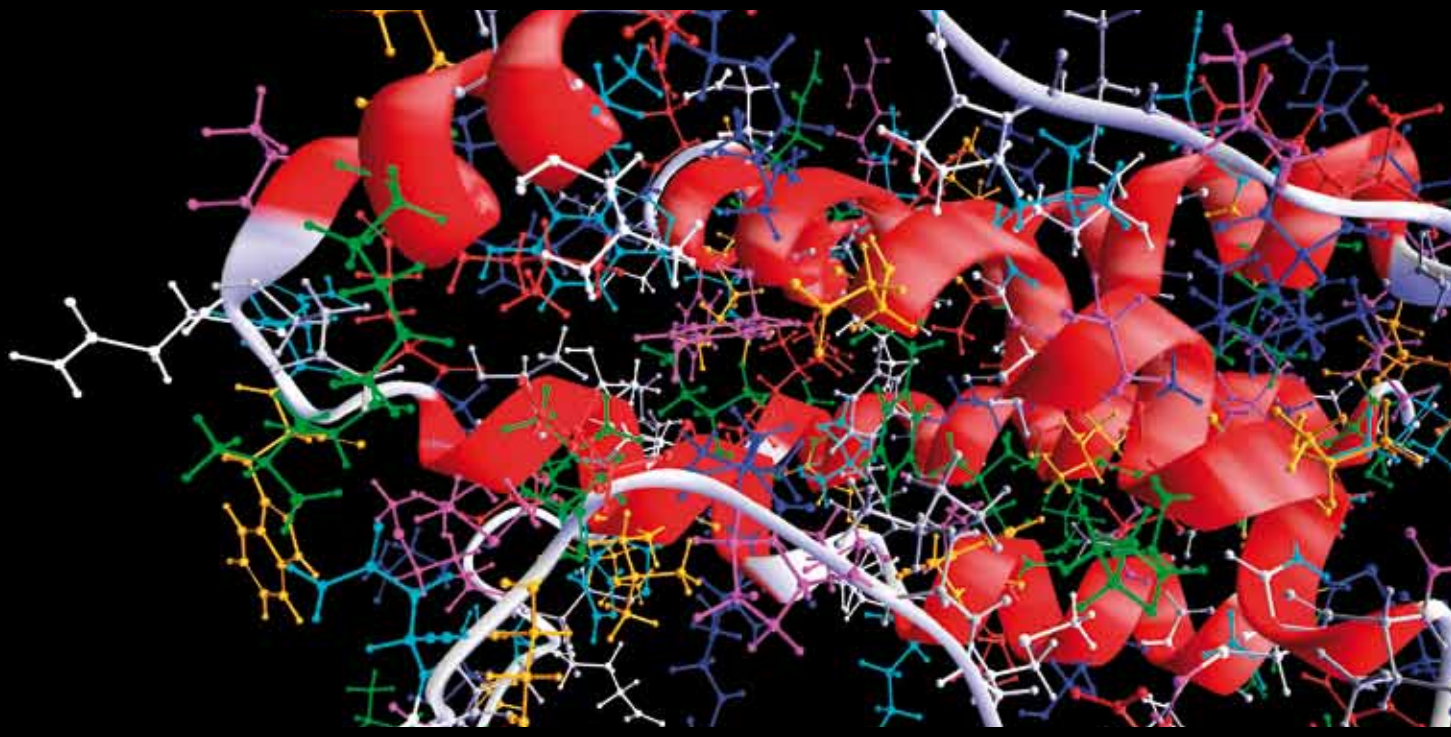
Computational and Mathematical Methods in Medicine

Advanced Medical Image Analysis

Guest Editors: Rong Chen, Zhongqiu Wang, and Yuanjie Zheng



Advanced Medical Image Analysis

Computational and Mathematical Methods in Medicine

Advanced Medical Image Analysis

Guest Editors: Rong Chen, Zhongqiu Wang, and Yuanjie Zheng

Editorial Board

Contents

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 975372, 1 page
http://dx.doi.org/10.1155/2014/975372

Editorial

Advanced Medical Image Analysis

Rong Chen,[1] Zhongqiu Wang,[2] and Yuanjie Zheng[3]

[1] *Department of Radiology, University of Maryland School of Medicine, 22 S. Greene Street, Baltimore, MD 21201, USA*
[2] *Department of Radiology, Affiliated Hospital of Nanjing University of Traditional Chinese Medicine, 155 Hanzhong Road, Nanjing, Jiangsu 210029, China*
[3] *Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA*

Correspondence should be addressed to Rong Chen; rong.chen.mail@gmail.com

Received 9 June 2014; Accepted 9 June 2014; Published 1 July 2014

Medical image analysis is performed in order to facilitate medical research and ultimately provide better healthcare. It is critical to the advancement of imaging-based medical research, for example, using magnetic resonance (MR) imaging to probe brain structural and functional changes related to a disease or cognitive process.

