Article ID 2187647, 11 pages, 2016.

[23] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[24] J. Chorowski, J. Wang, and J. M. Zurada, "Review and performance comparison of SVM- and ELM-based classifiers," *Neurocomputing*, vol. 128, pp. 507–516, 2014.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 6727290, 10 pages
http://dx.doi.org/10.1155/2016/6727290

Research Article

Multigrid Nonlocal Gaussian Mixture Model for Segmentation of Brain Tissues in Magnetic Resonance Images

Yunjie Chen,[1] Tianming Zhan,[2] Ji Zhang,[3] and Hongyuan Wang[3]

[1] *School of Math and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China*
[2] *Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science and Technology, Nanjing 210044, China*
[3] *School of Information Science and Engineering, Changzhou University, Changzhou 213164, China*

Correspondence should be addressed to Hongyuan Wang; hywang@cczu.edu.cn

Received 13 June 2016; Accepted 22 July 2016

Academic Editor: Yong Xia

We propose a novel segmentation method based on regional and nonlocal information to overcome the impact of image intensity inhomogeneities and noise in human brain magnetic resonance images. With the consideration of the spatial distribution of different tissues in brain images, our method does not need preestimation or precorrection procedures for intensity inhomogeneities and noise. A nonlocal information based Gaussian mixture model (NGMM) is proposed to reduce the effect of noise. To reduce the effect of intensity inhomogeneity, the multigrid nonlocal Gaussian mixture model (MNGMM) is proposed to segment brain MR images in each nonoverlapping multigrid generated by using a new multigrid generation method. Therefore the proposed model can simultaneously overcome the impact of noise and intensity inhomogeneity and automatically classify 2D and 3D MR data into tissues of white matter, gray matter, and cerebral spinal fluid. To maintain the statistical reliability and spatial continuity of the segmentation, a fusion strategy is adopted to integrate the clustering results from different grid. The experiments on synthetic and clinical brain MR images demonstrate the superior performance of the proposed model comparing with several state-of-the-art algorithms.

1. Introduction

Magnetic resonance imaging (MRI) is a helpful method for diagnosis of brain diseases, such as Alzheimer's disease and schizophrenia. Accurate tissues segmentation, including gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), plays an important role in clinical practice and hence has attracted extensive research attention.

Many methods have been proposed for MR image segmentation in the past several decades. These approaches can be classified in terms of different criteria. For example, edge based methods [1, 2], region based methods [3, 4], and clustering based methods [5–7]. Unfortunately, most segmentation methods are hindered by various imaging artifacts such as noise and intensity inhomogeneities.

Intensity inhomogeneity, also known as bias field, arises from the imperfections of the image acquisition process and changes the absolute intensity for a given tissue class in different locations, which usually makes the intensity distribution within a particular tissue class flatter. Most traditional intensity based methods cannot obtain satisfactory results due to the impact of intensity inhomogeneity.

The observed MRI signal Y is the product of the true signal X generated by the underlying anatomy and a spatially varying field factor B and an additive noise N:

$$Y_i = X_i B_i + N_i \quad \forall i \in \{1, 2, \ldots, M\}, \tag{1}$$

where Y_i, X_i, B_i, and N_i are the observed intensity, true intensity, bias field, and noise at the ith voxel, respectively. M is the total number of pixels in the MR image. Many techniques [6, 8, 9] often ignore the noise and take the logarithmic transform on both sides of (1):

$$\log(Y) = \log(X) + \log(B). \tag{2}$$

Many methods have been proposed to correct or estimate intensity inhomogeneities. Collewet et al. proposed a method based on measuring the coil sensitivity functions [8]. Based on the observation that the bias field is smooth, another group of methods overcome the impact of intensity inhomogeneity without estimating the bias field [9, 10]. However, most of them may lose edge information [11].

In this paper, we first propose an improved nonlocal Gaussian mixture model by introducing the nonlocal information into GMM model to reduce the effect of the noise. Then, a new multigrid generation method is presented, and we simplify the NGMM into a local version to eliminate the effect of the bias field and the intensity variation of intratissues. Finally, we propose a fusion strategy to integrate the results from different local regions. The experiments on both synthetic and clinical brain MR images show that our method can obtain more accurate results.

2. Nonlocal Gaussian Mixture Model

Gaussian mixture model (GMM) has been widely used in many applications due to its excellent approximation properties. Suppose a MR image has a mixture of K components, the mixture density function of pixel i can be written as

$$p(Y_i \mid \theta) = \sum_{k=1}^{K} \pi_k p(Y_i \mid \theta_k), \tag{3}$$

where π_k is the mixture weights and $\theta = (\theta_1, \theta_2, \ldots, \theta_K)$ is the parameter vector. $p(Y_i \mid \theta_K)$ is a standard Gaussian distribution of the kth component and $\theta_k = (\mu_k, \sigma_k)$ contains the parameters of the Gaussian distribution. μ_k and σ_k are the mean and variance, respectively. Then the entire distribution can be written as

$$p(Y \mid \theta) = \prod_{i=1}^{M} p(Y_i \mid \theta) = L(\theta \mid Y). \tag{4}$$

The problem is how to find the best parameters θ:

$$\theta^* = \arg\max L(\theta \mid Y). \tag{5}$$

Equation (5) can be calculated by using the expectation-maximization (EM) method [7]. In the E-Step, the algorithm calculates the expected value of the kth weight:

$$\begin{aligned} p(k \mid Y_i, \theta) &= \frac{\pi_k p_k(Y_i \mid \theta)}{\sum_{j=1}^{K} \pi_j p_j(Y_i \mid \theta)}, \\ \alpha_k &= \sum_{i=1}^{N} p(k \mid Y_i, \theta). \end{aligned} \tag{6}$$

In the M-Step, the parameters of the kth Gaussian distribution can be calculated:

$$\begin{aligned} \pi_k &= \frac{1}{N}\alpha_k, \\ \mu_k &= \frac{\sum_{i=1}^{N} Y_i p(k \mid Y_i, \theta)}{\sum_{i=1}^{N} p(k \mid Y_i, \theta)}, \\ \sigma_k &= \frac{\sum_{i=1}^{N} (Y_i - \mu_k)^2 p(k \mid Y_i, \theta)}{\sum_{i=1}^{N} p(k \mid Y_i, \theta)}. \end{aligned} \tag{7}$$

Based on the initialization, α and θ are calculated iteratively until the stop criteria are reached. Finally, the pixel i can be classified into the kth class when $\{k \mid p(k \mid Y_i, \theta) > p(j \mid Y_i, \theta),\ j \in \{1, 2, \ldots, K\}\ j \neq k\}$. From (3), we can find that the GMM only considers the intensity distribution, which makes the method sensitive to the intensity inhomogeneity and noise.

In order to reduce the effect of the intensity inhomogeneity, Wells et al. [6] proposed a method to simultaneously estimate the bias field and segment the image into different classes. However, the method only addressed the bias field without analyzing the inhomogeneities in inner tissues. In order to ensure the smoothness of the bias field, a low-pass filter is utilized to convolve the bias field, which makes the estimated bias field inaccurate. Furthermore, the method is sensitive to the noise.

In order to reduce the effect of the noise and preserve more detailed information, we improve the Gaussian mixture model by using nonlocal information. The nonlocal information has been widely used for denoising purposes [12, 13]. Following the idea of nonlocal means method [12], we use the nonlocal information to adapt $p(k \mid Y, \theta)$ which can be defined as

$$\mathrm{NL}(p(k \mid Y_i, \theta)) = \sum_{j=1}^{N} W(i, j)\, p(k \mid Y_j, \theta), \quad k = 1, \ldots, K, \tag{8}$$

where $W(i, j)$ is the weight function based on the similarity between the neighbor patch of each neighbor pixel to that of center pixel and satisfies the conditions $0 \leq W(i, j) \leq 1$ and $\sum_{j=1}^{N} W(i, j) = 1$. For pixel i and its neighbor pixel j, the weight function is defined as

$$W(i, j) = \frac{e^{-\|\Delta_i - \Delta_j\|_{2,r}^2 / h^2}}{\sum_{j=1}^{N} e^{-\|\Delta_i - \Delta_j\|_{2,r}^2 / h^2}}, \tag{9}$$

where Δ_i and Δ_j are the neighbor patches around pixels i and j with width $2 \times p + 1$. Broadly speaking, if the neighbor patches of two pixels i and j are similar, it is more probable that these pixels belong to the same tissue, and the corresponding values of weight function would be higher. Conversely, if these two pixels are quite different, the value of the weight function should be small. h acts as a filtering parameter to control the decay of the exponential function. $\|\cdot\|$ is the Euclidean distance and r is the standard

deviation of the Gaussian kernel. Due to the fast decay of the exponential kernel, large distances between estimated patches lead to nearly zero weights. Essentially, the weight function aims to take advantage of the redundancy present in natural structures. Therefore, by using nonlocal information, the nonlocal Gaussian mixture model can reduce the effect of noise and preserve details of the edges.

3. Multigrid Nonlocal Gaussian Mixture Method (MGMM)

Without considering the intensity inhomogeneity, the nonlocal Gaussian mixture model can only reduce the effect of noise but cannot obtain satisfactory results for the image containing severe intensity inhomogeneities. Zhu and Jiang [14] proposed a multicontext fuzzy clustering method (MCFC) to reduce the effect of intensity inhomogeneity by using fuzzy clustering method on each nonoverlapping regions and a fusion strategy to integrate the clustering outcomes form different regions. However, this method is sensitive to noise and it only uses traditional multigrid generation method, which makes the method inaccurate. Following the idea of MCFC, we improve the GMM by using the nonoverlapping multigrid. This idea is based on the following four assumptions:

(1) The brain image has been skull-stripped. In this paper, we use the cut based method [15].

(2) Bias field B is smooth and slowly varying.

(3) Within each grid, the number of clusters must equal K and there are considerable numbers of pixels in each tissue class.

(4) Within a grid, all pixels of the same tissue have similar true intensities.

The brain image without skull only has cerebrospinal fluid, gray matter, and white matter. Then, we set $K = 3$ with assumption (3). The bias field is smooth and slowly varying, which makes it probable that the bias field values in small local region be regarded as constant. Then the multigrid segmentation method is less sensitive to the bias field. However, each pixel is processed only in its single local grid, which makes the method very sensitive to the size of the grid and unable to preserve the statistical reliability and spatial continuity of segmentation results. This can be illustrated in Figure 1.

Figure 1 shows the segmentation results of GMM and MGMM applied on a synthetic 3 T MR image, which were created by using the MRI simulator (Brain Imaging Center at the Montreal Neurological Institute, McGill University, Montreal). There are many advantages for using these synthetic images rather than real image data volumes for validating segmentation methods. The simulator can provide full three-dimensional data volumes which have been simulated using three sequences (T1-, T2-, and PD-weighted) and a variety of slice thicknesses, noise levels, and intensity inhomogeneity levels, providing the ground truth of the image data. In this paper, the parameters of the simulated data sets are as follows: Phantom: normal; slice thickness: 1 mm; scan technique: SFLASH; TR = 18 msec; flip angle = 30 degrees; TE = 10 msec. The dimension of the image data sets is 181 × 217 × 181. The parameters of Figure 1(a) are as follows: noise level 0% and INU (RF) level 100%. Figure 1(a) shows the initial image, which is uniformly divided into 4 × 4 nonoverlapping grids. Figure 1(b) shows the results of GMM. Due to the effect of the bias field, some voxels of WM and CSF have been misclassified into GM. Figures 1(c)–1(e) are segmentation results of MGMM when the image is divided into 3 × 3, 4 × 4, and 5 × 5 grids, respectively.

It can be seen from the grids (2, 2) and (3, 2) in Figure 1(c) or from the grids (3, 2) and (4, 2) in Figure 1(d) that the variation of intensity distributions of neighbor patches would, more or less, lead to discontinuities across the grids' boundaries. Furthermore, when the bias field is severe, grids with large size can satisfy assumption (3); however, this would make the method sensitive to the bias field and does not satisfy assumption (4). It is hard to hold assumption (3) in some grids when the size of the grids is small. For example, there are no WM pixels in the grid (3, 1) of Figure 1(e), which makes some pixels of GM misclassified into WM. It can also be found in grid (3, 2) of Figure 1(e) that tissues of WM and GM predominate the grid, which yields a deviant intensity distribution far from that typical of the brain. The GMM misclassifies some GM pixels with relatively low intensity into the class of CSF.

3.1. New Multigrid Generation Method. In this paper we present a new method to generate the multigrid. Firstly, the boundary of the brain needs to be found, because there are a large number of pixels belonging to the background in brain MR images, which usually affect the accuracy of segmentation methods. Secondly, the brain region is divided into $N \times N$ small nonoverlapping grids. The generated nonoverlapping grids may not satisfy assumption (3). Figure 2 shows the generated multigrid on brain MR images. In this paper, we set $N = 6$. It can be seen from Figure 2(a) that the grids (1, 1) and (6, 3) only contain some CSF pixels and the grids (1, 6) and (6, 6) have no brain tissues. Furthermore, the grid (6, 1) has no pixels of the WM. The NGMM cannot obtain accurate results based on these grids. In order to deal with this problem, the small grids need to be combined. The combine process includes 6 steps as follows.

Step 1. Define a matrix L with the same size of the multigrid and a variable $N_{\text{number}} = 1$ to count the serial number of the final patches.

Step 2. Classify every grid into K classes.

Step 3. Find the worst grid (x, y) and set $L(x, y) = N_{\text{number}}$.

Step 4. Find the preferred grid (x', y') to be combined into the patch $\{(x, y) \mid L(x, y) = N_{\text{number}}\}$ and set $L(x', y') = N_{\text{number}}$. Repeat this process N_{search} times.

Step 5. Set $N_{\text{number}} = N_{\text{number}} + 1$ and go to Step 3 until all grids have been labeled except those nonbrain grids.

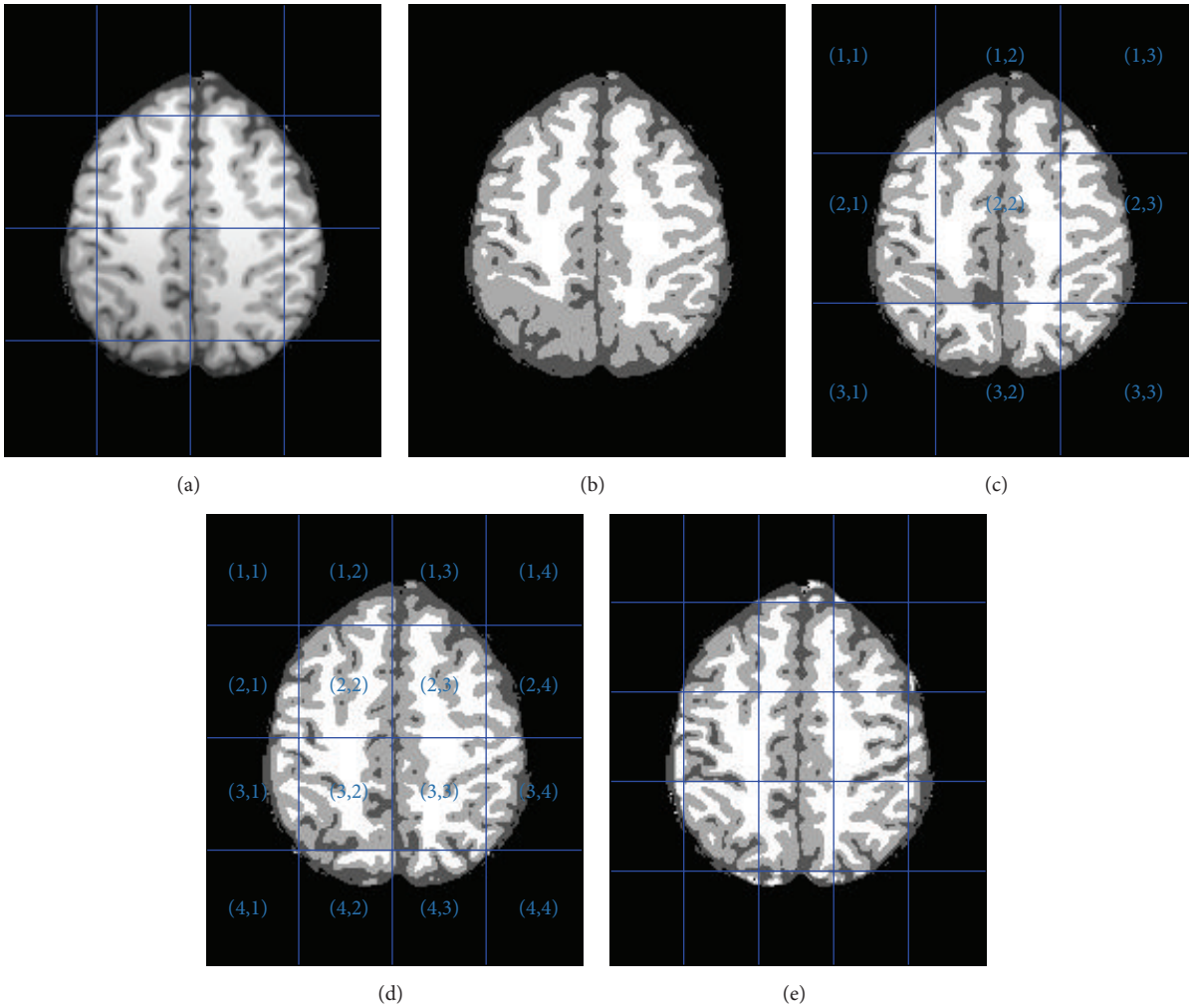


Figure 1: Brain image segmentation: (a) original T1-weighted image; (b) GMM segmentation; (c)–(e) MGMM segmentation.

Step 6. Combine single patches into the preferred patches nearby.

In Step 1, matrix L is defined to sign the multigrid and initialized to zero for $\forall x, y$. The variable N_{number} is used to count the serial numbers of the final patches and initialized as 1. In Step 2, every grid is classified into K classes by using Fuzzy Clustering Means (FCM) method, which can classify grids efficiently. We assume that assumption (3) holds in all grids and all patches are classified into 3 classes except those background grids. The histograms of the grids $(6, 2)$ and $(3, 3)$ are shown in Figures 2(b) and 2(c), respectively. There are only GM and some CSF pixels in the grid $(6, 2)$, which make the distribution of the pixels' intensities more compact than that of $(3, 3)$.

In Step 3, we first calculate an inner distance for every grid. The inner distance is defined as $\min(|C_{\text{CSF}} - C_{\text{GM}}|, |C_{\text{WM}} - C_{\text{GM}}|)$, where C_{CSF}, C_{GM}, and C_{WM} are the intensity centers of CSF, GM, and WM, respectively. The worst grid (x, y) with the smallest inner distance can be found easily and $L(x, y)$ is set as N_{number}. In Step 4, the grids $\{(\hat{x}, \hat{y}) \mid L(\hat{x}, \hat{y}) = 0\}$, which are next to the patch $\{(x, y) \mid L(x, y) = N_{\text{number}}\}$ need to be found. Then, every grid $(\hat{x}, \hat{y})$ is combined into the patch $\{(x, y) \mid L(x, y) = N_{\text{number}}\}$ independently and the inner distance of the corresponding combined patch is calculated. The grid with the largest inner distance is regarded as the preferred grid to be combined into patch $\{(x, y) \mid L(x, y) = N_{\text{number}}\}$. This process should be repeated for N_{search} times, where N_{search} satisfies $N_{\text{search}} * N_{\text{grid}} \leq \alpha N_{\text{brain}}$. N_{grid} is the size of the grids. N_{brain} is the amount of tissue pixels in the input image. α is the control parameter. It is hard to hold assumption (4) with large α. In this paper, we set $\alpha = 0.1$, then $N_{\text{search}} = 3$. Furthermore the width and height of every patch are no more than 1/2 of the brain region's width and height.

After Step 5, all grids have been labeled; however, there may be some grids that have not been combined with any neighbor patch. These grids are set as single grids. It is hard to hold assumption (3) in these grids. In order to deal with this problem, we use similar strategy shown in Step 4 to combine the single grids into best neighbor patch. Firstly, we find one of the single grids. Secondly, we analyze the neighbor patches,

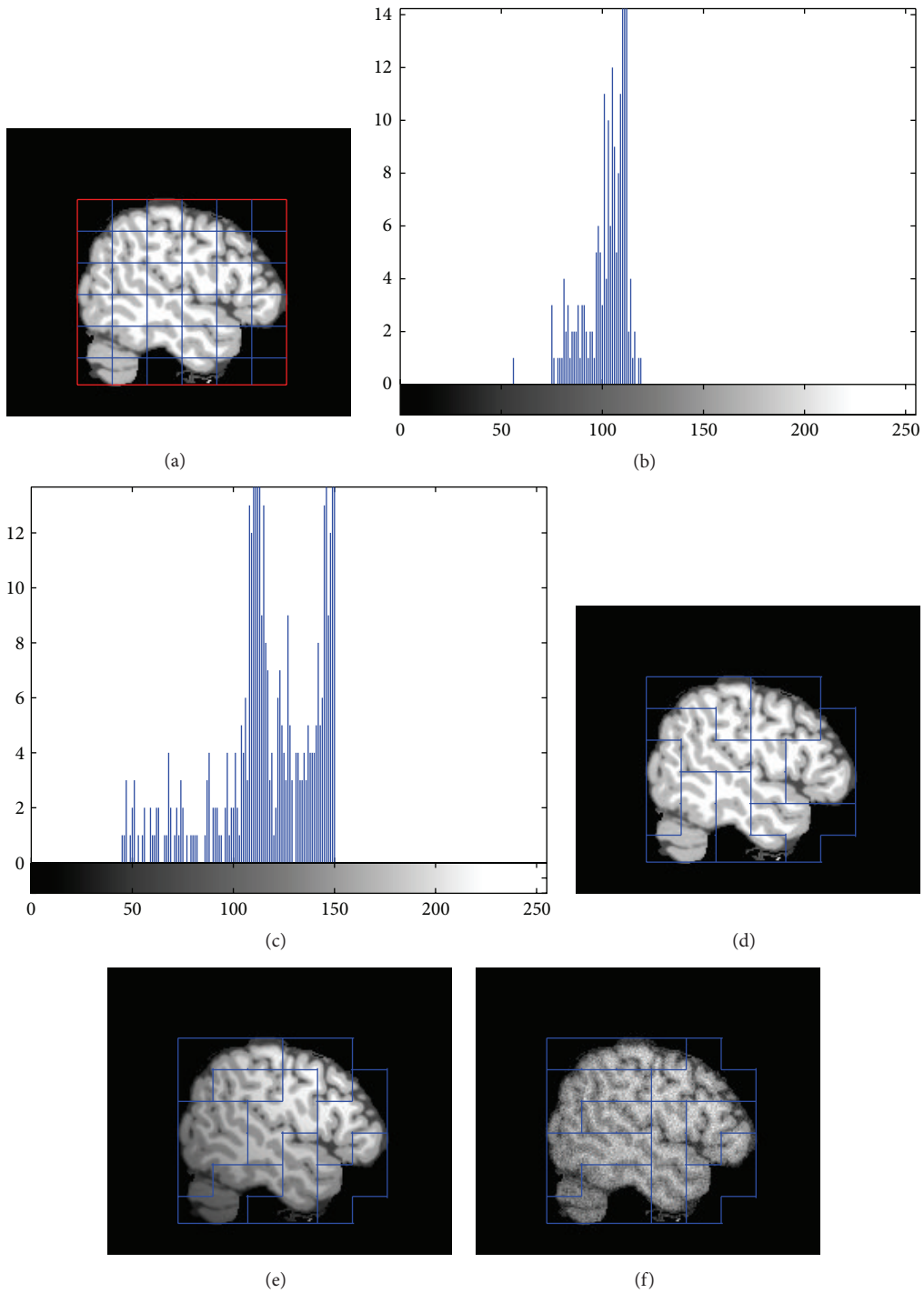


Figure 2: Generation of the grid: (a) initial results; (b) histogram of patch (6, 2); (c) histogram of patch (3, 3); (d)–(f) results of multigrid.

which are next to the single grid, to find the patch with largest inner distance as the preferred patch. The single grid is combined into the preferred patch. The process is repeated until all single grids have been combined into their preferred patch.

Figures 2(d)–2(f) show the results of the new multigrid generation method on three synthetic brain MR images, with the following parameters: noise levels are 0%, 0%, and 5%, respectively, and INU (RF) levels are 0%, 100%, and 40%, respectively. It can be seen that our method can generate good multigrid even when the image has strong noise or bias field.

3.2. Multigrid Nonlocal Gaussian Mixture Model (MNGMM). Assumption (3) can be retained by using the improved multigrid. However, due to the effect of the bias field, the variation of intensity distributions of neighbor patches

would, more or less, lead to boundaries discontinuous across edges of some grids. Despite this problem, the assumptions for simplifying the image model are still correct in most cases, due to the complicated and convoluted structure of human brain as mentioned above. In other words, most grids would yield good judgments based on the expected values $p(k \mid Y_i, \theta)$. A very natural idea is that the adverse impact of some grids could be overcome by all the good clustering results from other neighbor patches. Such a consideration leads to the development of a novel method called the multigrid nonlocal Gaussian mixture model (MNGMM) that can take advantage of the adaptation of local clustering but also keep the classifications spatially continuous and statistically reliable.

MNGMM includes two basic parts: clustering and information fusion. The rationale of MNGMM can be described as follows. First, all grids need to be classified by using NGMM to obtain the corresponding expected values $p(k \mid Y_i, \theta)$. For every pixel i, it is easy to find the corresponding grid and obtain its expected values $p(k \mid Y_i, \theta)$ in the grid. The distribution information of the neighbor grids Q_{ij}, $j = 1, \ldots, N_i$ next to i is used to calculate $p(k \mid Y_i, \theta_j)$, where θ_j is the distribution parameter of Q_{ij} and N_i is the total number of the neighbor grids. In the information fusion stage, all these expected values are integrated with a strategy F to adapt the current expected values. Consequently, for each pixel P, MNGMM can be summarized as follows.

Step 1. One has

$$Q_{ij} = \left\{ p\left(k \mid Y_i, \theta_j\right) \right\}, \quad k = 1, \ldots, K; \; j = 1, \ldots, N_i. \tag{10}$$

Step 2. One has

$$\begin{aligned} &p\left(k \mid Y_i, \theta\right) \\ &\quad = F\left(p\left(k \mid Y, \theta\right), p\left(k \mid Y, \theta_1\right), \ldots, p\left(k \mid Y, \theta_{N_i}\right)\right). \end{aligned} \tag{11}$$

Here, we name $p(k \mid Y_i, \theta)$ as $p_{k,0}$ and $p(k \mid Y_i, \theta_j)$, $j = 1, \ldots, N_i$ as $p_{k,j}$; the strategy is

$$F\left(p_{k,0}, p_{k,1}, \ldots, p_{k,N_i}\right) = \sum_{j=0}^{N_i} \frac{p\left(Y_i \mid \theta_{k,j}\right)}{\sum_{l=0}^{N_i} p\left(Y_i \mid \theta_{k,l}\right)} p_{k,j}. \tag{12}$$

In our method, only pixels on edge regions of the grid may be misclassified into other class. The intensities of the pixels are continuous across the edges of the neighbor grids, which makes it probable that the expected values of the misclassified pixels be weighted averaged by using the information of the neighbor grids. As a result, the weighted averaging values of different grids make the final result more reliable. Moreover, the weighted averaging operation can also preserve spatial continuity of the membership distributions of each tissue class. Finally, maximum expected values $p(k \mid Y_i, \theta)$ principle is used to obtain the segmentation results from the weighted averaging values.

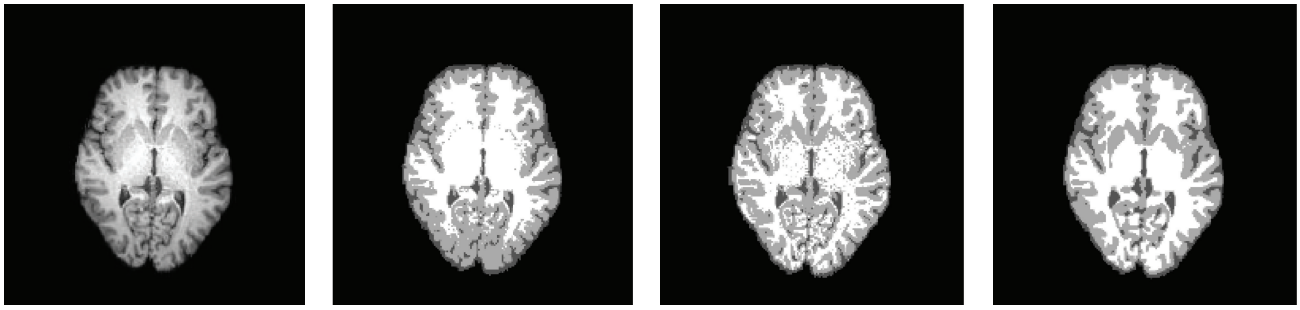
4. Results

In order to show the robustness of our method, we compared our method with GMM and Wells method on a clinical 3 T MR image, which has severe intensity inhomogeneity and noise. The acquisition parameters of the real data are as follows: slice thickness 1 mm, echo delay time (TE) 35 msec, repetition time (TR) 450 msec, and flip angle 90 deg. The size of the data is 256 × 256 × 170. From the image, we can find that most intensities of putamen are higher than those of the cortex, which also belong to the GM. GMM misclassifies putamen into WM since some intensities of putamen are closer to WM than to GM, which can be shown in Figure 3(b). The result of the Wells method [6] is shown in Figure 3(c). It can be seen from the result that the method can reduce the effect of the bias field; however, with the effect of the inhomogeneities in inner tissues, some pixels of putamen have been misclassified into WM and the method is sensitive to the noise. In contrast, MNGMM can yield satisfactory results, which are more compatible with human visual perception.

In order to quantitatively show the performances of the proposed model, we compared our method with GMM, Wells method, MCFC, and MNGMM on the synthetic data from MRI simulator with the following parameters: noise level 5% and INU level 80%. The results are shown in Figure 4. The left column shows the initial image. The end column shows the ground truth. The second column to the fifth column show the segment results of GMM, Wells, MCFC, and MNGMM, respectively. Due to the effect of the intensity inhomogeneity, many pixels of WM have been misclassified into GM in the result of GMM. Wells and MCFC can reduce the effect of the intensity inhomogeneities; however, they are affected by the noise. By using the nonlocal information and multigrid information, MNGMM can yield much more complete and continuous results, which are very similar to the ground truth.

We use Jaccard similarity value (JS) [6] to quantitatively evaluate the performance of a classification method. The value of JS ranges from 0 to 1, with a higher value representing a more accurate segmentation result. The statistical results (means and standard deviations of JS values for WM, GM, and CSF) are listed in Table 1. The results demonstrate that our method produces the most accurate results and has the best ability and robustness to the noisy images (with lower standard deviations of JS values and higher mean of JS values when the noise increases), especially in the area with abundant textures (with higher JS values for CSF tissue). We also apply the above five methods to the segmentation of 40 whole synthetic MR image data sets, in which the level of intensity inhomogeneity ranges from 20% to 100%. The segmentation accuracy is measured in terms of the average JS of WM, GM, and CSF delineation and is shown in Figure 5. Both visual and quantitative comparisons show that our method is more robust to the intensity inhomogeneity and can obtain more accurate results.

Figure 6 shows the segmentation results on a real brain MR data from the Internet brain segmentation repository (IBSR at http://www.cma.mgh.harvard.edu/ibsr/) with the name 12_3 (39th image). The intensity distribution of the basal ganglia is midway between the assumed distributions of GM and WM and the basal ganglia have low contrast. From the results, we can find that our method can obtain accurate



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

Figure 3: Segmentation results for a clinical 3 T MR image: (a) original image; (b) segmentation result of GMM; (c) segmentation result of Wells; (d) segmentation result of MNGMM.

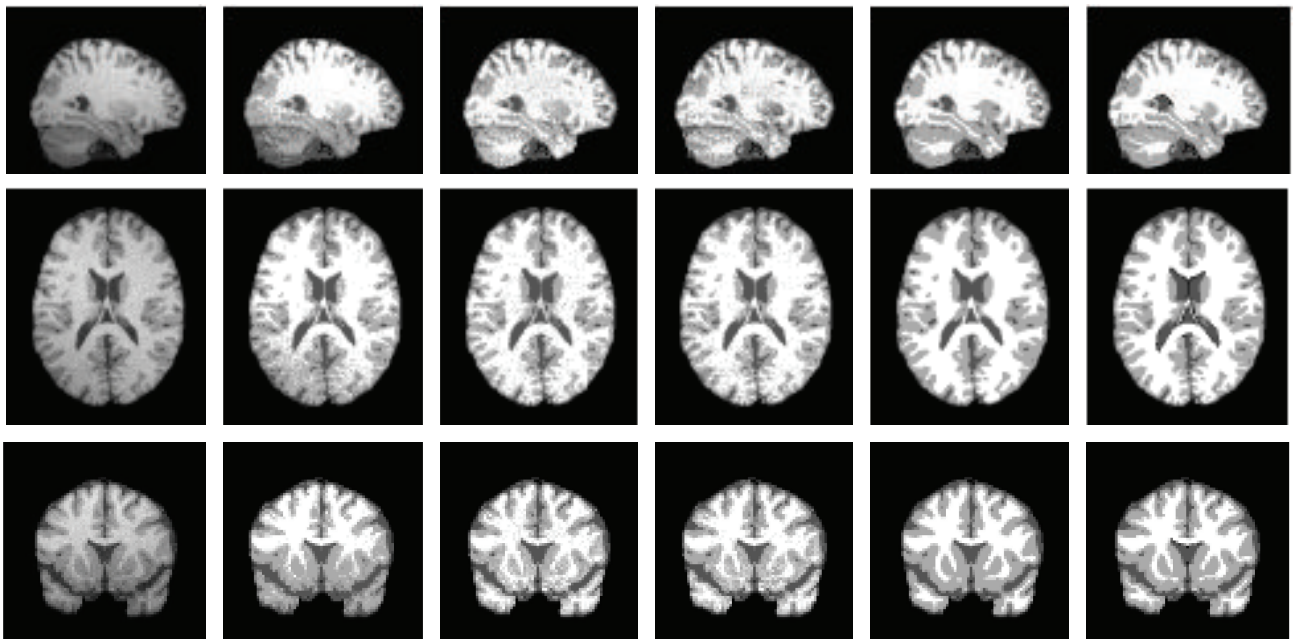


Figure 4: Segmentation results of synthetic data with the parameter noise level 5% and INU level 80%. The left column shows the initial image. The second column to the fifth column show the results of GMM, Wells, MCFC, and MNGMM, respectively. The right column is the ground truth.

Table 1: The average JS values (mean ± standard deviation) of GM, WM, and CSF segmentation obtained by applying four algorithms to T1-weighted brain MR images with increasing level of noise.

Algorithm	Tissues	3%	5%	7%	9%
GMM	WM	0.8512 ± 0.051	0.7532 ± 0.047	0.6574 ± 0.064	0.6135 ± 0.067
	GM	0.8478 ± 0.059	0.7231 ± 0.065	0.6326 ± 0.046	0.6012 ± 0.056
	CSF	0.8547 ± 0.048	0.7447 ± 0.054	0.6236 ± 0.043	0.6103 ± 0.055
Wells	WM	0.9201 ± 0.071	0.7932 ± 0.069	0.7154 ± 0.061	0.6843 ± 0.062
	GM	0.9102 ± 0.056	0.7863 ± 0.048	0.7001 ± 0.044	0.6632 ± 0.051
	CSF	0.8842 ± 0.052	0.7731 ± 0.059	0.7011 ± 0.061	0.6691 ± 0.058
MCFC	WM	0.9382 ± 0.051	0.8131 ± 0.053	0.7320 ± 0.046	0.6914 ± 0.038
	GM	0.9262 ± 0.048	0.7914 ± 0.046	0.7250 ± 0.047	0.6853 ± 0.032
	CSF	0.8937 ± 0.046	0.7724 ± 0.055	0.7123 ± 0.059	0.6749 ± 0.054
MNGMM	WM	0.9328 ± 0.019	0.9257 ± 0.032	0.9231 ± 0.031	0.9105 ± 0.032
	GM	0.9331 ± 0.017	0.9216 ± 0.038	0.9187 ± 0.034	0.9073 ± 0.038
	CSF	0.9293 ± 0.022	0.9211 ± 0.039	0.9127 ± 0.028	0.9005 ± 0.041

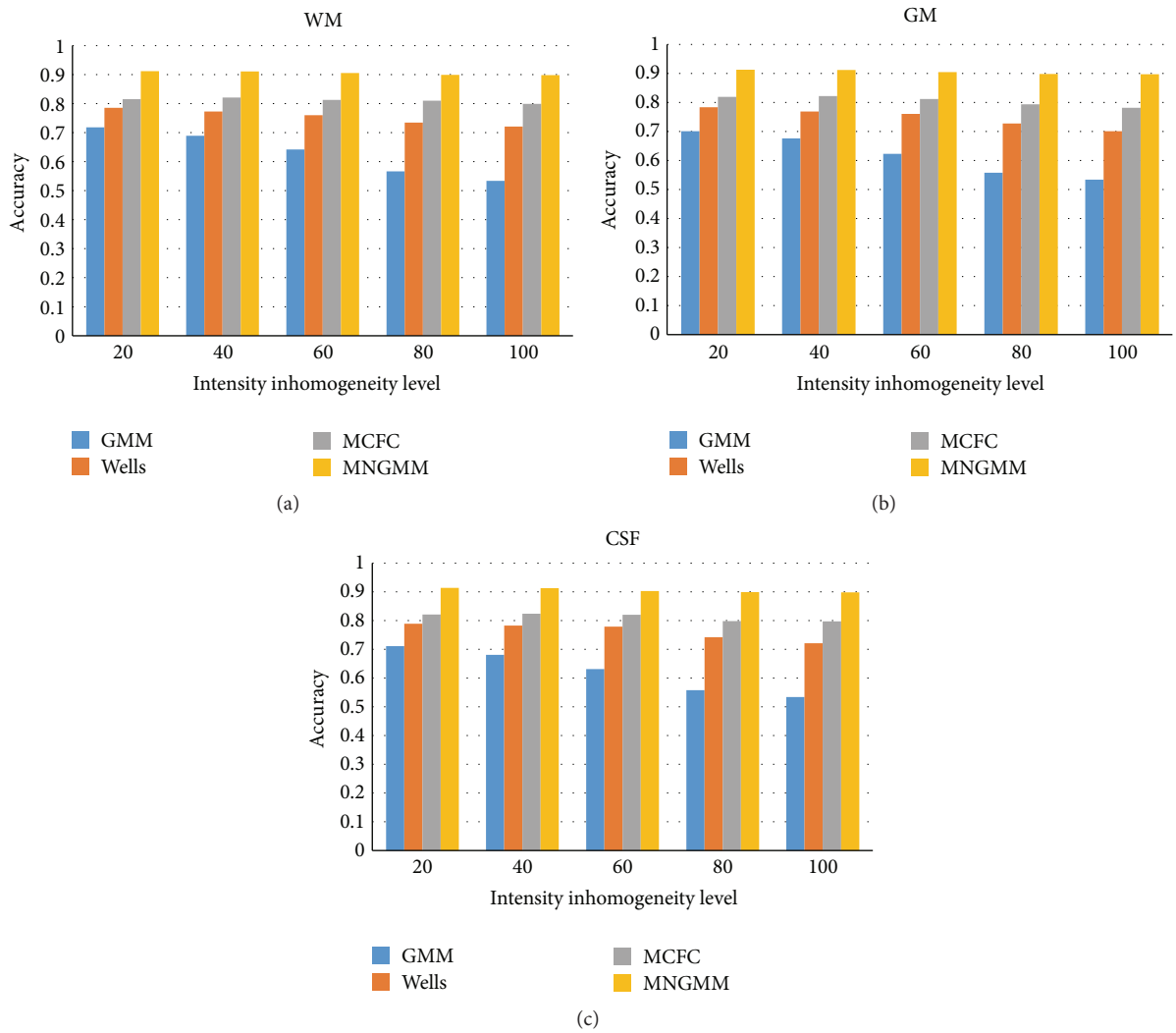


Figure 5: Average JS values of the segmentation results of WM (a), GM (b), and CSF (c) obtained by applying five segmentation methods to simulated brain MR images with increasing levels of intensity inhomogeneity.