This special issue highlights new methods of signal and image processing, computer vision, machine learning, and statistical analysis and their application in medical image analysis. It is organized into two groups of papers: predictive modeling and image processing.

In the first group, we include a set of papers for imaging-derived biomarker detection and clinical decision support systems. The overall theme is how to detect biomarkers characterizing a disease and to build decision support systems using computer algorithms. B. Tay et al. propose a classification scheme for identifying healthy individuals and patients with spinal cord injuries based on fractional anisotropy values obtained from diffusion tensor imaging data. H. Liu et al. address cirrhosis classification based on multimodality imaging and propose a new method to extract texture features for multilabel classification (normal, early, middle, and advanced stages). G.-P. Liu et al. employ deep learning and multilabel learning to construct the syndrome diagnostic model for chronic gastritis. M. Yang et al. describe a method to automatically detect corticospinal tract damage in chronic stroke patients and demonstrate that the detected biomarker is associated with motor impairment. Y. Liu et al. describe a novel machine learning algorithm which combines tract-based spatial statistics and Bayesian data mining to quantify white matter changes in mild traumatic brain injury.

The second group of papers consists of a wide range of medical image processing algorithms, including segmentation, visualization, and enhancement. For brain tissue segmentation, S. Ji et al. describe a multistage method based on superpixel and fuzzy clustering for brain MRI segmentation. Y. Zhong et al. propose a method for regularization of fMRI data to address the limitations of traditional spatial independent component analysis. J. Gao et al. describe a global search algorithm based method to separate MR images blindly. Y. P. Du et al. develop a method to reduce partial volume effect with voxel-shifted interpolation and this algorithm substantially improves the detection of BOLD signal in fMRI. Z. Jin et al. use a high-pass filter based method to enhance the visibility of the venous vasculature and reduce the artifacts in the venography. A framework for tracking left ventricular endocardium through 2D echocardiography image sequence is proposed by H. Ketout and J. Gu. E. Bengtsson and P. Malm review automated analysis of the cell samples to screen for cervical cancer.

This special issue is selective. Among 27 submissions, 12 were selected. It is our hope that this impressive group of papers will help the medical image analysis community in their efforts to advance imaging-based medical research.

Rong Chen
Zhongqiu Wang
Yuanjie Zheng

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 812785, 9 pages
http://dx.doi.org/10.1155/2014/812785

Research Article

Background-Suppressed MR Venography of the Brain Using Magnitude Data: A High-Pass Filtering Approach

Zhaoyang Jin,[1,2,3] Ling Xia,[1,2] Minming Zhang,[4] and Yiping P. Du[1,2]

[1] *Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China*
[2] *Key Laboratory for Biomedical Engineering of Education Ministry of China, Hangzhou, China*
[3] *College of Automation, Hangzhou Dianzi University, Hangzhou, China*
[4] *Department of Radiology, School of Medicine, Second Affiliated Hospital, Zhejiang University, Hangzhou, China*

Correspondence should be addressed to Yiping P. Du; yipingdu@zju.edu.cn

Received 24 October 2013; Revised 10 February 2014; Accepted 14 February 2014; Published 10 June 2014

Academic Editor: Yuanjie Zheng

Conventional susceptibility-weighted imaging (SWI) uses both phase and magnitude data for the enhancement of venous vasculature and, thus, is subject to signal loss in regions with severe field inhomogeneity and in the peripheral regions of the brain in the minimum-intensity projection. The purpose of this study is to enhance the visibility of the venous vasculature and reduce the artifacts in the venography by suppressing the background signal in postprocessing. A high-pass filter with an inverted Hamming window or an inverted Fermi window was applied to the Fourier domain of the magnitude images to enhance the visibility of the venous vasculature in the brain after data acquisition. The high-pass filtering approach has the advantages of enhancing the visibility of small veins, diminishing the off-resonance artifact, reducing signal loss in the peripheral regions of the brain in projection, and nearly completely suppressing the background signal. The proposed postprocessing technique is effective for the visualization of small venous vasculature using the magnitude data alone.

1. Introduction

MR venography (MRV) has demonstrated substantial clinical significance in the diagnosis of intracranial vascular diseases, such as venous thrombosis and arteriovenous malformations [1–3]. Flow-sensitive 2D time-of-flight and phase-contrast techniques have been commonly used for intracranial MRV [1–3]. MRV with administration of contrast media has been used for the visualization of dural sinuses and small veins with low blood flow and diagnoses of cerebral vasoocclusive diseases [4–7]. MR susceptibility-weighted imaging (SWI) has been demonstrated to be an excellent technique in the visualization of the cerebral venous vasculature without exogenous contrast media [8–11]. The susceptibility difference between deoxygenated hemoglobin in venous blood and brain parenchyma and the field inhomogeneity within the surrounding tissue around the veins result in partial volume signal cancellation in a voxel and generate a phase shift in the voxel [12, 13]. SWI data have usually been acquired using a 3D spoiled gradient-echo (SPGR) pulse sequence with a relatively long echo time (TE). In conventional SWI processing, both signal reduction and phase shift are used for the display of venous vasculature. Excellent venous contrast can be obtained by combining the high-pass filtered phase data with the magnitude data.

Current SWI methods either use a homodyne filtering algorithm or a combination of phase unwrapping and high-pass filtering [14, 15]. These methods, however, are subject to the off-resonance artifact in regions with severe field inhomogeneity because of the incomplete suppression of background phase in the high-pass filtered phase images in these regions [8, 10, 16, 17]. This artifact can be reduced by increasing the filter size in both methods [14–17], which may have the drawback of reducing the contrast of small veins [17]. Signal loss is also commonly observed in the peripheral regions of the brain in the minimum-intensity projection (mIP) of the 3D SWI data [17]. In addition, the contrast of veins in the high-pass filtered phase images is highly dependent on the orientation of the veins relative to the main field, voxel size, and voxel aspect ratio [12, 13].

Several techniques have been developed to reduce the off-resonance artifact in regions with severe field inhomogeneity in SWI [17]. These techniques include phase unwrapping [14, 15], a regional suppression of phase artifact using local field gradient mapping [17], and the use of a susceptibility model [18]. Signal loss in the peripheral region of the brain was reduced by the use of tissue-air volume segmentation algorithms [17, 19–21]. Despite the effectiveness of these techniques, residual image artifacts often remain in SWI.

In this study, we present a background-suppressed MRV (BS-MRV) technique for the visualization of the cerebral venous vasculature using 3D SPGR magnitude data acquired without exogenous contrast media. In this technique, visualization of the venous vasculature in the magnitude data is enhanced by background suppression through postprocessing. This technique has a major advantage of simplicity since only the magnitude data are used. This eliminates the need for phase data and the possible technical challenges in dealing with phase errors and phase unwrapping. The BS-MRV technique has the advantages of (1) the enhanced visibility of small veins, (2) diminished off-resonance artifacts, (3) reduced signal loss in the peripheral regions of the brain in the mIP display, and (4) a nearly completely suppressed background signal.

2. Materials and Methods

2.1. Background Suppression Using a High-Pass Filter. In this study, a high-pass (HP) filtering technique was applied in the Fourier domain of the magnitude data to enhance the signal of venous vasculature through background suppression. The implementation of the BS-MRV technique includes the following steps. (1) Form a 3D complex image dataset with the real component filled with the magnitude image data and the imaginary component filled with zeroes. (2) Apply Fourier transform of the 3D complex data. (3) Apply a 2D HP filter to the k_x-k_y plane for each partition of the 3D data for background suppression, where k_x is the readout direction and k_y is the phase encoding direction. Because a 2D HP filter is applied to the phase data along the in-plane directions in conventional SWI, we decided to use the same 2D HP filtering approach in this study. (4) Apply an inverse Fourier transform to obtain a 3D complex dataset in the image domain. (5) Take the real component of the 3D complex dataset to form a 3D HP-filtered dataset, $I_{\mathrm{HP}}(x, y, z)$. Two types of HP filters were studied for background suppression in Step 3, namely, an inverted Hamming (iH) and an inverted Fermi (iF) filter. The iH filter, with a full size of $2H_x$ and $2H_y$, was applied at the center of k-space:

$$\mathrm{iH}\left(n_x, n_y\right) = \begin{cases} 0.46\left[1.0 - \cos\left(\pi\sqrt{\dfrac{n_x^2}{H_x^2} + \dfrac{n_y^2}{H_y^2}}\right)\right], \\ \qquad\qquad \text{for } \dfrac{n_x^2}{H_x^2} + \dfrac{n_y^2}{H_y^2} \le 1, \\ 1.0, \qquad \text{elsewhere.} \end{cases} \tag{1}$$

The iF filter for square FOV ($N_x = N_y$) is given by

$$\mathrm{iF}\left(n_x, n_y\right) = \begin{cases} 1 - \dfrac{1}{\left[1 + \exp\left(\sqrt{n_x^2 + n_y^2} - R\right)/W\right]}, \\ \qquad\qquad \text{for } \sqrt{n_x^2 + n_y^2} \le R + W, \\ 1.0, \qquad \text{elsewhere,} \end{cases} \tag{2}$$

where $2R$ is the size of the filter (i.e., R is the radius of the filter) and W is the width of the transition. When $N_x \neq N_y$, $2R_x$ and $2R_y$ are the size of the filter along the k_x and k_y directions, respectively. An iF filter is expected to have more effective background suppression than an iH filter with the same size because the former one has a broader stop-band than the latter one. Another Fermi filter was applied to improve SNR and reduce the angular dependency of spatial resolution by suppressing the data on the corners of the rectangular region in k-space [22]. Zero-filled interpolation was applied along both k_x and k_y directions to improve the smoothness of the vessel contrast by reducing the partial volume effect.

2.2. Scaling and Display of BS-MRV Data. The background signal in the $I_{\mathrm{HP}}(x, y, z)$ images is expected to be nearly completely suppressed with intensity close to zero, while the venous vasculature is expected to have negative values because veins have a negative contrast in the original magnitude data. The contrast of the veins in the $I_{\mathrm{HP}}(x, y, z)$ data, however, is dependent on the size of the HP filter. Using a larger filter size would result in reduced vascular signal because of the greater signal suppression near the central region of k-space. In the conventional display of the MRV data, the brightness and contrast of the veins are affected by the voxels with signal at the high and low ends of the intensity range of the data. Instead of using the intensity range of the data, the standard deviation (STD), σ, of the HP-filtered data, $I_{\mathrm{HP}}(x, y, z)$, was used to scale the data, so that the brightness and contrast of the veins have substantially reduced dependency on the voxels with intensity near the maximum and minimum in the data. On the other hand, the brightness of the background is affected by the mean intensity, I_m, in $I_{\mathrm{HP}}(x, y, z)$. For this reason, I_m was subtracted from $I_{\mathrm{HP}}(x, y, z)$ before the scaling to minimize the dependency of the brightness in background on filter size. The calculations of σ and I_m were repeated twice based on the histogram of $I_{\mathrm{HP}}(x, y, z)$ in a selected region of interest in the brain. The initial calculations of σ and I_m were used to exclude the voxels with the low intensity $< I_m - 3\sigma$ and high intensity $> I_m + 3\sigma$ for improved estimation of STD and mean in the refined calculation.

Table 1: The standard deviation and mean intensity of the 3D BR-MRV datasets obtained with the inverted Hamming and Fermi filters ($\times 10^{-5}$). The maximum intensity of the original 3D magnitude dataset prior to the filtering was normalized to 1.0.

Filter size	Inverted Hamming						Inverted Fermi					
	σ_t	σ_a	σ	$I_{m,t}$	$I_{m,a}$	I_m	σ_t	σ_a	σ	$I_{m,t}$	$I_{m,a}$	I_m
12×9	247	68.5	260	147	1.6	24.2	237	68.6	225	30.8	−6.3	−13.4
16×12	240	68.5	241	98	−1.8	7.8	229	68.3	207	21.5	−6.6	−13.4
24×18	227	68.5	216	56	−4.4	—	210	68.4	182	9.2	−6.5	−11.6
32×24	216	68.2	198	37	−5.4	—	196	68.0	169	3.1	−6.5	−10.0
48×36	200	67.6	176	19	−6.2	—	181	66.9	153	−1.0	−6.4	−8.2
64×48	189	67.3	163	11	−6.2	—	171	64.7	141	−2.3	−6.7	−7.1

σ_t: standard deviation of intensity in brain tissue; σ_a: standard deviation of intensity in air; σ: standard deviation of intensity in the entire dataset; $I_{m,t}$: mean of intensity in brain tissue; $I_{m,a}$: mean of intensity in air; I_m: mean of intensity in the entire dataset.