Table 2: The average JS values for the real MR images segmentation (%).

	GMM	Wells	MCFC	MNGMM
JS of WM	61.32	72.13	80.87	90.31
JS of GM	59.23	71.35	78.31	88.37
JS of CSF	57.87	78.32	76.45	89.52

result. In order to quantitatively evaluate the benefits, we segmented 20 standard sets of real brain MR data from IBSR by using GMM, Wells method, MCFC, and MNGMM. The average quantitative results of GM, WM, and CSF are listed in Table 2. It can be seen that our method is more accurate than others.

5. Discussion

In our work, the number of grids N is set as a constant. The choices of the divided number should be based on the size of the brain region and intensity inhomogeneity level. A smaller divided number will make the brain region only divided into few patches, which makes the proposed method sensitive to intensity inhomogeneity. A larger divided number will make the method less efficient. We have studied the relationship between size of patches and segmentation accuracies. Figure 7 shows the accuracies of the segmentation to simulated images with different parameters. In fact, the larger the patch, the more the data to be clustered; the greater the similarity between the intensity distribution of the patch and that of the input image is, the more reliable the clustering results are. However, assumptions 2 and 4 require smaller patches. The left column of Figure 6 shows the accuracies of the results using different N when generating the multigrid on the input data with the following parameters: noise level, 0%, and intensity inhomogeneity level, 0%, 10%, 20%, 40%, 80%, and 100%, respectively. It can be seen that with the increase of intensity inhomogeneities the accuracies decrease when N is small, which also means that the patch is bigger.

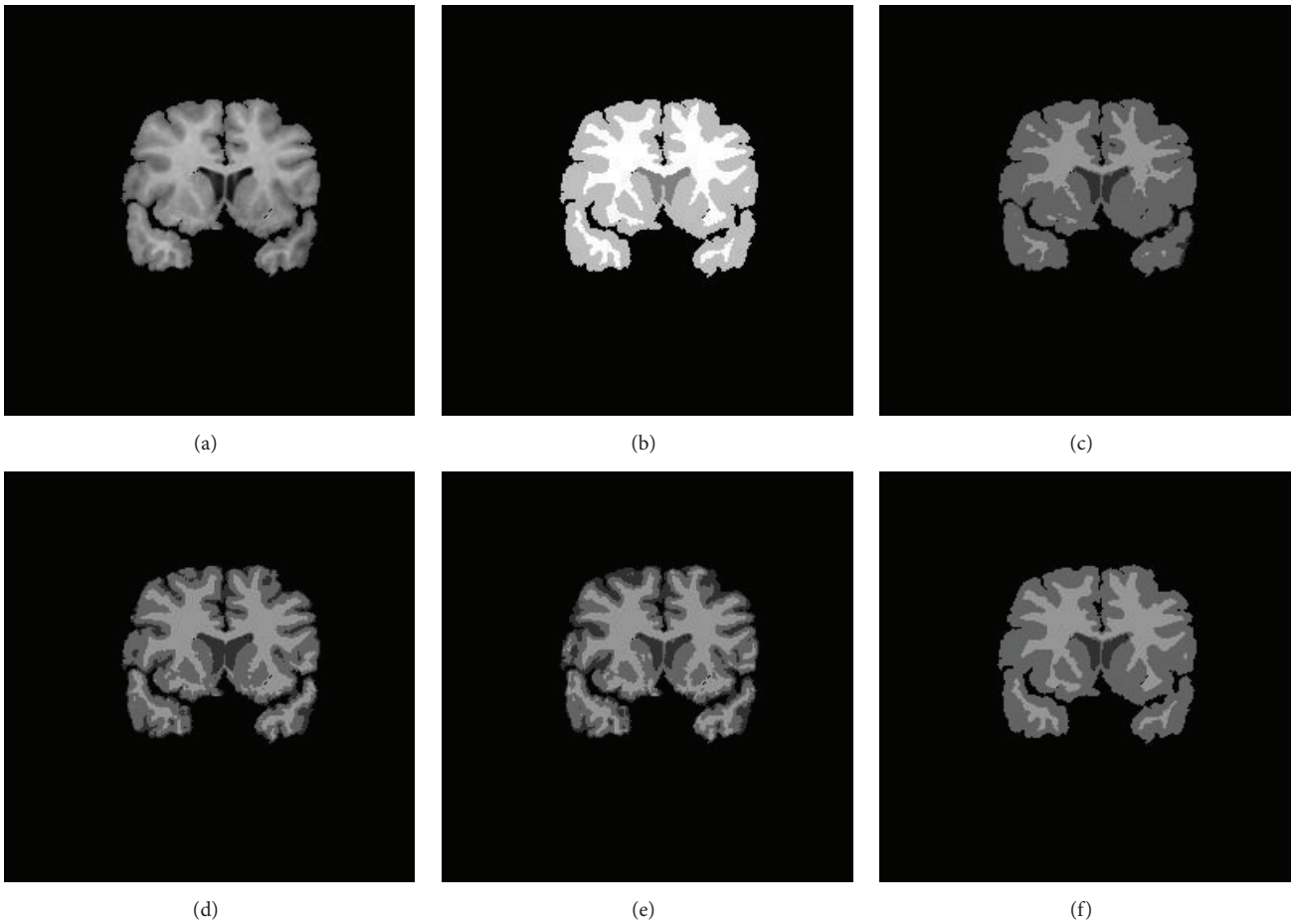


Figure 6: Segmentation results of real data from IBSR with the name 12_3 (39th image). (a) Initial image, (b) ground truth, and (c–f) the segmentation results of GMM, Wells, MCFC, and MNGMM.

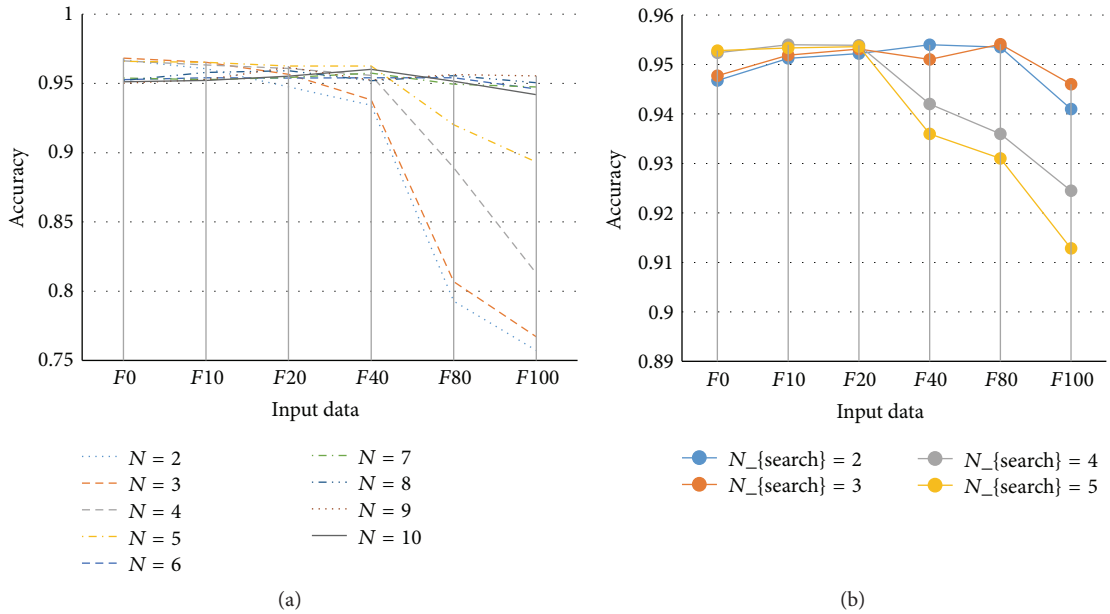


Figure 7: Accuracies of the segmentation with different N and N_{search}.

At this time, assumptions (2) and (4) cannot be satisfied and the accuracies decrease very quickly. When $N = 6, 7, 8, 9$ the results have similar accuracies. We also analyzed the influence of N_{search}. The right column of Figure 7 shows the accuracies with different N_{search} when $N = 6$. It can be seen from the results that the accuracy is decreasing with the increase of N_{search} when the intensity inhomogeneities increase. This is because the bigger N_{search} leads to bigger patches, which means that assumptions (2) and (4) cannot hold.

6. Conclusions

In this paper, we have presented a theoretically simple approach to automatically segment 2D or 3D human brain MRI data. To reduce the effect of intensity inhomogeneity, a multigrid Gaussian mixture model has been proposed. In order to reduce the effect of noise, we improve the Gaussian mixture by using the nonlocal information, which can preserve more detailed information. Experimental results have proved that our method outperforms other segmentation methods when segmenting images with severe intensity inhomogeneities and noise.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China (61572257 and 61502206); National Research Foundation of Korea under Grant NRF-2013K2 A2S2000777; Natural Science Foundation of Jiangsu Province under Grant nos. BY2014007-04, BK20150523, and BK20150923; and Open Project of Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science and Technology (KDXS1404).

References

[1] S. Zhu and R. Gao, "A novel generalized gradient vector flow snake model using minimal surface and component-normalized method for medical image segmentation," *Biomedical Signal Processing and Control*, vol. 26, pp. 1–10, 2016.

[2] C. Li, J. C. Gore, and C. Davatzikos, "Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation," *Magnetic Resonance Imaging*, vol. 32, no. 7, pp. 913–923, 2014.

[3] H. Lu, Z. Cao, Y. Xiao, Y. Li, and Y. Zhu, "Region-based colour modelling for joint crop and maize tassel segmentation," *Biosystems Engineering*, vol. 147, pp. 139–150, 2016.

[4] Y. Zheng, B. Jeon, D. Xu, Q. M. J. Wu, and H. Zhang, "Image segmentation by generalized hierarchical fuzzy C-means algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 28, no. 2, pp. 961–973, 2015.

[5] N. Kim, M. Heo, R. Fleysher, C. A. Branch, and M. L. Lipton, "Two step Gaussian mixture model approach to characterize white matter disease based on distributional changes," *Journal of Neuroscience Methods*, vol. 270, no. 1, pp. 156–164, 2016.

[6] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 429–442, 1996.

[7] H. Zhang, Q. M. J. Wu, and T. M. Nguyen, "Incorporating mean template into finite mixture model for image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 2, pp. 328–335, 2013.

[8] G. Collewet, A. Davenel, C. Toussaint, and S. Akoka, "Correction of intensity nonuniformity in spin-echo T1-weighted images," *Magnetic Resonance Imaging*, vol. 20, no. 4, pp. 365–373, 2002.

[9] Z. Ji, J. Liu, G. Cao, Q. Sun, and Q. Chen, "Robust spatially constrained fuzzy c-means algorithm for brain MR image segmentation," *Pattern Recognition*, vol. 47, no. 7, pp. 2454–2466, 2014.

[10] E. A. Vokurka, N. A. Thacker, and A. Jackson, "A fast model independent method for automatic correction of intensity nonuniformity in MRI data," *Journal of Magnetic Resonance Imaging*, vol. 10, no. 4, pp. 550–562, 1999.

[11] U. Vovk, F. Pernuš, and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI," *IEEE Transactions on Medical Imaging*, vol. 26, no. 3, pp. 405–421, 2007.

[12] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 425–441, 2008.

[13] Q. Chen, Q.-S. Sun, and D.-S. Xia, "Homogeneity similarity based image denoising," *Pattern Recognition*, vol. 43, no. 12, pp. 4089–4100, 2010.

[14] C. Zhu and T. Jiang, "Multicontext fuzzy clustering for separation of brain tissues in magnetic resonance images," *NeuroImage*, vol. 18, no. 3, pp. 685–696, 2003.

[15] S. A. Sadananthan, W. Zheng, M. W. L. Chee, and V. Zagorodnov, "Skull stripping using graph cuts," *NeuroImage*, vol. 49, no. 1, pp. 225–239, 2010.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 1480423, 13 pages
http://dx.doi.org/10.1155/2016/1480423

Research Article

Many Is Better Than One: An Integration of Multiple Simple Strategies for Accurate Lung Segmentation in CT Images

Zhenghao Shi,[1] Jiejue Ma,[1] Minghua Zhao,[1] Yonghong Liu,[2] Yaning Feng,[1] Ming Zhang,[3] Lifeng He,[4] and Kenji Suzuki[5]

[1] *School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China*
[2] *Xianyang Hospital, Yan'an University, Xianyang 712000, China*
[3] *First Affiliated Hospital of School of Medicine, Xian Jiaotong University, Xian 710061, China*
[4] *School of Information Science and Technology, Aichi Prefectural University, Nagakute, Aichi 480-1198, Japan*
[5] *Medical Imaging Research Center, Illinois Institute of Technology, Chicago, IL 60616-3793, USA*

Correspondence should be addressed to Zhenghao Shi; ylshi@xaut.edu.cn

Received 28 April 2016; Accepted 19 July 2016

Academic Editor: Weidong Cai

Accurate lung segmentation is an essential step in developing a computer-aided lung disease diagnosis system. However, because of the high variability of computerized tomography (CT) images, it remains a difficult task to accurately segment lung tissue in CT slices using a simple strategy. Motived by the aforementioned, a novel CT lung segmentation method based on the integration of multiple strategies was proposed in this paper. Firstly, in order to avoid noise, the input CT slice was smoothed using the guided filter. Then, the smoothed slice was transformed into a binary image using an optimized threshold. Next, a region growing strategy was employed to extract thorax regions. Then, lung regions were segmented from the thorax regions using a seed-based random walk algorithm. The segmented lung contour was then smoothed and corrected with a curvature-based correction method on each axis slice. Finally, with the lung masks, the lung region was automatically segmented from a CT slice. The proposed method was validated on a CT database consisting of 23 scans, including a number of 883 2D slices (the number of slices per scan is 38 slices), by comparing it to the commonly used lung segmentation method. Experimental results show that the proposed method accurately segmented lung regions in CT slices.

1. Introduction

Accurate lung segmentation is very important to ensure the performance of computer-aided lung diseases diagnosis (CAD) systems [1]. A recent study shows that 17% of true positives were missed because of poor lung segmentation [2]. Hence, there has been a growing interest in automated and accurate segmentation methods for lung CT images in recent years. Studies have reported on many methods, which are generally classified as threshold-based [3], region-based [4–6], and deformable-model-based methods [7–10]. Though they provide good results, no method has demonstrated robust and accurate results across the wide range of clinical imaging parameters and pathology faced in clinical practice. And so it remains an ongoing challenge to segment lung regions accurately in CT slices because of the complex anatomy of the thorax and image artifacts.

In this paper, a novel method for accurately segmenting lung in CT (computed tomography) slice based on the integration of multiple segmentation strategies was proposed. The method was derived of several different simple strategies. Firstly, in order to avoid noise disturbances, the input CT slice was smoothed using a guided filter. Then, the smoothed slice was transformed into a binary image using an optimized threshold. Next, a region growing and random walk strategy was employed to obtain the masks of the lungs. Finally, with the lung masks, lung regions were automatically segmented from a CT slice. The proposed method was assessed on 23 lung CT scans with 883 2D slices. Experiments indicate that this method achieved an average absolute border distance of 0.62 mm compared to manually segmented ground truths.

The primary contributions of the paper are summarized as follows:

(1) A fully automatic approach for accurate lung segmentation is developed by effective integrating multiple well developed simple image preprocessing strategies. The whole process does not need any user interaction.

(2) The affection of image noise on integrity of lung segmentation is counted.

(3) To make the segmentation of lung region more accurate, the contours of initial segmented lungs are refined using a curvature-based correction method. The proposed method can include all juxtapleural nodules near the mediastinum.

The remainder of this paper is organized as follows. To easy understand the value of this work, related works are discussed in Section 2. The proposed method is described in Section 3. Subsequently, experimental results are presented and discussed in Sections 4 and 5, respectively. Finally, conclusions and further discussion are provided in Section 6.

2. Related Works

One of the most commonly used lung segmentation methods for CT images is the threshold-based method [3], where the contrast between the low-density lungs and the surrounding high-density chest wall is usually used to guide the segmentation process, and by which tissues having higher gray levels than the selected threshold are excluded from the thoracic region. Although threshold-based strategies can be used to extract the lung parenchyma, these methods may fail in intensity variation condition and incorrectly exclude some vital regions, for example, juxtapleural nodules, from the lung area. Region-based methods, such as region growing [4] and graph cutting [5], are useful for dealing with intensity variations. However, with the presence of dense pathology in the lung field, it is not enough for successful delineation only with intensity. Deformable-model-based methods are widely used for the segmentation of pathological lungs [9, 10]. For example, in [9], an initial contour close to the lung boundary is firstly obtained, and then the contour reaches the object border. The limitation of this method is that some edge information might be lost when the edge of the contour is cracked. In [10], the level set method is employed to find the lung boundaries using energy minimization procedures. Though these methods are accurate in segmenting lung regions, they might lead to local minima states.

Recently, random-walk-based segmentations [12], in which graph optimization is employed to obtain accurate segmentation with user interaction based on selected seed pixels, are attracting increasing interest. Though this method shows excellent performance in image segmentation, it is sensitive to seed pixels.

3. The Proposed Method

Figure 1 shows a flow of the proposed method. It effectively integrates six simple strategies in three operation steps, CT image processing, lung region initial segmentation, and lung region refinement. For the whole processing, the method began with image denoising using the guided filter [13]. Then, a threshold is selected to binarize the filtered CT images using the Otsu algorithm, and the thorax regions are then extracted by region growing. In this step, the artifacts external to the patients bodies are removed. Next, prior knowledge is used for the automated selection of foreground seeds, defined as the lung confidence region. After the estimating of seed points and that of the lung intensity range, a seed-based random walk algorithm is applied to segment lung regions from the thorax region. Finally, holes in the segmented lung region are filled with a rolling-ball algorithm, and an iterative weighted averaging and adaptive curvature threshold is used to smooth and correct the segmented lung contour on each axis slice.

Details of each operation are described in the following sections.

3.1. CT Image Denoising Using Guided Filter. The aim of this operation is to smooth intraregion and to preserve the interregion edges of the images, which is of benefit to the following processing, because all operations in following steps including thorax extraction by region growing and lung segmentation with random walk are sensitive to image noise. Conventional filtering methods, such as Gaussian filter, mean filter, and the median filter, often obtain poor results as they incur more edge blurring and detail loss. However, some improved anisotropic diffusion filtering methods, such as guided filters and bilateral filters, can overcome this drawback by introducing an implicit edge detection step into the filtering process to encourage intraregion smoothing and preserve the interregion edges [13]. Guided filters [13] are widely used in image smoothing as an alternative to bilateral filters, as was done in this paper. Not only can a guided filter reduce computing time compared to a bilateral filter, but also image noise which can be incorrectly regarded as lung borders in many cases can be removed from lung parenchyma. The main idea of a guided filter is to filter input images by considering the content of the guidance image. Formally, given a guidance image, a guided filter is defined as follows:

$$q_i = a_k I_i + b_k, \quad \forall i \in \omega_k, \tag{1}$$

where q_i is a linear transformation of I_i in a window ω_k centered at the pixel (x, y) and a_k and b_k are the linear coefficients of local area ω_k and are supposed to be a constant. In this section, ω_k is assigned as 15×15. This local linear model ensures that q_i has an edge only if I_i has an edge. To make the difference between the output q_i and the input I_i as small as possible, the cost function $E(a_k, b_k)$ is minimized in window ω_k:

$$E(a_k, b_k) = \sum_{i \in \omega_k} \left(\left(a_k I_i + b_k - p_i \right)^2 + \varepsilon a_k^2 \right), \tag{2}$$

where ε is a regularization parameter keeping a_k from being too large. In this work, the value of ε is assigned as 0.008 according to our experience.

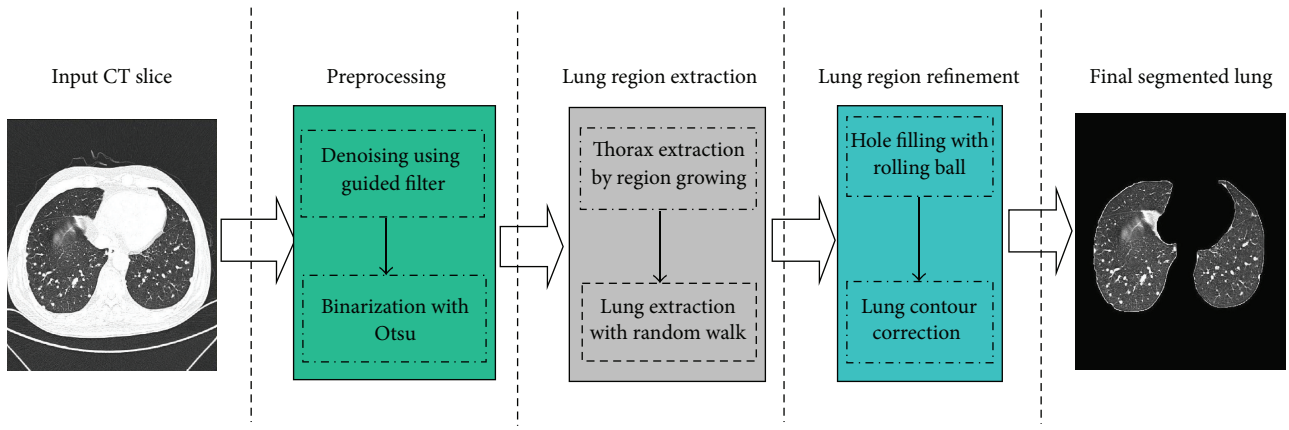
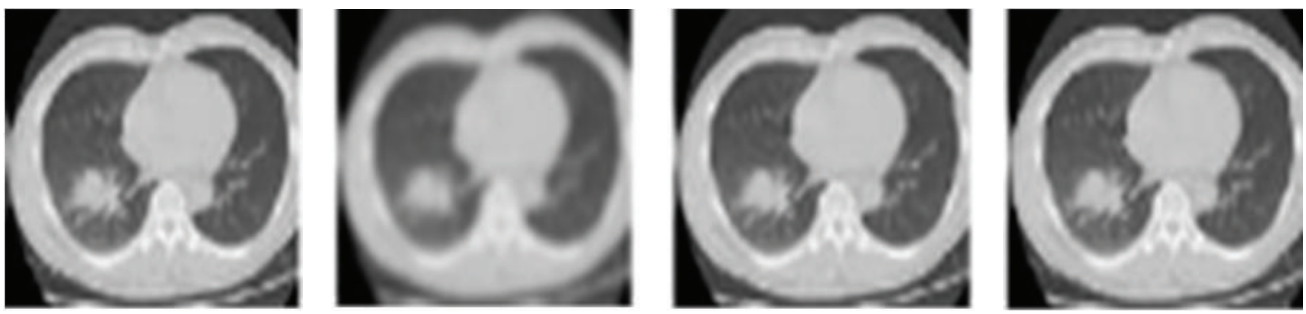


Figure 1: Flow chart of proposed method.



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

Figure 2: Comparison of CT image smoothing using the guided filter against other filtering methods. (a) Input CT image. (b) Smoothed CT image by Gaussian filter. (c) Smoothed CT image by mean filter. (d) Smoothed CT image by guided filter.

Through (2), the linear coefficients a_k and b_k can be computed as follows:

$$a_k = \frac{(1/\omega)\sum_{i\in\omega_k} I_i p_i - u_k \overline{p}_k}{\sigma_k^2 + \varepsilon}, \qquad b_k = \overline{p}_k - a_k u_k, \tag{3}$$

where u_k and σ_k^2 are the average value and the variance of the input image I_i in window ω_k. $|\omega|$ is the pixels number of window ω_k. $\overline{p}_k$ is the mean of the guided image in window ω_k. ε is the regularization parameter which is used to determine the intensity of changes in the pixels values.

In our work, the guided image filter is used to filter the 3 channels of RGB image, respectively, and the guided image is selected as the corresponding original channel component. Figure 2 shows a comparison of CT image smoothing using the guided filter against other filtering methods. It is obvious that the boundary of the smoothed image using the guided filter was clearer than when the other filters were used.

A quantitative comparison between the aforementioned filters in terms of PSNR (peak signal-to-noise ratio) was also conducted, as shown in Table 1. A higher PSNR value means that the image has high quality with less noise. It can be seen from Table 1 that the guided filter performed better than other denoise filters with the maximum PSNR of 63.1342. By contrast, the Gauss filter was inferior to both the guided filter and mean filter with a smaller PSNR of 61.4568. These results were consistent with those shown in Figure 2.

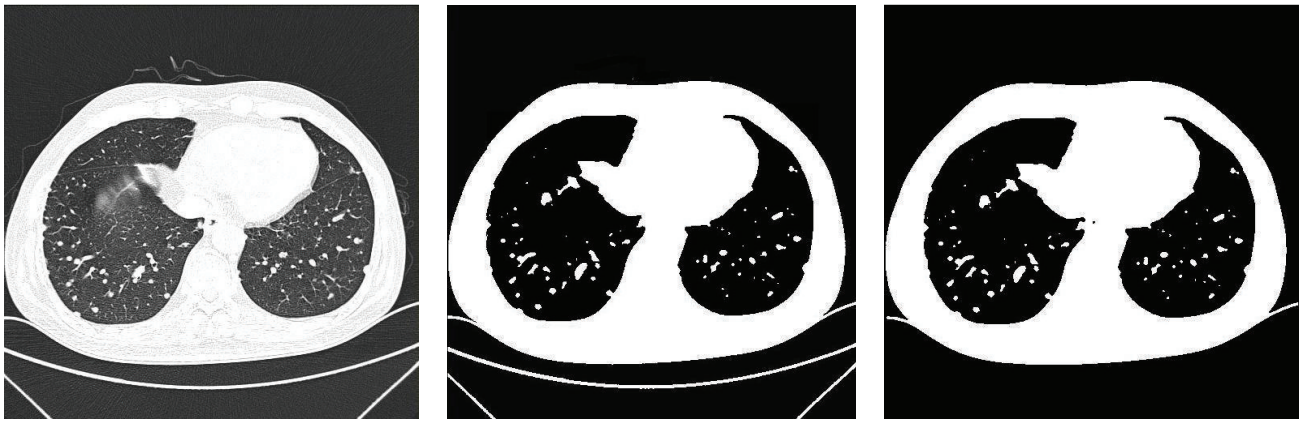
Table 1: Quantitative comparison of different denoise filters.

	Ground-truth marker	The number of locations recognized by the system
Images during day	100	98
Images during night	100	90

3.2. CT Image Binarization. In this step, Otsu's adaptive thresholding method [3] is employed to obtain a binarized CT image. The purpose of this operation is to simply follow operations for lung segmentation. For a given image, let L represent the grey level of the pixels $[1, 2, \ldots, L]$. By choosing a threshold at grey level k, the pixels are divided into object class C_0 and background class C_1.

Let ω_0 and ω_1 be the probabilities of C_0 and C_1 separated by a defined threshold, and let σ_0^2 and σ_1^2 be the variances of the two kinds. The variance of intrakind is defined as the weighted sum of the two variances [3], as in the following:

$$\sigma_{\text{intra}}^2(k) = \omega_0(k)\,\sigma_0^2(k) + \omega_1(k)\,\sigma_1^2(k). \tag{4}$$



(a) (b) (c)

Figure 3: Thorax extraction by region growing. (a) Input CT image. (b) Binarized CT image. (c) Extracted thorax using region growing.

The optimal threshold T is calculated as the value minimizing $\sigma^2_{\text{intra}}(k)$, as in the following:

$$T = \underset{k\in[1,L]}{\text{argmin}}\ \sigma^2_{\text{intra}}(k). \tag{5}$$

3.3. Thorax Extraction by Region Growing. In chest CT images, there are two main basic regions with different density distributions. The first is the low-density region, which contains background air, lungs, and airways, and the second is the high-density region, which includes the chest wall and bed and lung nodules.

The goal of this operation is to reduce artifacts external to the patients bodies to a certain extent. Based on the density of chest CT images, to extract the thorax from the CT images in this step, region growing [4] is used in the thresholded chest CT images to discard the background. For this purpose, seed pixels are selected from the four corners of the background in each axial CT image firstly, which then grows to all pixels in the four neighborhoods. The region growing process is repeated until there are no more adjacent pixels with a lower density than threshold value T. Then, the background region is obtained. The background image is subtracted from the binarization CT image and then the thorax region is extracted. Figure 3 shows an example of thorax extraction using region growing.

3.4. Extracting Lung from Thorax Region with Random Walk Algorithm. Considering the distribution of density of thorax tissues in CT images, random walk [12] strategy is employed for extracting lung from the thorax region in this step. Random walk is a seed-based graph method in which an image is considered a discrete object described with a weighted graph, where image pixels are taken as nodes connected by undirected edges. Taking an undirected graph

$$G = (V, E), \tag{6}$$

where V describes the set of vertex and E is the set of edges, and letting ω_{ij} be the edge weight that represents the probabilities between two neighboring nodes, the weight ω_{ij} can be defined as follows:

$$\omega_{ij} = \exp\left(-\beta\left(g_i - g_j\right)^2\right), \tag{7}$$

where g_i indexes an image feature at pixel i such as intensity gradients, which indicates the relationship of pixels to an image. The parameter β is the only parameter that can be adjusted in this method. The weights of ω_{ij} edges range from 0 to 1, letting 1 represent similar pixels and 0 represent dissimilar pixels. Given a small number of seeds in different locations, the random-walk-based lung segmentation will start its task at a pixel that reaches prelabeled seeds first by measuring the greatest transition probability. Essentially, the exact solution to the desired random walk is to minimize the Dirichlet energy with boundary conditions. The Dirichlet integral can be described as follows:

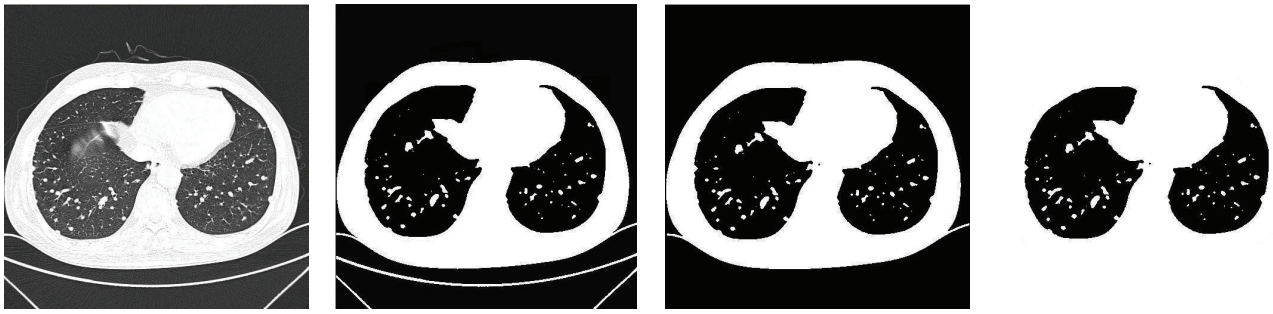
$$D[x] = \frac{1}{2}x^T L x = \frac{1}{2}\sum_{\omega_{ij}\in E}\omega_{ij}\left(x_i - x_j\right)^2, \tag{8}$$

where the function x is only the critical points, which will be minima, and L is a Laplacian matrix described as follows:

$$D[x] = \frac{1}{2}x^T L x = \frac{1}{2}\sum_{\omega_{ij}\in E}\omega_{ij}\left(x_i - x_j\right)^2. \tag{9}$$

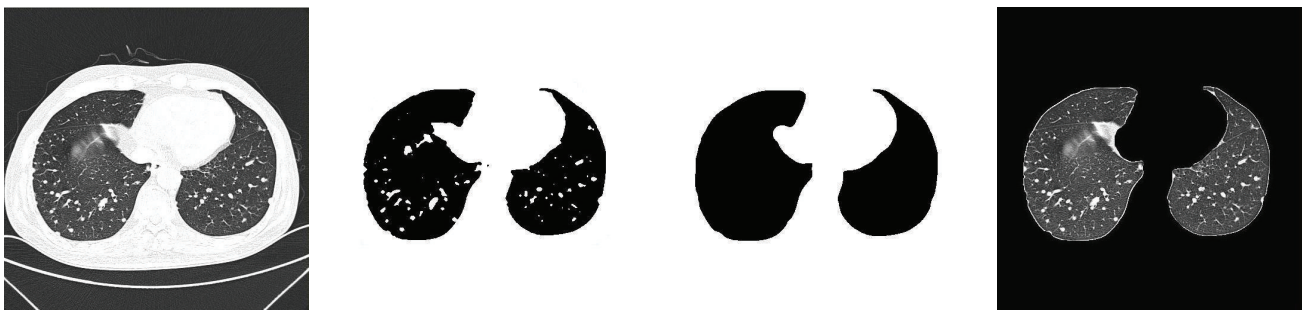
Derived from the Laplacian graph expressing the image, the analysis and computation of the probabilities are obtained by resolving a set of sparse and positive definite linear equations. In random walk processing, each step usually works with previous steps independently. And then its behavior is absolute according to a transition probability matrix L.

Research shows that the random walk method demonstrates good performance in image segmentation and is sensitive to initial seeds. In CT images, the intensity of lung tissue is usually in 400 HU to 600 HU, while the chest wall, blood, and bone are usually above 100 HU [4]. Aimed at the issue mentioned above, pixels with 500 HU within the thorax region are selected as initial seeds firstly, and then a set of



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

Figure 4: Final result for lung extraction. (a) Input CT image. (b) Binarized CT image. (c) Extracted thorax by region growing. (d) Extracted lung region by random walk.



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

Figure 5: Example of final refined lung mask and final segmented lung region. (a) Input CT image. (b) Extracted lung region using random walk. (c) Refined lung segmentation mask. (d) Final segmented lung region.

pixels with minimum HU values surrounding the initial seeds are sampled as seeds. Once seeds and affinity parameters for the random walk are set, lung delineation is performed. In this work, three initial seeds are automatically selected. Figure 4 shows a final result for this step.

3.5. Lung Region Refinement. As can be seen in Figure 4, it is obvious that though the lung region was extracted, there were many holes in the extracted lung regions, and parts of the regions were also excluded from the extracted lung region, which may lead to important tissue information being lost. In order to overcome such problems, holes in the segmented lung region are filled with a rolling-ball algorithm, and the contours of the segmented lungs are refined by a curvature-based correction method [14] in which a scan line search is used to calculate the curvatures of scanned points on the preliminary contour described by the random walk. In order to cut down computation time, each CT slice is scanned in the horizontal direction with a predefined interval l by seeking the intersection points, which is experimentally set to 3 pixels. The intersection points of the scanned edge are classified into three species, namely, the first point (P_i^{first}), the last point (P_i^{last}), and the middle point (P_i^{middle}). The middle point should be removed, as it is frequently found around the mediastinum or indentations that include the lung nodules. Thus, only the first and last intersection points are retained as they correspond to the lateral and medial lung contours. The curvature of the first and the last intersection points is computed as follows:

$$k_i = (x_{i-1} - x_i) \times (y_{i-1} - 2y_i + y_{i+1}) \times \frac{(y_{i-1} - y_i) \times (x_{i-1} - 2x_i + x_{i+1})}{\left[(x_{i-1} - x_i)^2 + (y_{i-1} - y_i)^2\right]^{3/2}}, \quad (10)$$

where (x_i, y_i) denotes an intersection point on i_{th} search line. Similarly, (x_{i-1}, y_{i-1}) and (x_{i+1}, y_{i+1}) are the same on the previous and following scan lines. For most natural images, taking into account high curvatures located at small perturbations, such as at the base and areas of normal lungs, the differences are used to rule out unnecessary points. Figure 5 shows an example of the final refined lung mask and final segmented lung region. As seen, the proposed method worked well.

4. Experiments

In this section, the clinical materials used in this work and the evaluation criteria are described. Then, the detailed results are presented, which include the visualization of segmentation errors and quantitative and statistical accuracy comparisons. Finally, the issues and limitations that were observed in the experiments are discussed. All methods were implemented

in Matlab and tested using a 2.3 GHz Intel Core i3 computer with 2 GB RAM.