The BS-MRV data, $I_{\text{BS-MRV}}(x, y, z)$, was generated using the following linear scaling procedure to reduce the dependency of the display vessel contrast on filter size:

$$I_{\text{BS-MRV}}(x, y, z) = \frac{[I_{\text{HP}}(x, y) - I_m]}{\sigma},$$
$$\text{if } -\eta\sigma < [I_{\text{HP}}(x, y, z) - I_m] < 0,$$
$$I_{\text{BS-MRV}}(x, y, z) = -\eta, \quad \text{if } [I_{\text{HP}}(x, y, z) - I_m] < -\eta\sigma, \tag{3}$$
$$I_{\text{BS-MRV}}(x, y, z) = 0, \quad \text{if } [I_{\text{HP}}(x, y, z) - I_m] > 0.$$

The venous vasculature has a negative contrast in $I_{\text{BS-MRV}}(x, y, z)$ with an intensity range from $-\eta$ to 0.

2.3. Data Acquisition. The MRV data were acquired on a GE 3T scanner (Milwaukee, WI, USA) in a volume slightly above the circle of Willis. Studies were performed on six healthy subjects who provided written informed consent approved by the local Institutional Review Board. High-order shim was applied prior to the scans to improve the overall field inhomogeneity.

The first dataset, with a size of $512 \times 384 \times 64$, was acquired along the transverse direction using a 3D SPGR pulse sequence with a birdcage head coil. The scan parameters were field of view (FOV) = $26 \times 19.5\,\text{cm}^2$, 1.0 mm slice thickness, full echo acquisition, readout bandwidth = ±31.3 kHz, and TE/TR/α = 20 ms/34 ms/20°. The frequency encoding direction had a FOV of 26 cm with 512 sampling points. The scan time was 14.8 minutes. Flow compensation was applied along the readout direction to reduce the phase variation caused by blood flow. The 3D complex SWI data were reconstructed offline using MATLAB (The MathWorks, Inc., Natick, MA, USA).

The second dataset was acquired with five subjects using an eight-channel phased-array RF coil with the same scan parameters as in the first dataset, except for the use of a smaller flip angle of 12 degrees for reduced partial T1 saturation of signal in cerebrospinal fluid (CSF). The magnitude data acquired from each of the coils was combined to form a composite dataset using the square root of the sum of squares.

The third dataset was acquired at a lower resolution, with a size of $384 \times 312 \times 32$, using a 4-echo MR angiography and venography pulse sequence [23] with the same birdcage coil. The MRA data were acquired at the first echo and three volumes of MRV data were acquired at the second, third, and fourth echoes. The FOV was $20 \times 16\,\text{cm}^2$ and the slice thickness was 1.6 mm. All four echoes were acquired with a 66.25% partial echo and a flip angle of 20 degrees. The partial echo data were acquired in a readout window of 8.14 ms with a bandwidth of ±15.6 kHz. The TEs of these four echoes were TE1/TE2/TE3/TE4 = 5.5/19.7/24.8/39.0 ms. The TR was 47 ms and the scan time was 7 minutes, 53 seconds. Flow compensation was applied along the readout and in-plane phase-encoding directions to the first echo to reduce flow artifact. The partial-echo datasets were reconstructed offline using the POCS [24] algorithm.

3. Results

The mean and STD of the HP-filtered 3D images are shown in Table 1. These images were obtained by applying the iH and iF filters with different filter sizes to the 3D SPGR (512×384×64, zero-filled to $1024 \times 768 \times 128$) dataset. The following was observed. (1) The intensity of the filtered datasets had tight distributions with small STDs. For example, the overall STD of the entire 3D imaging volume in the filtered dataset (σ) with iH 32×24 ($2H_x = 32$, $2H_y = 24$) was less than 0.2% of the maximum intensity of the original dataset. (2) The STDs in air, σ_a, were about one-third smaller than that in tissue, σ_t. Thus, the overall STDs of the filtered datasets were dominated by the STDs in tissue. (3) Both σ and σ_t were reduced as the filter size was increased. (4) The mean intensity in the datasets (I_m) was near zero in brain tissue, air, and the entire imaging volume. For example, the mean of the filtered dataset with iH 32×24 was less than 0.01% of the maximum intensity of the original dataset. (5) Both σ and σ_t of an iH filtered dataset were higher than that of an iF filtered dataset with the same filter size.