4.1. Materials. The database used in this study consisted of 23 CT scans, including a total of 883 2D slices, which were acquired using MDCT scanner (GE Light-Speed Ultra, Milwaukee, WI, USA) with 120 kVp and 100 mA in the medical school of Xian Jiaotong University. The number of slices per scan is about 38 slices per scan. Each CT slice had an image matrix of 512 by 512 (16-bit depth) pixels. Pixel size ranged between 0.625 mm and 0.742 mm, with a mean value of 0.692 mm, depending on the physical size of the patient.

4.2. Evaluation Method. Quantitative evaluation of lung segmentation is important because it not only provides a reliable basis for clinical application but also indicates its relative performance with respect to other used methods [4]. However, conducting an evaluation of a lung segmentation method is still difficult. One reason is that the true lung boundary is unknown, and the reference standard often refers to several experts consensus [6]. In this study, the reference standard was produced in the following way to address this issue: all lung contours were first manually marked by an experienced radiograph expert and then reviewed by another radiologist. If the opinion of the second radiologist was different from that of the first one, the lung contours were corrected by the two radiologists under collaboration and the results were used as the reference standard. Although only limited radiologists involved in the manual segmentation might lead to bias, the difference between the lung boundaries obtained by this method and the reference standard can reflect the errors of the proposed method with respect to an expert.

Another reason for difficulty is that even though there are many metrics used to evaluate a lung segmentation method, such as dice similarity coefficient, jacquard similarity, false positive rate, and false negative rate, they do not provide both local and global impressions of the segmentation performance [6]. Because of this, the following three metrics for measuring the segmentation performance of the proposed method were employed: (1) oversegmentation rate, (2) undersegmentation rate, and (3) the average of absolute border distance. The experiments showed that these three metrics not only demonstrated an overview of the oversegmentation and undersegmentation but also confirmed the whole statistical distribution of segmentation error distances.

The oversegmentation rate is termed as the number of voxels in a segmented region which are included as part of the ROI but are not in the reference standard. Let V_{auto} represent the volume of the binary mask generated using the proposed approach and let V_{mannual} be the volume of the reference standard. The oversegmentation rate of $\text{OR}(V_{\text{auto}}, V_{\text{mannual}})$ can be found using the following:

$$\text{OR}\left(V_{\text{auto}}, V_{\text{mannual}}\right) = \left|\frac{V_{\text{auto}}/V_{\text{mannual}}}{V_{\text{mannual}}}\right|, \tag{11}$$

where $V_{\text{auto}}/V_{\text{mannual}}$ represents the relative complement of V_{auto} in V_{mannual}. Similarly, the undersegmentation rate of $\text{UR}(V_{\text{auto}}, V_{\text{mannual}})$ is defined as the relative lung volume amount which is regarded as lung tissue in the reference standard but not in a segmented image region with an automatic segmentation method:

$$\text{UR}\left(V_{\text{auto}}, V_{\text{mannual}}\right) = \left|\frac{V_{\text{auto}}/V_{\text{mannual}}}{V_{\text{mannual}}}\right|. \tag{12}$$

The average of the absolute border distance (ABD) is a statistical measurement of the fitting between the lung surfaces generated by a segmentation method and the lung surfaces in the reference standard. It is used to measure the spatial similarity between the lung boundaries generated by a segmentation approach and that of the reference standard. The shortest distance between a point on the lung surface obtained by the proposed algorithm and the lung surface of the reference standard was used to generate the absolute border distance.

Table 2: Quantitative comparison of segmentation results.

Methods	OR	UR	ABD (mm)
Region growing	2.1%	2.7%	0.72
Active contour	1.9%	2.38%	0.64
Proposed method	1.87%	2.36%	0.62

4.3. Qualitative Results. Figure 6 shows the results between ground truth and the proposed method. Figure 6(a) shows the input CT slices. The ground truth is in Figure 6(b), which was manually marked by an experienced radiograph expert. Figure 6(c) displays the segmented results using the proposed method. As can be seen, the proposed method's segment results were closest to the ground truth, which indexes the effectiveness of the proposed method.

A comparison of the proposed approach with two often-used state-of-the-art lung segmentation methods, the active-contour-based method [9] and the region growing-based method [4], is shown in Figure 7. Figure 7(a) shows the input CT images, and Figure 7(b) shows the segmented results using the active-contour-based method. In Figure 7(c), the region growing-based method is shown. The proposed method's segmentation results are shown in Figure 7(d). From Figure 7(c), it can be seen that even though the lung boundaries are well smoothed, part of the pleural regions in the mediastinum was excluded. In Figure 7(b), although no lung regions were excluded, parts of nonlung regions were erroneously included. However, Figure 7(d) shows sufficient pleural nodule regions and diffuse areas are also included. This indicates that the proposed method exhibits a more powerful discriminating ability compared to other methods.

4.4. Quantitative Results. Table 2 shows comparisons between the proposed method and two state-of-the-art techniques, the region growing method and the active contour method, to the manually defined ground truth using prior-mentioned metrics. It can be seen that the average of the absolute border distance, the oversegmentation rate, and the undersegmentation rate of the proposed method

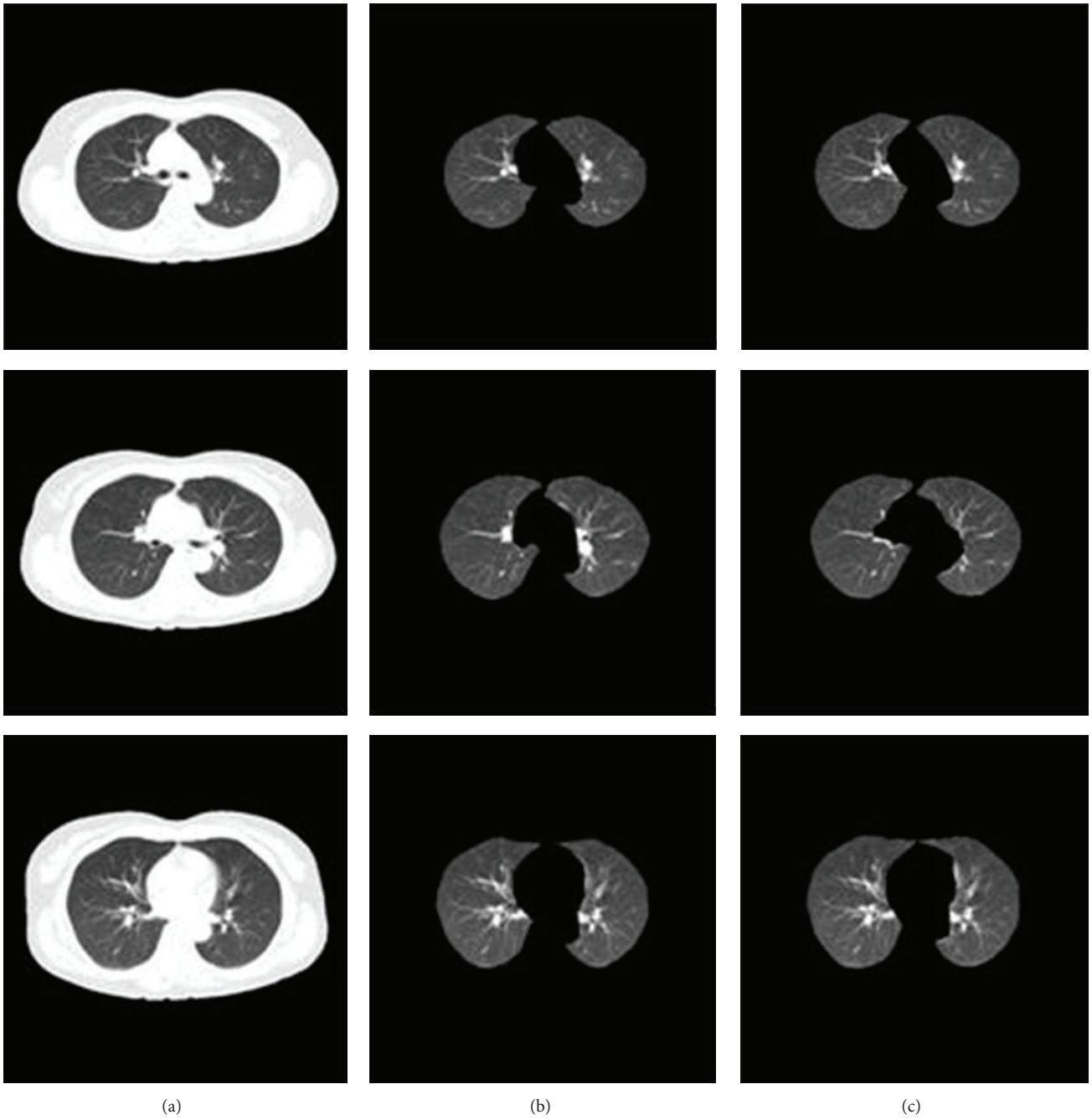


Figure 6: Comparisons between ground truth and the proposed method. (a) Input CT images. (b) Ground truth. (c) Results using proposed method.

were 0.62 mm, 1.87%, and 2.36%, respectively. These results were better than those of the region growing-based method (0.72 mm, 2.1%, and 2.7%) and the active-contour-based method (0.64 mm, 1.9%, and 2.38%). This indicated that the proposed method achieved more accurate and robust results than the other approaches.

5. Discussion

As an integration of multiple simple image segmentation strategies, the method proposed in this paper possesses several advantages over single segmentation strategy methods, as illustrated in Figure 8. Figure 8(a) shows input CT slices, Figure 8(b) shows the results obtained with the threshold-based method, Figure 8(c) is the region growing-based method, Figure 8(d) is the active-contour-based method, Figure 8(e) is the random walk method, Figure 8(f) is the active contour and curvature correction method, and Figure 8(g) shows the segmented results using the proposed method. Figure 8(h) shows the reference standard. As can be seen in the images, compared with the reference standard, the main trachea was excluded from the segmented lung

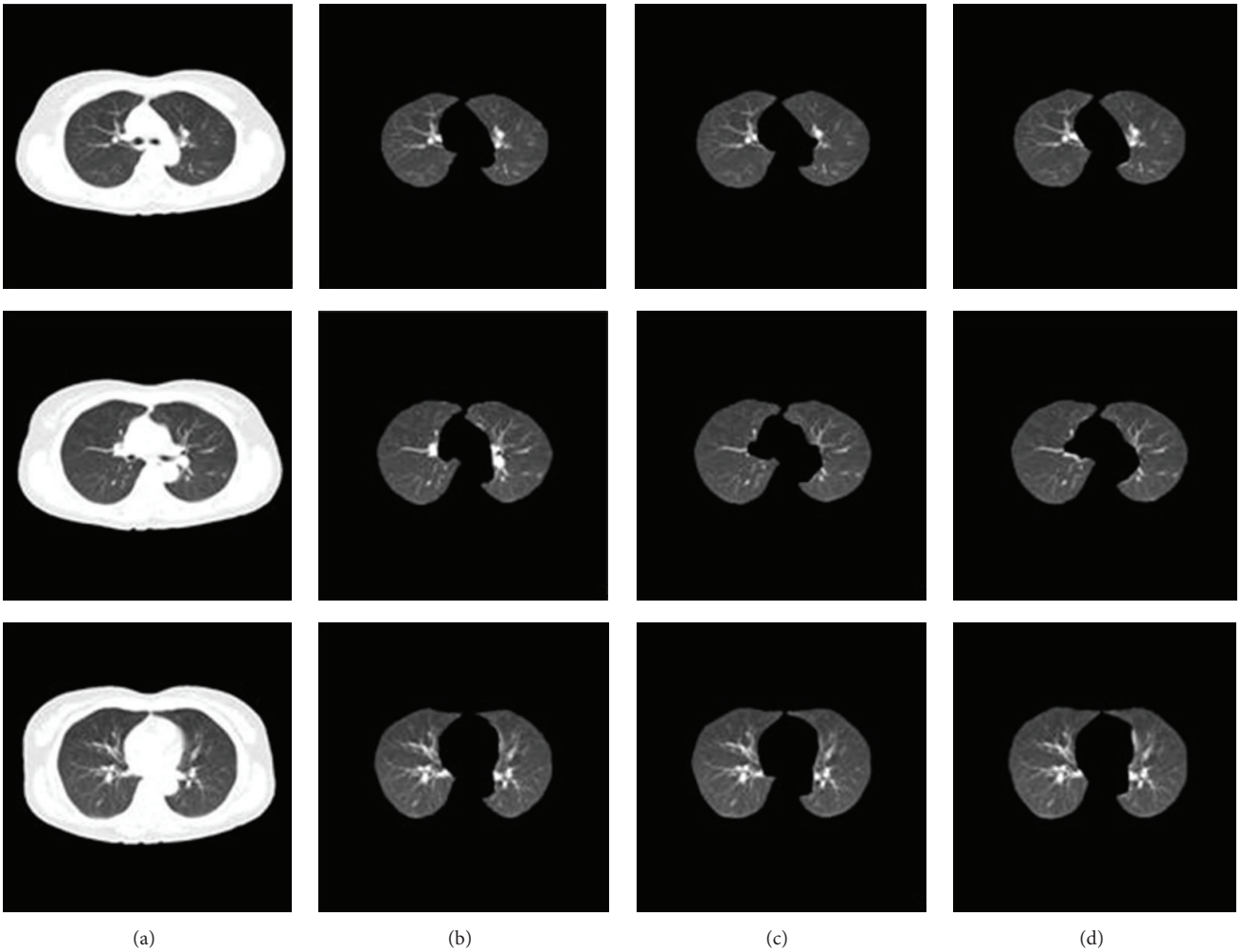


Figure 7: Comparison of segmentation results. (a) Original CT images. (b) Active-contour-based method. (c) Region growing-based method. (d) The proposed method.

area using the threshold-based method. The region growing-based method made the lung edge rough, and good segmentation results of lung regions were not achieved due to the limiting of growth rule. In the evolution of the active contour curves, the diffuse area was left out. The random-walk-based method produced inaccurate segmentation results. As can be seen from the results of the active contour and curvature-based correction method, the lung boundary was smoothed with undersegmentation in the mediastinum. By comparison, the developed approach demonstrated a more powerful discriminating ability and included sufficient pleural nodule regions and diffuse areas.

A quantitative comparison between the segmentation results obtained with an assemblage of multiple segmentation strategies against other segmentation strategies was also performed using the overlap ratio between the manually outlined contours and computer-defined outlines, as shown in Table 3. The overlap ratio can be defined using the following:

$$\text{overlap ratio} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}}, \qquad (13)$$

where N_{TP} represents the number of correctly segmented pixels in a lung region, N_{TN} stands for the number of correctly segmented pixels in the background area, and N_{FP} and N_{FN} are the missegmented lung regions and background area, respectively. A high overlap ratio indicates accurate segmentation results. As can be seen from Table 3, the average overlap ratio obtained using the proposed method was 98.4%, whereas that obtained using the threshold-based method was 94.1%, the region growing overlap ratio was 95.3%, the active contour was 94.4%, the active contour with a curvature-based correction method was 95.8%, and the random-walk-based method was only 93.8%. These results were consistent with those shown in Figure 8. The conclusion from these results is also consistent with the outcome mentioned in Table 2.

A comparison between our method and some recently published, independent methods, such as the graph-cut based method presented in [5] and the method proposed in [11], is shown in Figure 9 and Table 4. It can be seen from Figure 9 that the segmented lung images are very similar, and the difference between the segmented lung image with our method and that with the graph-cut based method presented

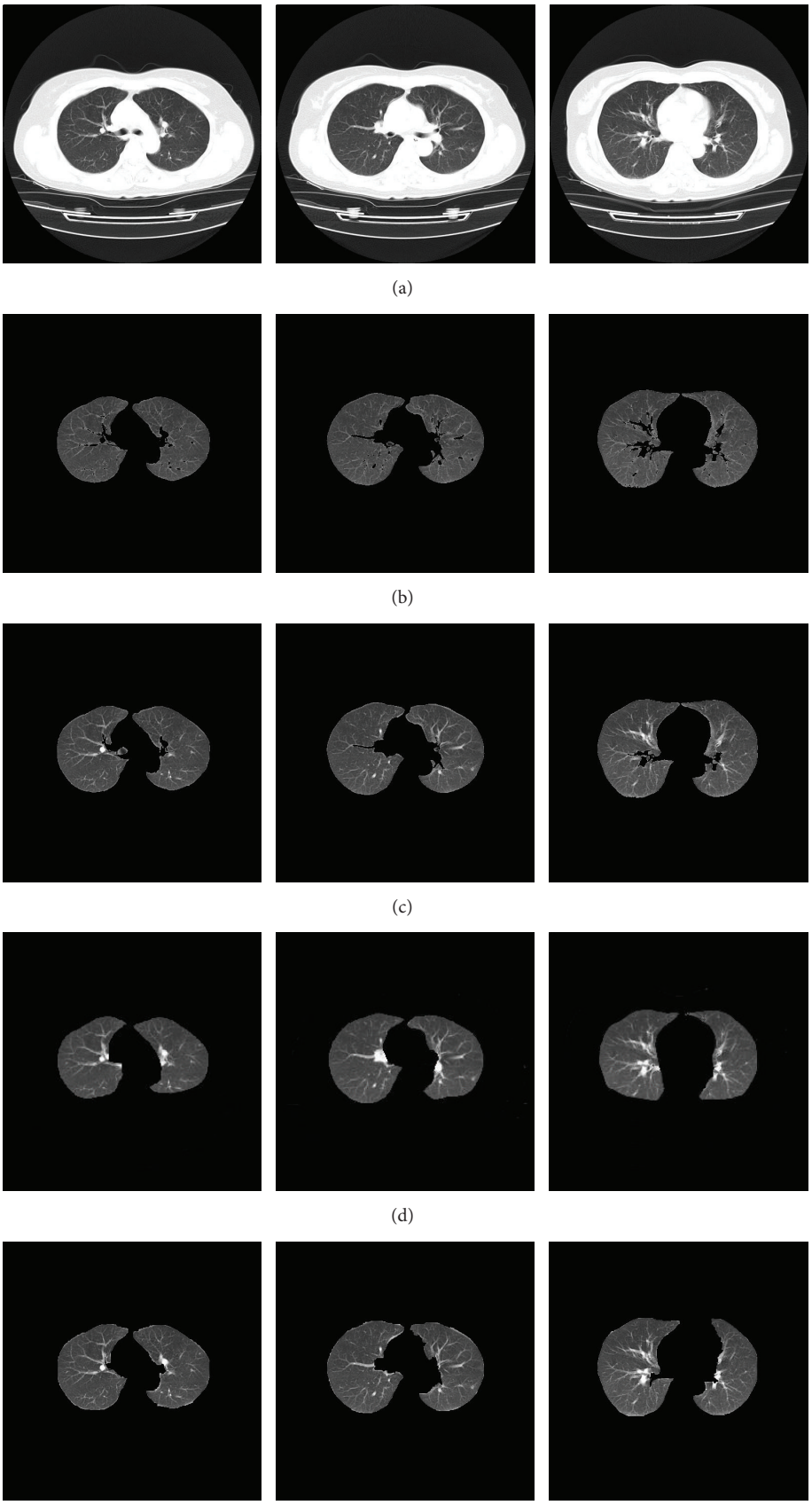


Figure 8: Continued.

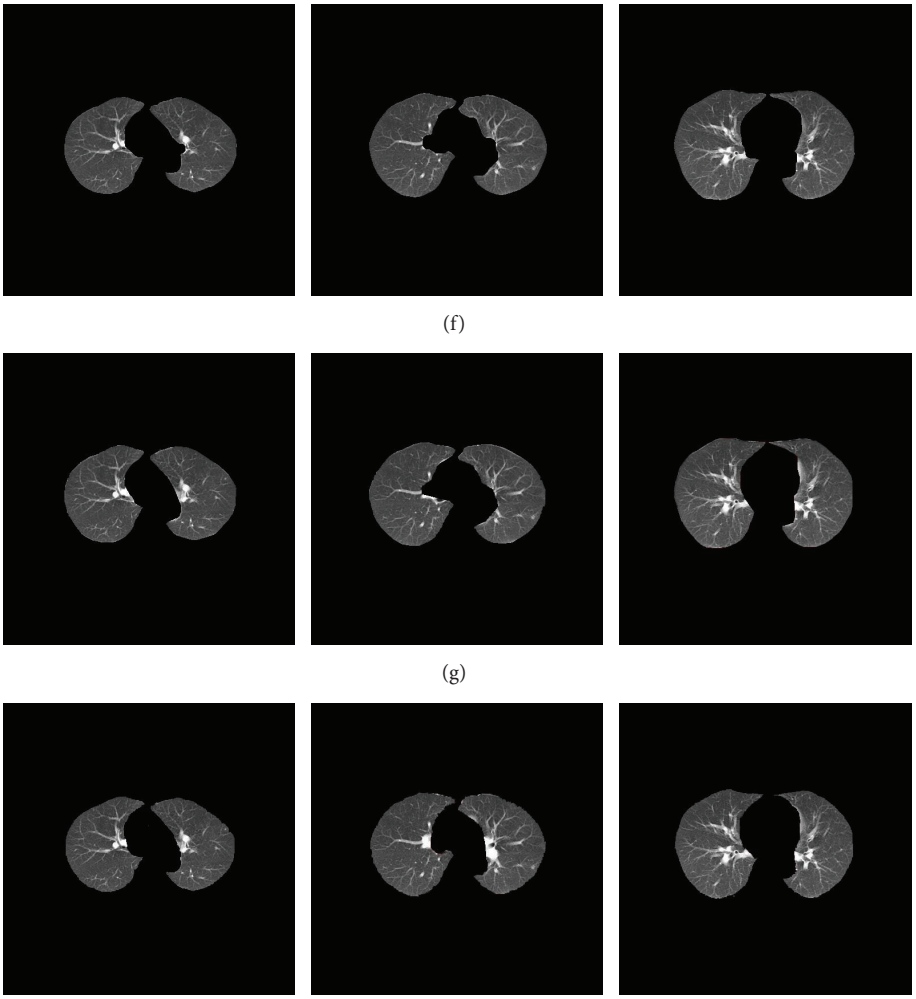


Figure 8: Comparisons of segmentation results. (a) CT images. (b) Threshold method. (c) Region growing method. (d) Active contour method. (e) Random walk method. (f) Active contour with curvature-based correction method. (g) Proposed method. (h) Reference standard.

Table 3: Quantitative comparison between segmentation results.

Methods	Overlap ratio (%)
Threshold	94.1
Region growing	95.3
Active contour	94.4
Random walk	93.8
Active contour with curvature correction	95.8
Proposed method	98.4

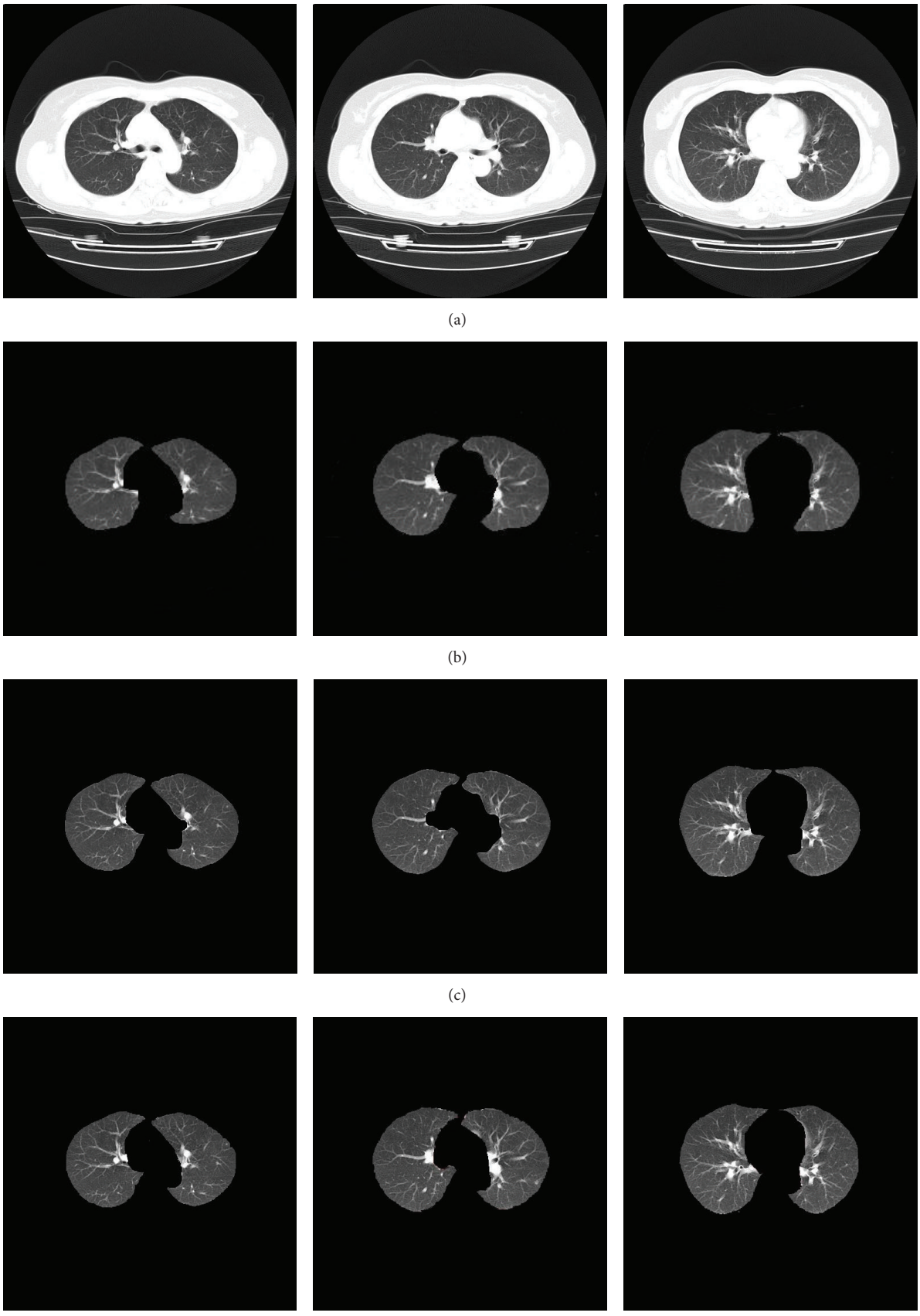
in [5] and also that with the method proposed in [11] are too tiny to observe.

Table 4 shows the comparison in terms of the oversegmentation rate and the undersegmentation rate. As can be seen, the oversegmentation rates of our proposed method, of the graph-cut method [5], and of the method proposed in [11] are 1.87%, 1.88%, and 1.86%, respectively, whereas the undersegmentation rates of these are 2.36%, 2.34%, and 2.37%, respectively. This indicated that the segmentation accuracies of the three methods mentioned above are similar. This conclusion is consistent with the outcome demonstrated in Figure 9.

Table 4: Quantitative comparison of segmentation results in terms of the oversegmentation rate and the undersegmentation rate.

Methods	OR	UR
Method in [11]	1.86%	2.37%
Method in [5]	1.88%	2.34%
Proposed method	1.87%	2.36%

A comparison of running time for all methods mentioned above is given in Table 5. It can be seen that the proposed method requires longer working times (1.68 s for one slice segmentation) compared to previously well-established methods, such as threshold-based method (0.28 s for one slice segmentation), region growing (0.38 s for one slice



(a)

(b)

(c)

(d)

Figure 9: Continued.

(e)

Figure 9: Comparisons of segmentation results. (a) Original CT images. (b) The method proposed in [11]. (c) The graph-cut based method presented in [5]. (d) Our method. (e) Reference standard.

Table 5: Working times of different methods.

Methods	Average running time (seconds)
Threshold	0.28
Region growing	0.38
Active contour	0.54
Random walk	0.58
Active contour with curvature correction	0.64
Method in [11]	2.58
Method in [5]	3
Proposed method	1.68

segmentation), active contour (0.54 s for one slice segmentation), random walk (0.58 s for one slice segmentation), and active contour with curvature correction (0.64 s for one slice segmentation), whereas compared to the graph-cut based method [5] (3 s for one slice segmentation) and the method proposed in [11] (2.58 s for one slice segmentation), the running time of our method is significantly short. What is the reason for this? For the proposed method and previously well-established methods, the reason is that the new method is an integration of six simpler image segmentation strategies meaning more time is needed to carry out all the steps. For the proposed method and the graph-cut based method, the reason is that the use of expectation maximization to calculate the weight that each pixel belongs to the foreground object leads to a longer running time than that of our proposed method. For the proposed method and the method proposed in [11], the reason is that the fuzzy c-means method which is used for lung identification might lead to a longer running time.

Considering the compromise between the accuracy of the segmented results and the computing time of the whole method, it is obvious that our proposed method is more efficient.

6. Conclusions

In this paper, an assemblage of several simple image segmentation strategies was proposed for segmenting lung regions in chest CT images. The effectiveness of this approach was demonstrated on 23 CT scans, and the results were compared to the manual segmentations of an expert and results obtained with the two most esteemed techniques. Experimental results showed that this method was more accurate in lung segmentation compared to other methods.

It should be noted that accurate segmentation of lung regions in the presence of severe pathologies, such as lung cancer, is still a challenging task. Future work will mainly focus on the segmentation algorithm of lung tissue characterized with severe abnormalities.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by a grant from the National Natural Science Foundation of China (no. 61202198 and no. 61401355), a grant from the China Scholarship Council (no. 201608610048), and the Nature Science Foundation of Science Department of PeiLin count at Xian (GX1619).

References

[1] T. Doel, D. J. Gavaghan, and V. Grau, "Review of automatic pulmonary lobe segmentation methods from CT," *Computerized Medical Imaging and Graphics*, vol. 40, pp. 13–29, 2015.

[2] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen, "Random walk and graph cut for co-segmentation of lung tumor on PET-CT-images," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5854–5867, 2015.

[3] S. S. Mokri, M. I. Saripan, M. H. Marhaban, and A. J. Nordin, "Lung segmentation in CT for thoracic PET-CT registration

through visual study," in *Proceedings of the 2nd IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES '12)*, pp. 550–554, Langkawi, Malaysia, December 2012.

[4] Y. Iwao, T. Gotoh, S. Kagei, T. Iwasawa, and M. de Sales Guerra Tsuzuki, "Integrated lung field segmentation of injured region with anatomical structure analysis by failure-recovery algorithm from chest CT images," *Biomedical Signal Processing and Control*, vol. 12, no. 1, pp. 28–38, 2014.

[5] S. Dai, K. Lu, J. Dong, Y. Zhang, and Y. Chen, "A novel approach of lung segmentation on chest CT images using graph cuts," *Neurocomputing*, vol. 168, pp. 799–807, 2015.

[6] A. Mansoor, U. Bagci, Z. Xu et al., "A generic approach to pathological lung segmentation," *IEEE Transactions on Medical Imaging*, vol. 33, no. 12, pp. 2293–2310, 2014.

[7] S. Sun, C. Bauer, and R. Beichel, "Automated 3-D segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 449–460, 2012.

[8] S. Sun, C. Bauer, and R. Beichel, "Automated 3-D segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 449–460, 2012.

[9] Y. Li, Z. Miao, and B. Wang, "Segmentation of lung CT with pathologies based on adapt active appearance models," in *Proceedings of the 3rd International Conference on Computer Science and Network Technology (ICCSNT '13)*, pp. 1119–1121, Dalian, China, October 2013.

[10] M. R. D. Raj and C. H. Sulochana, "An efficient lung segmentation approach for interstitial lung disease," in *Proceedings of the 2014 International Conference on Circuits, Power and Computing Technologies (ICCPCT '14)*, pp. 1211–1216, IEEE, Nagercoil, India, March 2014.

[11] S. Zhou, Y. Cheng, and S. Tamura, "Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonary vessels on chest CT images," *Biomedical Signal Processing and Control*, vol. 13, no. 1, pp. 62–70, 2014.

[12] J. Zhou, Z. Yan, G. Lasio et al., "Automated compromised right lung segmentation method using a robust atlas-based active volume model with sparse shape composition prior in CT," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 47–55, 2015.

[13] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[14] Y. Yim and H. Hong, "Correction of segmented lung boundary for inclusion of pleural nodules and pulmonary vessels in chest CT images," *Computers in Biology and Medicine*, vol. 38, no. 8, pp. 845–857, 2008.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 4851401, 9 pages
http://dx.doi.org/10.1155/2016/4851401

Research Article

Estimation of Response Functions Based on Variational Bayes Algorithm in Dynamic Images Sequences

Bowei Shan

School of Information Engineering, Chang'an University, Shaanxi 710064, China

Correspondence should be addressed to Bowei Shan; bwshan@chd.edu.cn

Received 16 April 2016; Revised 8 June 2016; Accepted 17 July 2016

Academic Editor: Zexuan Ji

We proposed a nonparametric Bayesian model based on variational Bayes algorithm to estimate the response functions in dynamic medical imaging. In dynamic renal scintigraphy, the impulse response or retention functions are rather complicated and finding a suitable parametric form is problematic. In this paper, we estimated the response functions using nonparametric Bayesian priors. These priors were designed to favor desirable properties of the functions, such as sparsity or smoothness. These assumptions were used within hierarchical priors of the variational Bayes algorithm. We performed our algorithm on the real online dataset of dynamic renal scintigraphy. The results demonstrated that this algorithm improved the estimation of response functions with nonparametric priors.

1. Introduction

Highly rapid development of machine learning technique offers an opportunity to obtain information about organ function from dynamic medical images, instead of invasive intervention. The unknown input function can be obtained by deconvolution of the organ time-activity curve and organ response function. Typically, both the input function and the response functions are unknown. Moreover, the time-activity curves are also not directly observed since the recorded images are observed as superposition of multiple signals. Analysis of the dynamic image sequences thus require to separate the original sources images and their weights over the time forming the time-activity curves (TACs). The TACs are then decomposed into input function and response functions. Success of the procedure is dependent on the model of the image sequence.

The common model for dynamic image sequences is the factor analysis model [1], which assumes linear combination of the source images and TACs. Another common model is that TAC arises as a convolution of common input function and source-specific kernel [2, 3]. The common input function is typically the original signal from the blood and the role of convolution kernels varies from application area: impulse response or retention function in dynamic renal scintigraphy [4]. In this paper, we will refer to the source kernels as the response functions; however other interpretations are also possible.

Analysis of the dynamic image sequences can be done with supervision of experienced physician or technician, who follows recommended guidelines and uses medical knowledge. However, we aim at fully automated approach where the analysis fully depends on the used model. The most sensitive parameter of the analysis is the model of the response functions (i.e., the convolution kernels). Many parametric models of response functions have been proposed, including the exponential model [5] or piecewise linear model [6, 7]. An obvious disadvantage of the approach is that the real response function may differ from the assumed parametric models. Therefore, more flexible classes of models based on nonparametric ideas were proposed such as averaging over region [8], temporal regularization using finite impulse response filters [9], or free-form response functions using automatic relevance determination principle in [10].

In this paper, we will study the probabilistic models of response functions using Bayesian methodology within the general blind source separation model [11]. The Bayesian approach was chosen for its inference flexibility and for

its ability to incorporate prior information of models [12, 13]. We will formulate the prior model for general blind source separation problem with deconvolution [10] where the hierarchical structure of the model allows us to study various versions of prior models of response functions. Specifically, we design different prior models of the response functions with more parameters than the number of points in the unknown response function. The challenge is to regularize the estimation procedure such that all parameters are estimated from the observed data. We will use the approximate Bayesian approach known as the variational Bayes method [14]. The resulting algorithms are tested on synthetic as well as on real datasets.

2. Probabilistic Model of Image Sequences

A probabilistic model of image sequences is introduced in this section. Estimation of the model parameters yields an algorithm for Blind Source Separation and Deconvolution. Prior models of all parameters except for the response functions are described here while the priors for the response functions will be studied in detail in the next section.

2.1. Model of Observation. Each recorded image is stored as a column vector $\mathbf{d}_t \in \mathbf{R}^{p\times 1}$, $t = 1,\dots,n$, where n is the total number of recorded images. Each vector $\mathbf{d}_t$ is supposed to be an observation of a superposition of r source images $\mathrm{a}_k \in \mathbf{R}^{p\times 1}$, $k = 1,\dots,r$, stored again columnwise. The source images are weighed by their specific activities in time t denoted as $x_{1,t},\dots,x_{r,t} \equiv \overline{x}_t \in \mathbf{R}^{1\times r}$. Formally,

$$\mathbf{d}_t \in \mathrm{a}_1 x_{1,t} + \mathrm{a}_2 x_{2,t} + \cdots + \mathrm{a}_r x_{r,t} + \mathrm{e}_t = A\overline{x}_t^T + \mathrm{e}_t, \quad (1)$$

where e_t is the noise of the observation, $A \in \mathbf{R}^{p\times r}$ is the matrix composed of source images as its columns $A \in [\mathrm{a}_1,\dots,\mathrm{a}_r]$, and symbol $()^T$ denotes transposition of a vector or a matrix. Equation (1) can be rewritten in the matrix form. Suppose that the observation matrix $D = [\mathrm{d}_1,\dots,\mathrm{d}_n] \in \mathbf{R}^{p\times r}$ and the matrix with TACs in its columns $X = [\overline{x}_1^T,\dots,\overline{x}_1^T]^T \in \mathbf{R}^{n\times r}$. Note that we will use the bar symbol, $\overline{x}_k$, to distinguish the kth row of matrix X, while x_k will be used to denote the kth column. Then, (1) can be rewritten into the matrix form as

$$D = AX^T + E. \quad (2)$$

The tracer dynamics in each compartment is commonly described as convolution of common input function, vector $b \in \mathbf{R}^{n\times 1}$, and source-specific response function (convolution kernel, mathematically), vector $\mathbf{u}_k \in \mathbf{R}^{n\times 1}$, $k = 1,\dots,r$ [5, 6, 15]. Using convolution assumption, each TAC can be rewritten as

$$x_k = B\mathbf{u}_k, \quad \forall k = 1,\dots,r, \quad (3)$$

where the matrix $B \in \mathbf{R}^{n\times n}$ is composed of elements of input function b as

$$B = \begin{pmatrix} b_1 & 0 & 0 & 0 \\ b_2 & b_1 & 0 & 0 \\ \cdots & b_2 & b_1 & 0 \\ b_n & \cdots & b_2 & b_1 \end{pmatrix}. \quad (4)$$

Suppose that the aggregation of response function $U = [\mathbf{u}_1,\dots,\mathbf{u}_r] \in \mathbf{R}^{n\times r}$. Then, $X = BU$ and model (2) can be rewritten as

$$D = AU^T B^T + E. \quad (5)$$

The task of subsequent analysis is to estimate the matrices A and U and the vector b from the data matrix D.