The percentage of voxels excluded by the second condition in (3), as shown in a function of η in Figure 1, was less than 0.026% when $\eta \geq 6$. Therefore, $\eta = 6$ was selected in this study because it only excludes a very small portion of voxels.

Figure 2 shows the mIP of the STD-scaled BS-MRV datasets, $I_{\text{BS-MRV}}(x, y, z)$, processed with the iH (a) and iF (b) filters at different filter sizes. The venous vasculature was well depicted in these images. The visibility of the veins in

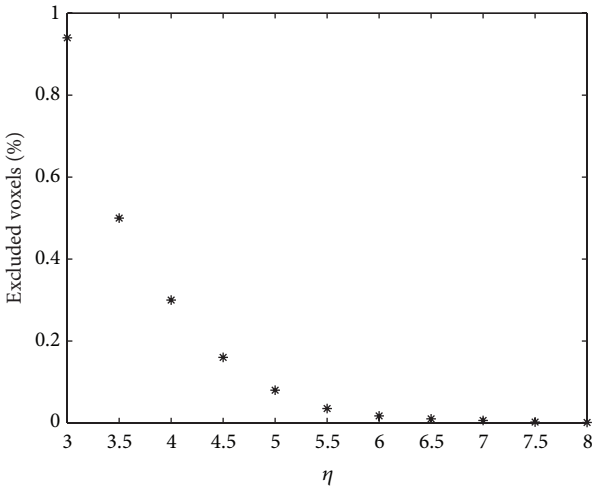


Figure 1: This figure shows the percentage of voxels excluded by the second condition in (3) as a function of η. The excluded voxels were only 0.026% when $\eta = 6$.

the images processed with an iF filter was comparable with that in the images processed with an iH filter. It was observed that the filter size only had a minor effect on the overall image quality. The shade in the peripheral regions of the brain, as indicated by the thick short arrows, and at the third ventricle, as indicated by the thin short arrows, was more conspicuous with smaller filter sizes. The BS-MRV datasets at the middle two columns appear to have similar vessel visibility compared to the other images. Although not shown here, the BS-MRV datasets processed with larger filter sizes than the ones shown in this figure demonstrated gradually increased noise level and decreased venous contrast.

Figure 3 shows the $512 \times 384 \times 50$ (interpolated to $1024 \times 768 \times 100$) dataset, the single slice (top row), and mIP (bottom row) of the magnitude images (a and e), MRV obtained by applying the conventional SWI postprocessing algorithm (homodyne filter size 64×48) (b and f), conventional SWI with a median filter ($5 \times 5 \times 5$) (c and g), and BS-MRV technique with iH filtering ($2H_x \times 2H_y = 32 \times 24$) (d and h). Severe off-resonance artifacts (as indicated by the bright arrows) and signal loss at the peripheral regions of the brain were observed in (b) and (f). Slight off-resonance artifacts (as indicated by the arrows) were also observed in (c) and (g). The visibility of larger veins was reduced in (c) and (g). In contrast, the BS-MRV images in (d) and (h) show good visibility of the venous vasculature, effective background suppression, and the absence of off-resonance artifacts. The veins adjacent to the regions of low signal intensity, such as around the globus pallidus, were largely obscured in the mIP of the magnitude data and SWI. But they were well depicted in the BS-MRV images.