2.2. Noise Model. We assume that the noise has homogeneous Gaussian distribution with zero mean and unknown precision parameter ω, $e_{i,j} = \mathcal{N}_{e_{i,j}}(0,\omega^{-1})$. Then, the data model (2) can be rewritten as

$$f(D \mid A, X, \omega) = \prod_{t=1}^{n} \mathcal{N}_{d_t}\left(A\overline{x}_t, \omega^{-1} I_p\right), \quad (6)$$

where symbol $\mathcal{N}$ denotes Gaussian distribution and I_p is identity matrix of the size given in its subscript. Since all unknown parameters must have their prior distribution in the variational Bayes methodology, the precision parameter (inverse variance) ω has a conjugate prior in the form of the Gamma distribution

$$f(\omega) = \mathcal{G}_\omega(\vartheta_0, \rho_0) \quad (7)$$

with chosen constants shape parameter ϑ_0 and scale parameter ρ_0, due to the homogeneous noise model.

2.3. Probabilistic Model of Source Images. The only assumption on source images is that they are sparse; that is, only some pixels of source images are nonzeros. The sparsity is achieved using prior model that favors sparse solution depending on data [16]. We will employ the automatic relevance determination (ARD) principle [17] based on joint estimation of the parameter of interest together with its unknown precision. Specifically, each pixel $a_{i,j}$ of each source image has Gaussian prior truncated to positive values (see Appendix A.1) with unknown precision parameter $\xi_{i,j}$ which is supposed to have conjugate Gamma prior as

$$\begin{gathered} f(a_{i,k} \mid \xi_{i,k}) = t\mathcal{N}_{a_{i,j}}\left(0, \xi_{i,k}^{-1}\right), \\ f(\xi_{i,k}) = \mathcal{G}_{\xi_{i,k}}(\phi_0, \psi_0) \end{gathered} \quad (8)$$

for $\forall i = 1,\dots,p$, $\forall k = 1,\dots,r$, and ϕ_0, ψ_0 are chosen constants. The precisions $\xi_{i,j}$ form the matrix Ξ of the same size as A.

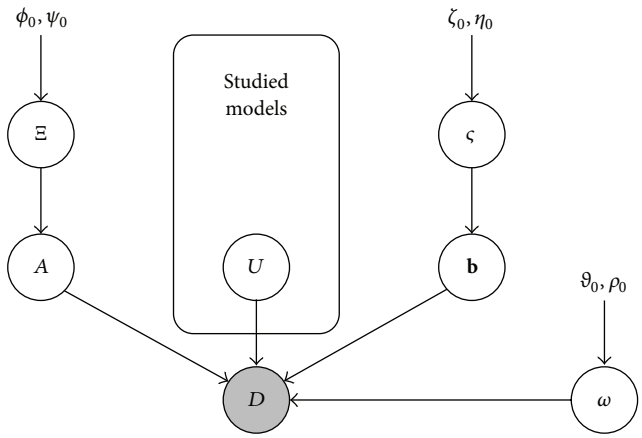


Figure 1: Hierarchical model for image sequences.

2.4. Probabilistic Model of Input Function. The input function $\mathbf{b}$ is assumed to be a positive vector; hence, it will be modeled as truncated Gaussian distribution to positive values with scaling parameter $\varsigma \in \mathbf{R}$ as

$$f(\mathbf{b} \mid \varsigma) = t\mathcal{N}\left(\mathbf{0}_{n,1}, \varsigma^{-1} I_n\right), \qquad f(\varsigma) = \mathcal{G}(\zeta_0, \eta_0), \tag{9}$$

where $\mathbf{0}_{n,1}$ denotes zeros matrix of the given size and ζ_0, η_0 are chosen constants.

2.5. Models of Response Functions. So far, we have formulated the prior models for source images A and input function $\mathbf{b}$ from decomposition of the matrix D. The task of this paper is to propose and study prior models for response functions U as illustrated in Figure 1. Different choices of the priors on the response functions have strong influence on the results of the analysis which will be studied in the next section.

3. Nonparametric Prior Models of Response Function

Here, we will formulate several prior models of response functions. Our purpose is not to impose any parametric form as it was done, for example, in [5, 6], but to model response function as a free-form curve with only influence from their prior models. The motivation is demonstrated in Figure 2, where a common parametric model [6] is compared to an example of response function obtained from real data. While the basic form of the response function is correct, exact parametric form of the function would be very complex. Therefore, we prefer to estimate each point on the response function individually. However, this leads to overparameterization and poor estimates would result without regularization. All models in this section introduce regularization of the nonparametric function via unknown covariance of the prior with hyperparameters.

3.1. Orthogonal Prior. The first prior model assumes that each response function $\mathbf{u}_k$, $k = 1, \ldots, r$, is positive and each response function is weighed by its own precision relevance parameter $\nu_k \in \mathbf{R}$ which has a conjugate Gamma prior:

$$f(\mathbf{u}_k \mid \nu_k) = t\mathcal{N}_{\mathbf{u}_k}\left(\mathbf{0}_{n,1}, \nu_k^{-1} I_n\right), \qquad f(\nu_k) = \mathcal{G}_{\nu_k}(\alpha_0, \beta_0) \tag{10}$$

for $\forall k = 1, \ldots, r$ and where α_0, β_0 are chosen constants.

The precision parameters ν_k serve for suppression of weak response functions during iterative computation and therefore as parameters responsible for estimation of the number of relevant sources.

3.2. Sparse Prior. The model with sparse response functions has been introduced in [10]. The key assumption of this model is that the response functions are most likely sparse which is modeled similarly as in case of source images, Section 2.2, using the ARD principle. Here, each element of response function $u_{k,j}$ has its relevance parameter $\nu_{k,j}$ which is supposed to be conjugate Gamma distributed. In vector notation, each response function $\mathbf{u}_k$ has its precision matrix Υ_k with precision parameters $\nu_{k,j}$ on its diagonal and zeros otherwise. Then

$$f(\mathbf{u}_k \mid \Upsilon_k) = t\mathcal{N}_{\mathbf{u}_k}(\mathbf{0}_{n,1}, \Upsilon_k), \qquad f\left(\nu_{k,j}\right) = \mathcal{G}_{\nu_{k,j}}(\alpha_0, \beta_0), \quad \forall j = 1, \ldots, n, \tag{11}$$

where α_0, β_0 are chosen constants.

The employed ARD principle should suppress the noisy parts of response functions which should lead to clearer response functions and subsequently to clearer TACs.

3.3. Wishart Prior. So far, we have modeled only the first or the second diagonal of the precision matrix Υ_k. Each of these approaches has its advantages which we would like to generalize into estimation of several diagonals of the prior covariance matrix. However, this is difficult to solve analytically. Instead, we note that it is possible to create the model for the full prior covariance matrix of the response functions as well as their mutual interactions. For this task, we use vectorized form of response functions denoted as $\mathbf{u} \in \mathbf{R}^{nr\times 1}$, $\mathbf{u} = \mathrm{vec}(U) = [\mathbf{u}_1^T, \ldots, \mathbf{u}_r^T]^T$. This rearranging allows us to model mutual correlation between response functions. The full covariance matrix $\Upsilon \in \mathbf{R}^{nr\times nr}$ can be modeled as follows:

$$f(\mathbf{u} \mid \Upsilon) = t\mathcal{N}_{\mathbf{u}}\left(\mathbf{0}_{nr,1}, \Upsilon^{-1}\right), \qquad f(\Upsilon) = \mathcal{W}_{\Upsilon}(\alpha_0 I_{nr}, \beta_0), \tag{12}$$

where $\mathcal{W}$ is the Wishart distribution with parameters α_0, β_0. See Appendix A.2.

The advantage of this parameterization is obvious: the full covariance matrix is estimated. The disadvantage in this model is that, for estimation of nr parameters in vector $\mathbf{u}$, we need to estimate n^2r^2 additional parameters in covariance structure. The problem is regularized by the prior on Υ, which is relatively weak regularization with potential side effects. We try to suppress these side effects in the next section.

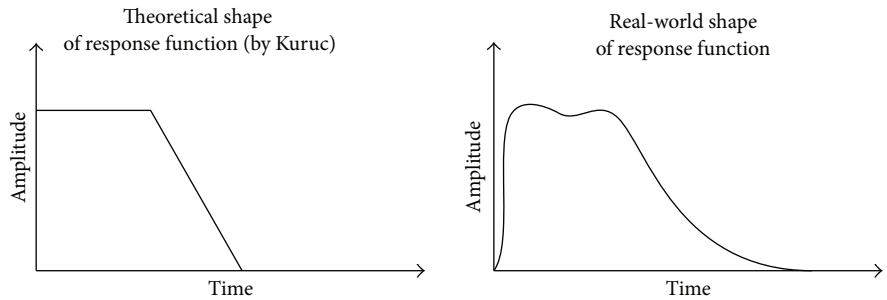


Figure 2: Example of theoretical shape of response function (by [6]), left, and corresponding real-world shape of convolution kernels, right.

3.4. Variational Bayes Approximate Solution. The whole probabilistic model comprises (6)-(7), (8), and (9) and selected response functions model from Sections 3.1–3.3. The probabilistic model is solved using variational Bayes (VB) method. Here, the solution is found in the form of probability densities of the same type of the priors. The shaping parameters of the posterior densities form a set of implicit equations, Appendix B, which is typically analytically intractable and has to be solved iteratively.

The algorithms are summarized in Algorithm 1. We named our algorithms as Nonparametric Variational Bayes Approximation (NVBA) algorithm. All prior parameters are set to 10^{-10} or 10^{+10} in order to obtain noninformative priors. The initial response functions are selected as pulses with different lengths with respect to covering the typical structures while the initial input function is selected as an exponential curve since the iterative solution could converge only to a local minimum.

Algorithm 1 (iterative NVBA algorithm).

(1) Initialization:

(a) Set prior parameters $\alpha_0, \beta_0, \vartheta_0, \rho_0, \phi_0, \psi_0, \zeta_0, \eta_0$.

(b) Set initial values for $\widehat{A}, \widehat{A^T A}, \widehat{\Xi}, \widehat{\mathbf{u}}, \widehat{\mathbf{u}^T \mathbf{u}}, \widehat{\Upsilon}, \widehat{\mathbf{b}}, \widehat{\mathbf{b}^T \mathbf{b}}, \widehat{\zeta}, \widehat{\omega}$.

(c) Set the initial number of sources $r_{\max}$.

(2) Iterate until convergence is reached using computation of shaping parameters from Appendix B:

(a) Source images $\mu_{\bar{a}_i}$, $\Sigma_{\bar{a}_i}$ and their variances $\phi_i, \psi_i, \forall i$ using (B.1)–(B.4).

(b) Response functions $\mu_{\mathbf{u}}$, $\Sigma_{\mathbf{u}}$ and their hyperparameters depending on version of the prior:

(i) OrthogonalRF: (B.11) and (B.12),

(ii) SparseRF: (B.11) and (B.13),

(iii) WishartRF: (B.11) and (B.14).

(c) Input function $\mu_{\mathbf{b}}$, $\Sigma_{\mathbf{b}}$ and its variance ζ, η using (B.5)–(B.7).

(d) Variance of noise ϑ, ρ using (B.8)–(B.9).

(3) Report estimates of source images $\widehat{A}$, response functions $\widehat{U}$, and input function $\widehat{\mathbf{b}}$.

4. Experiments and Discussion

We proposed three versions of model of nonparametric response functions within the model of probabilistic blind source separation model in Sections 3.1–3.3. The proposed algorithms are tested on simulated phantom study as well as on representative clinical data set from dynamic renal scintigraphy.

4.1. Synthetic Dataset. Performance of the proposed models of response functions is first studied on a synthetic dataset generated according to model (5). The size of each image is 50×50 pixels and the number of simulated time points is $n = 50$. We simulate 3 sources which are given in Figure 3, top row, using their source images and response functions together with generated input function $\mathbf{b}$ (top row, right). We generate homogeneous Gaussian noise with standard deviation 0.3 of the signal strength.

The results of the three proposed models are given in Figure 3 in the rowwise schema. Note that all algorithms are capable of estimating the correct number of sources. It can be seen that all methods estimated the source images correctly. The main differences are in estimated response functions, the fourth to the sixth columns, and estimated input function, the seventh column. Note that only the first prior, orthogonal, was not able to respect the sparse character of the modeled response functions; all other priors were able to do so.

4.2. Competing Methods. There are some other methods which provide solution to estimate the response functions in a nonparametric fashion as well.

(1) *FIR Filter.* A semiparametric approach based on finite impulse response filter (FIR Filter) is used to model the haemodynamic response functions [9].

(2) *S-BSS-vecDC.* The sparse blind source separation and vectorized deconvolution (S-BSS-vecDC) algorithm is used in hierarchical models [10].

Quality of estimation of the proposed methods is validated with quantitative results using mean square error (MSE). Here the MSE (μ^k_{MSE}) is computed between the

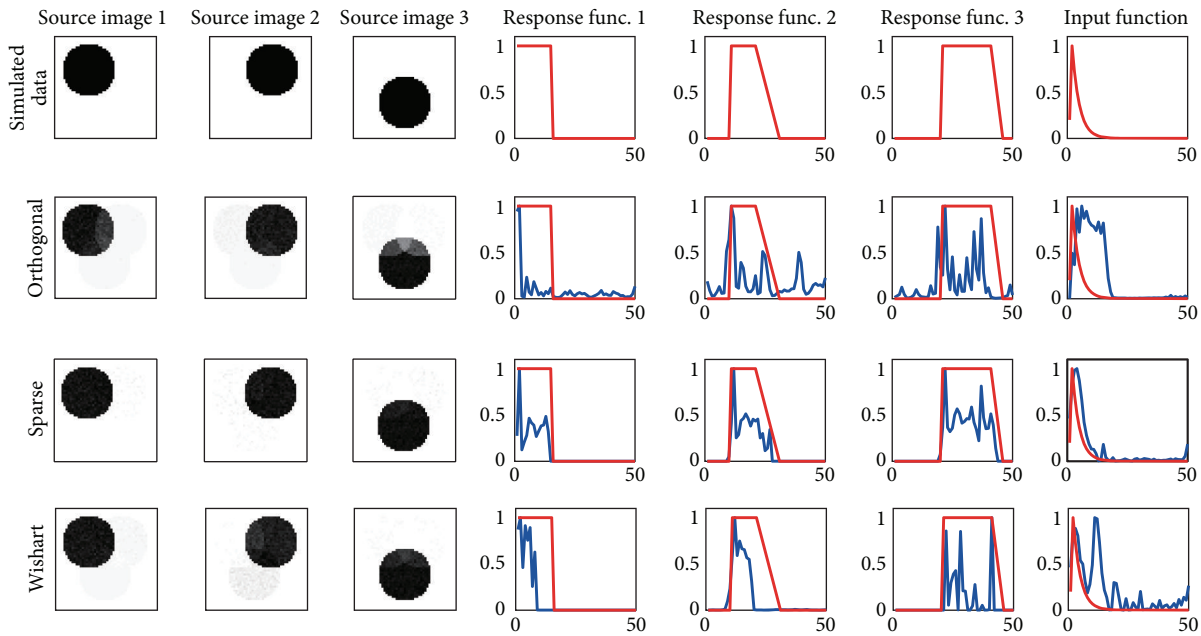


FIGURE 3: The results of the three studied methods on synthetic dataset (the first row). The red lines are generated data while the blue lines are estimated results from the respective methods.

TABLE 1: Comparison of MSE of response functions with different sets of images.

Algorithm	k	μ^k_{MSE}
FIR filter	1	0.044
	2	0.0214
	3	0.115
	4	0.234
S-BSS-vecDC	1	0.0057
	2	0.0021
	3	0.2401
	4	0.4722
NVBA with Wishart prior	1	0.003
	2	0.0016
	3	0.0077
	4	0.0056

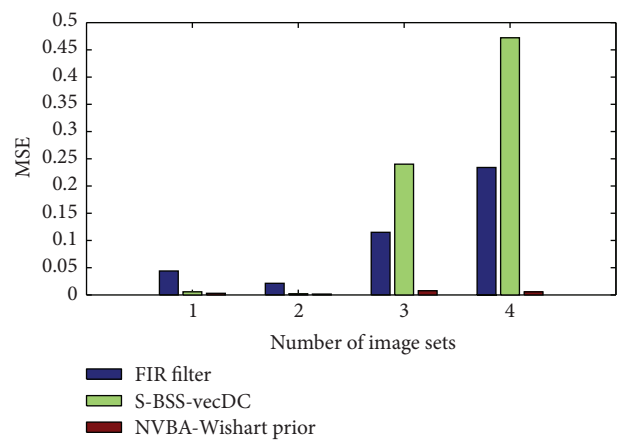


FIGURE 4: Comparison of MSE of response functions with different sets of images.

estimated response functions $\widehat{U}_j$ and their simulated values U_j^{sim}:

$$\mu^k_{\text{MSE}} = \frac{1}{n}\sum_{j=1}^{n}\left(\widehat{U}_{j,k} - U^{\text{sim}}_{j,k}\right)^2. \tag{13}$$

$k = 1, \ldots, 4$ is the set number of testing image. We compare the estimation results of the FIR filter, S-BSS-vecDC, and our NVBA with Wishart Prior with 4 sets of images [18]. Figure 4 gives the bar figure of Table 1.

For the four sets of images, the proposed NVBA with Wishart prior algorithm provided the best estimate of the response function (in terms of the MSE).

4.3. Datasets from Dynamic Renal Scintigraphy. The methods from Sections 3.1–3.3 were tested on real data from dynamic renal scintigraphy taken from online database (http://www.dynamicrenalstudy.org/). We illustrate the possible outcome of the method on two distinct datasets, numbers 84 and 42. Each dataset represents different behavior of the methods.

Both sequences consist of 50 frames taken after 10 seconds and both were preprocessed by selection region of the left kidney. The data are expected to contain three sources of activity: (i) parenchyma, the outer part of a kidney where the tracer is accumulated at the first, (ii) pelvis, the inner part of a kidney where the accumulation has physiological delay, and (iii) background tissues which is typically active

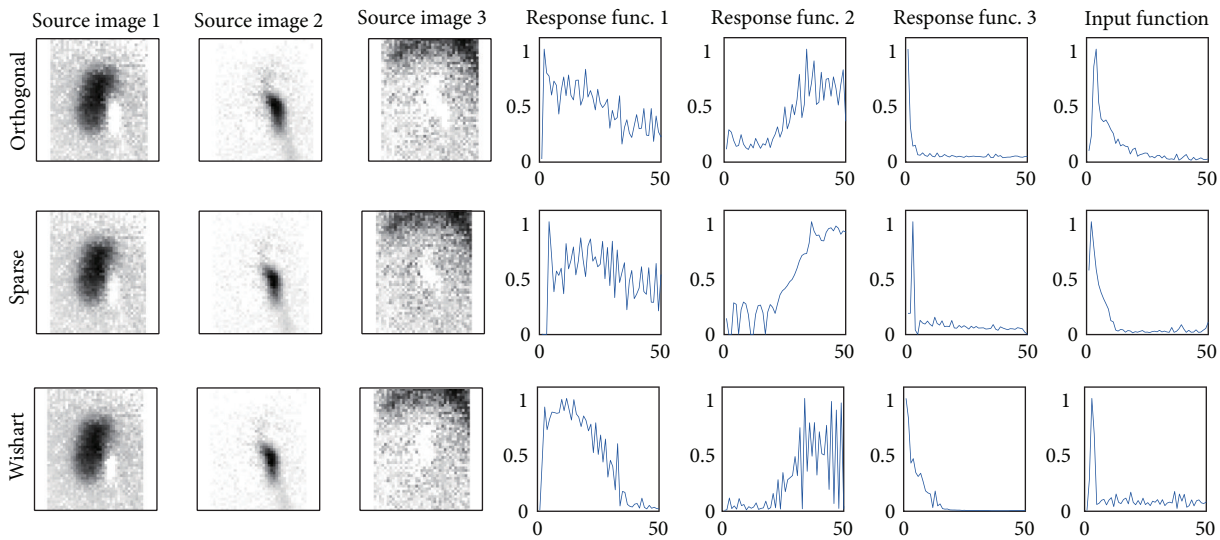


Figure 5: Estimated source images (columns 1–3), response functions (columns 4–6), and input functions (column 7) using priors: orthogonal, sparse, and Wishart.

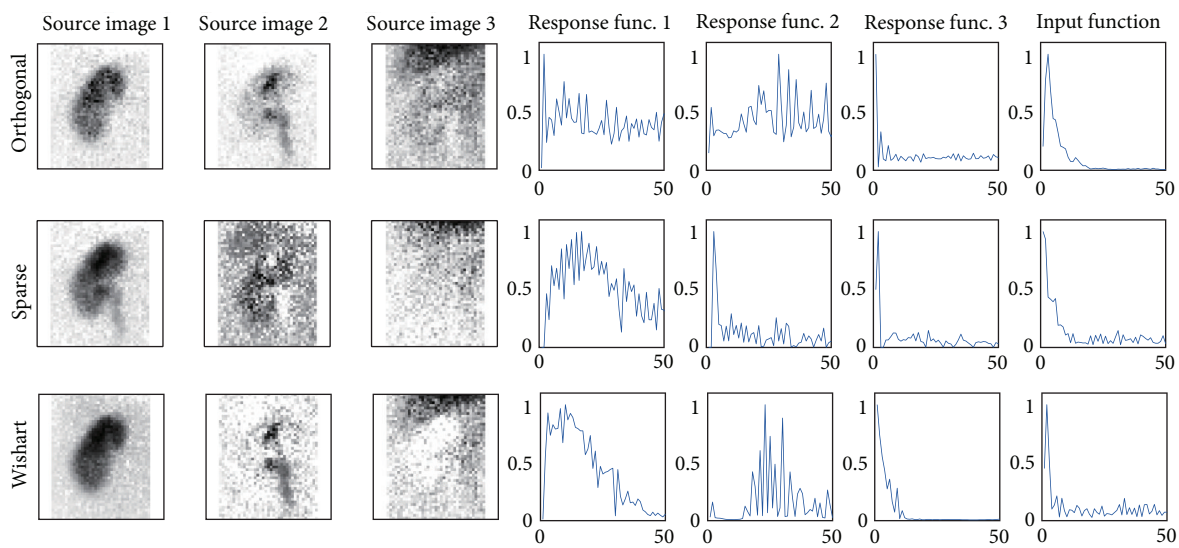


Figure 6: Estimated source images (columns 1–3), response functions (columns 4–6), and input functions (column 7) using priors: orthogonal, sparse, and Wishart.

at the beginning of the sequence. Since the noise in scintigraphy is Poisson distributed, the assumption of homogeneous Gaussian noise (6) can be achieved by asymptotic scaling known as the correspondence analysis [19] which transforms the original data D_{orig} as

$$d_{ij} = \frac{d_{ij,\text{orig}}}{\sqrt{\sum_{i=1}^{p} d_{ij,\text{orig}} \sum_{j=1}^{n} d_{ij,\text{orig}}}}. \tag{14}$$

First, we applied the methods from Sections 3.1–3.3 on dataset number 84 as a typical noncontroversial case. The results are shown in Figure 5 using the estimated source images (columns 1–3), the estimated related response functions (columns 4–6), and the estimated input function (column 7). The results of all three methods are comparable with the main difference being in the smoothness or nonsmoothness of the estimated response functions. This is most remarkable in the fifth column corresponding to the response functions of the pelvis. The sparse prior prefers sparse solution with many zeros, while the Wishart prior models full covariance of response function where no smoothness is incorporated. However, the differences in this case are relatively minor. Second, we apply the methods 3.1–3.3 on dataset number 42 where different methods yield more distinct results; see Figure 6. Note that the sparse priors were not able to separate the pelvis which is mixed with the parenchyma in the first column while the orthogonal prior estimated the source images reasonably; however, the response functions

of the parenchyma and the pelvis are clearly mixed. The Wishart prior was able to separate the parenchyma and the pelvis correctly together with meaningful estimates of their response functions. In this case, the use of more complex prior models significantly outperformed the simpler models. Indeed, the analysis of the full database would be of interest in concrete application; however, it is not a goal of this paper.

5. Conclusion

A common model in functional analysis of dynamic image sequences assumes that the observed images arise from superposition of the original source images weighed by their time-activity curves. Each time-activity curve is assumed to be a result of common input function and source-specific response function, both unknown. Estimation of the model parameters yields an algorithm for blind source separation and deconvolution. The focus of this study is the prior model of the response functions while the models of the source images and the input function are the same. We propose three prior models of the response functions. The advantage of all three models is their flexibility in estimation of various shapes of response functions since we do not impose any parametric form of them. The formulated probabilistic models in the form of hierarchical priors are solved using the variational Bayes methodology. The performance of the proposed methods is tested on simulated dataset as well as on representative real datasets from dynamic renal scintigraphy. It is shown that the behaviors of the methods well correspond with their prior expectations. We compared our algorithm with the other competing methods, and our method achieved the most accurate result. We conclude that the most complex model, that is, the Wishart model, provides also the most desirable results in the sense of mean square errors to the original simulated data as well as in sense of biologically meaningfulness of the results on the real datasets. Notably, the methods have no domain-specific assumptions; hence, they can be used in other tasks in dynamic medical imaging.

Appendix

A. Required Probability Distributions

A.1. Truncated Normal Distribution. Truncated normal distribution, denoted as $t\mathcal{N}$, of a scalar variable x on interval $[a;b]$ is defined as

$$t\mathcal{N}\left(\mu,\sigma\left[a,b\right]\right)=\frac{\sqrt{2}\exp\left(\left(x-\mu\right)^{2}\right)}{\sqrt{\pi\sigma}\left(\operatorname{erf}\left(\beta\right)-\operatorname{erf}\left(\alpha\right)\right)}\chi_{[a,b]}\left(x\right), \quad \text{(A.1)}$$

where $\alpha=(a-\mu)/\sqrt{2\sigma}$, $\beta=(b-\mu)/\sqrt{2\sigma}$, function $\chi_{[a,b]}(x)$ is a characteristic function of interval $[a,b]$ defined as $\chi_{[a,b]}(x)=1$ if $x\in[a,b]$, and $\chi_{[a,b]}(x)=0$ otherwise. erf() is the error function defined as $\operatorname{erf}(t)=(2/\sqrt{\pi})\int_0^t e^{-u^2}du$.

The moments of truncated normal distribution are

$$\begin{aligned}\widehat{x}&=\mu-\sqrt{\sigma}\frac{\sqrt{2}\left[\exp\left(-\beta^{2}\right)-\exp\left(-\alpha^{2}\right)\right]}{\sqrt{\pi\sigma}\left(\operatorname{erf}\left(\beta\right)-\operatorname{erf}\left(\alpha\right)\right)},\\ \widehat{x^{2}}&=\sigma+\mu\widehat{x}-\sqrt{\sigma}\frac{\sqrt{2}\left[b\exp\left(-\beta^{2}\right)-a\exp\left(-\alpha^{2}\right)\right]}{\sqrt{\pi}\left(\operatorname{erf}\left(\beta\right)-\operatorname{erf}\left(\alpha\right)\right)}.\end{aligned} \quad \text{(A.2)}$$

A.2. Wishart Distribution. Wishart distribution $\mathcal{W}$ of the positive-definite matrix $X\in\mathbf{R}^{p\times p}$ is defined as

$$\begin{aligned}\mathcal{W}_{p}\left(\Sigma,\nu\right)&=\left|X\right|^{(\nu-p-1)/2}2^{-\nu p/2}\left|\Sigma\right|^{-\nu/2}\Gamma_{p}^{-1}\left(\frac{\nu}{2}\right)\\ &\quad\cdot\exp\left(-\frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}X\right)\right),\end{aligned} \quad \text{(A.3)}$$

where $\Gamma_p(\nu/2)$ is the Gamma function. The required moment is

$$\widehat{X}=\nu\Sigma. \quad \text{(A.4)}$$

B. Shaping Parameters of Posteriors

Shaping parameters of posterior distributions are given as

$$\Sigma_{\bar{\mathrm{a}}_i}=\left(\widehat{\omega}\sum_{j=1}^{n}\left(\widehat{\overline{x}_j^T\overline{x}_j}\right)+\operatorname{diag}\left(\widehat{\Xi}_i\right)\right)^{-1}, \quad \text{(B.1)}$$

$$\mu_{\bar{\mathrm{a}}_i}=\Sigma_{\bar{\mathrm{a}}_i}\widehat{\omega}\sum_{j=1}^{n}\left(\widehat{\overline{x}}_j d_{i,j}\right), \quad \text{(B.2)}$$

$$\phi_i=\phi_{i,0}+\frac{1}{2}\mathbf{1}_{r,1}, \quad \text{(B.3)}$$

$$\psi_i=\psi_{i,0}+\frac{1}{2}\operatorname{diag}\left(\widehat{\overline{\mathrm{a}}_i^T\overline{\mathrm{a}}_i}\right), \quad \text{(B.4)}$$

$$\Sigma_{\mathbf{b}}=\left(\widehat{\varsigma}I_n+\omega\sum_{i,j=1}^{r}\left(\widehat{\overline{\mathrm{a}}_j^T\overline{\mathrm{a}}_j}\right)\left(\sum_{k,l=0}^{n-1}\Delta_k^T\Delta_l\widehat{u_{k+1,j}u_{l+1,i}}\right)\right)^{-1}, \quad \text{(B.5)}$$

$$\mu_{\mathbf{b}}=\Sigma_{\mathbf{b}}\widehat{\omega}\sum_{k=1}^{r}\left(\sum_{j=0}^{n-1}\Delta_j\widehat{u_{j+1,k}}\right)^{T}D^T\widehat{\mathrm{a}_k}, \quad \text{(B.6)}$$

$$\begin{aligned}\zeta&=\zeta_0+\frac{n}{2},\\ \eta&=\eta_0+\frac{1}{2}\operatorname{tr}\left(\widehat{\mathbf{b}^T\mathbf{b}}\right),\end{aligned} \quad \text{(B.7)}$$

$$\vartheta=\vartheta_0+\frac{np}{2}, \quad \text{(B.8)}$$

$$\begin{aligned}\rho&=\rho_0+\frac{1}{2}\operatorname{tr}\left(DD^T-\widehat{A}\widehat{D}^TD^T-D\widehat{X}^T\widehat{A}^T\right)\\ &\quad+\frac{1}{2}\operatorname{tr}\left(\widehat{A^TA}\widehat{X^TX}\right).\end{aligned} \quad \text{(B.9)}$$

Here, $\widehat{x}$ denotes a moment of respective distribution, tr() denotes a trace of argument, diag() denotes a square matrix with argument vector on diagonal and zeros otherwise or a vector composed from diagonal element of argument matrix, and $\mathbf{1}_{n,1}$ denotes the matrix of ones of dimension $n \times 1$; the auxiliary matrix $\Delta_k \in \mathbf{R}^{n\times n}$ is defined as

$$(\Delta_k)_{i,j} = \begin{cases} 1, & \text{if } i - j = k \\ 0, & \text{otherwise,} \end{cases} \tag{B.10}$$

and standard moments of required probability distributions are given in Appendices A.1 and A.2.

The shaping parameters for response functions are given in the following subsections while the parameter $\mu_{\mathbf{u}}$ is common for all methods as

$$\begin{aligned} &\mu_{\mathbf{u}} \\ &= \Sigma_{\mathbf{u}} \left(\widehat{A^T A} \otimes \widehat{\omega} \widehat{B^T B} \right) \operatorname{vec} \left(\widehat{B^T B}^{-1} \widehat{B}^T D^T \widehat{A} \widehat{A^T A}^{-1} \right). \end{aligned} \tag{B.11}$$

B.1. Shaping Parameters for Orthogonal Prior. Consider

$$\begin{aligned} \Sigma_{\mathbf{u}} &= \left(\widehat{A^T A} \otimes \widehat{\omega} \widehat{B^T B} + I_n \otimes \widehat{\Upsilon} \right)^{-1}, \\ \alpha_k &= \alpha_{k,0} + \frac{n}{2}, \\ \beta_k &= \beta_{k,0} + \frac{1}{2} \operatorname{tr} \left(\mathbf{u}_k \mathbf{u}_k^T \right). \end{aligned} \tag{B.12}$$

B.2. Shaping Parameters for Sparse Prior. Consider

$$\begin{aligned} \Sigma_{\mathbf{u}} &= \left(\widehat{A^T A} \otimes \widehat{\omega} \widehat{B^T B} + \operatorname{diag} \left(\operatorname{vec} \left(\widehat{\Upsilon} \right) \right) \right)^{-1}, \\ \alpha &= \alpha_0 + \frac{1}{2} \mathbf{1}_{nr,1}, \\ \beta &= \beta_0 + \frac{1}{2} \operatorname{diag} \left(\widehat{\mathbf{u}\mathbf{u}^T} \right). \end{aligned} \tag{B.13}$$

B.3. Shaping Parameters for Wishart Prior. Consider

$$\begin{aligned} \Sigma_{\mathbf{u}} &= \left(\widehat{A^T A} \otimes \widehat{\omega} \widehat{B^T B} + \widehat{\Upsilon} \right)^{-1}, \\ \Sigma_{\Upsilon} &= \left(\widehat{\mathbf{u}\mathbf{u}^T} + (\alpha_0 I_{nr})^{-1} \right)^{-1}, \\ \beta &= \beta_0 + 1. \end{aligned} \tag{B.14}$$

Competing Interests

The author declares that he has no competing interests.

Acknowledgments

This research work was supported by the National Natural Science Foundation of China (Grants nos. 61473047, 61271262, and 61201407) and Undergraduate Training Program for Innovation and Entrepreneurship (Grant no. 201610710143).

References

[1] A. L. Martel, A. R. Moody, S. J. Allder, G. S. Delay, and P. S. Morgan, "Extracting parametric images from dynamic contrast-enhanced MRI studies of the brain using factor analysis," *Medical Image Analysis*, vol. 5, no. 1, pp. 29–39, 2001.

[2] J. S. Fleming and P. M. Kemp, "A comparison of deconvolution and the Patlak-Rutland plot in renography analysis," *Journal of Nuclear Medicine*, vol. 40, no. 9, pp. 1503–1507, 1999.

[3] T. Taxt, R. Jiřík, C. B. Rygh et al., "Single-channel blind estimation of arterial input function and tissue impulse response in DCE-MRI," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 1012–1021, 2012.

[4] E. Durand, M. D. Blaufox, K. E. Britton et al., "International Scientific Committee of Radionuclides in Nephrourology (ISCORN) consensus on renal transit time measurements," *Seminars in Nuclear Medicine*, vol. 38, no. 1, pp. 82–102, 2008.

[5] L. Chen, P. L. Choyke, T.-H. Chan, C.-Y. Chi, G. Wang, and Y. Wang, "Tissue-specific compartmental analysis for dynamic contrast-enhanced MR imaging of complex tumors," *IEEE Transactions on Medical Imaging*, vol. 30, no. 12, pp. 2044–2058, 2011.

[6] A. Kuruc, W. J. Caldicott, and S. Treves, "An improved deconvolution technique for the calculation of renal retention functions," *Computers and Biomedical Research*, vol. 15, no. 1, pp. 46–56, 1982.

[7] O. Tichý, V. Šmídl, and M. Šámal, "Model-based extraction of input and organ functions in dynamic scintigraphic imaging," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 4, no. 3-4, pp. 135–145, 2014.

[8] J. Kershaw, S. Abe, K. Kashikura, X. Zhang, and I. Kanno, "A bayesian approach to estimating the haemodynamic response function in event-related fMRI," *NeuroImage*, vol. 11, no. 5, p. S474, 2000.

[9] C. Goutte, F. Å. Nielsen, and L. K. Hansen, "Modeling the haemodynamic response in fMRI using smooth FIR filters," *IEEE Transactions on Medical Imaging*, vol. 19, no. 12, pp. 1188–1201, 2000.

[10] O. Tichy and V. Šmídl, "Bayesian blind separation and deconvolution of dynamic image sequences using sparsity priors," *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 258–266, 2015.

[11] J. Miskin, *Ensemble learning for independent component analysis [Ph.D. thesis]*, University of Cambridge, 2000.

[12] M. W. Woolrich, "Bayesian inference in FMRI," *NeuroImage*, vol. 62, no. 2, pp. 801–810, 2012.

[13] D. M. Steinberg, O. Pizarro, and S. B. Williams, "Hierarchical bayesian models for unsupervised scene understanding," *Computer Vision and Image Understanding*, vol. 131, pp. 128–144, 2015.

[14] V. Šmidl and A. Quinn, *The Variational Bayes Method in Signal Processing*, Springer, Berlin, Germany, 2006.

[15] B. L. Diffey, F. M. Hall, and J. R. Corfield, "The 99mTc DTPA dynamic renal scan with deconvolution analysis," *Journal of Nuclear Medicine*, vol. 17, no. 5, pp. 352–355, 1976.

[16] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.

[17] C. Bishop and M. Tipping, "Variational relevance vector machines," in *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pp. 46–53, Stanford, Calif, USA, June-July 2000.

[18] Database of Dynamic Renal Scintigraphy, http://www.dynamicrenalstudy.org.

[19] H. Benali, I. Buvat, F. Frouin, J. P. Bazin, and R. Di Paola, "A statistical model for the determination of the optimal metric in factor analysis of medical image sequences (FAMIS)," *Physics in Medicine and Biology*, vol. 38, no. 8, pp. 1065–1080, 1993.