The results of applying the BS-MRV technique to a 4-echo MRAV dataset are shown in Figure 4. The data of Echo 2 (TE = 19.7 ms), Echo 3 (TE = 24.8 ms), and Echo 4 (TE = 39.0 ms) are shown in the left, middle, and right column, respectively. The mIP images of original 3D magnitude data are shown at the top row. The mIP images of the 3D MRV processed by applying the conventional SWI technique to the 3D complex data are shown in the middle row. The bottom row shows the mIP images of the 3D data obtained by applying the BS-MRV technique to the same 3D magnitude data as shown in the top row. The off-resonance artifact introduced by the phase distortion in the frontal region with severe field inhomogeneity in the SWI images, as indicated by an arrow, was absent in the BS-MRV data. The regions with high level of iron deposits, such as the globus pallidus regions, indicated by the dashed circles, did not obscure the visibility of veins in the BS-MRV data either. The veins in the peripheral regions of the brain were obscured in the mIP display of the magnitude data and SWI data but were well depicted in the mIP of the BS-MRV data. In most of the brain regions, the visibility of the veins in the BS-MRV data was much better than that in the magnitude data and was comparable to that in the SWI data, especially at a shorter TE. The results demonstrate that the BS-MRV algorithm can provide adequate venous visibility in a relatively wide range of TE, even in regions of severe field inhomogeneity.

The BS-MRV algorithm can be readily applied to 3D SPGR acquired with a phased-array coil. The composite images were constructed by the squared sum of the 8 data sets acquired from each of the coil elements. Figure 5 shows the mIPs of magnitude images (top row) and MRV data in a subvolume of 18 slices obtained by applying the BS-MRV algorithm to the 3D composite images in five subjects.

4. Discussion

Background suppression is a commonly used approach in MR vascular imaging [15, 25, 26]. In TOF MRA, background signal is saturated by repeated RF excitations in the imaging volume. The contrast of vessels can be further enhanced by the additional saturation of brain tissue using a magnetization transfer technique, in which off-resonance RF pulses with a relatively high power are repeatedly applied [27]. Median filters have been applied in the image domain for the suppression of background tissue in angiography [25] and BOLD venography [26]. In such spatial filtering approaches, the selection of the kernel size of a median filter will affect the visibility of the veins when the size of the kernel is comparable with the size of the veins, as shown in Figures 3(c) and 3(g). As an extension to these approaches of background suppression, the BS-MRV technique was developed to enhance the visibility of small veins through the digital suppression of the background signal by applying HP filters in the Fourier domain in postprocessing. The results shown in this study indicate that background suppression achieved by using a Fourier domain HP filter can effectively enhance the visibility of small veins when the size of the HP filter is in a relatively wide range.

In SWI, the contrast of the veins is obtained primarily from the susceptibility induced phase shift. This phase shift is enhanced in the phase mask by applying an HP filter in the Fourier domain to suppress the background phase. The contrast of the veins is further enhanced by multiplying the phase mask to the magnitude images several times.

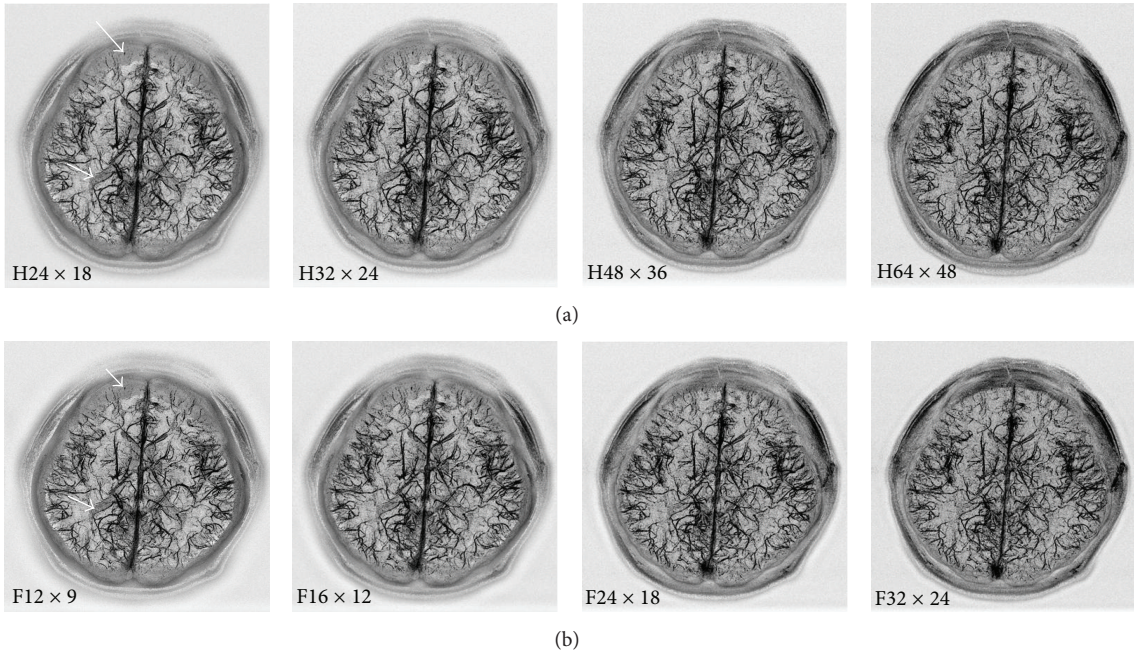


Figure 2: The mIP of the STD-scaled BS-MRV (512 × 384 × 50, interpolated to 1024 × 768 × 100) processed with the iH (top row) and iF (bottom row) filters at different filter sizes. At the top row, the numbers following the label "iH" indicate the filter size $2H_x \times 2H_y$. At the bottom row, the numbers following the label "iF" indicate the filter size $2R_x \times 2R_y$. In some of the peripheral regions of the brain, as indicated by thick long arrows, the visibility of the veins was obscured by the scalp.

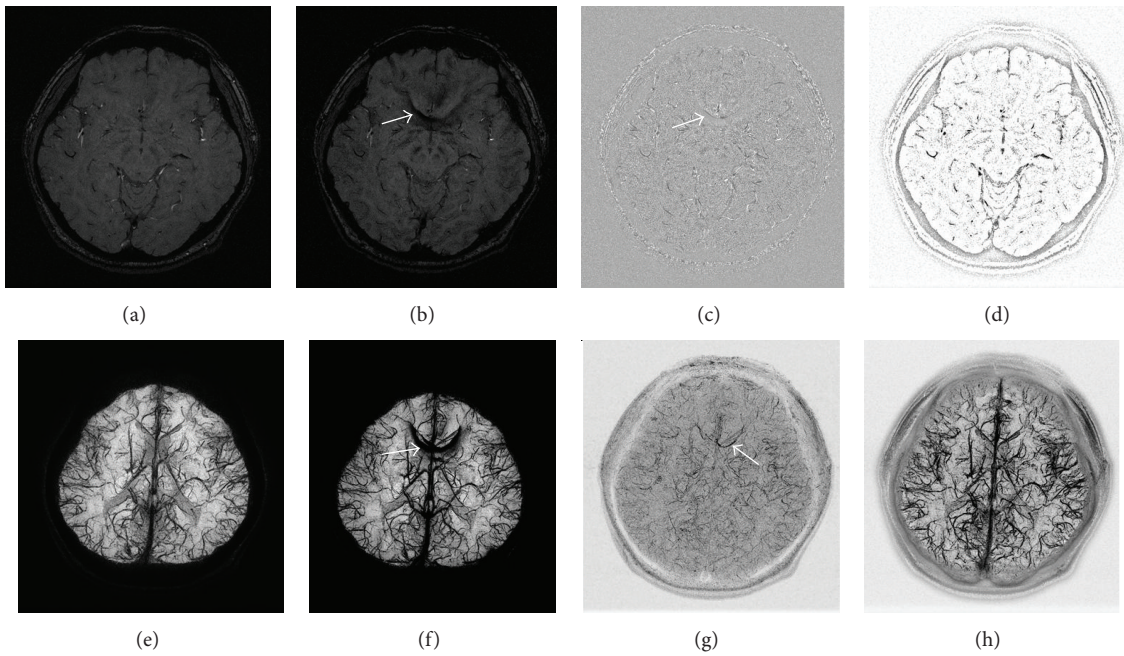


Figure 3: This figure shows the 512 × 384 × 50 (interpolated to 1024 × 768 × 100) dataset, the single slice (top row), and mIP (bottom row) of the magnitude images (a and e), MRV obtained by applying the conventional SWI postprocessing algorithm (homodyne filter size 64 × 48) (b and f), conventional SWI with a median filter (5 × 5 × 5) (c and g), and BS-MRV technique with iH filtering ($2H_x \times 2H_y = 32 \times 24$) (d and h). Severe off-resonance artifact (as indicated by bright arrows) and signal loss at the peripheral regions of the brain are observed in (b) and (f). Slight off-resonance artifacts (as indicated by arrows) were also observed in (c) and (g).

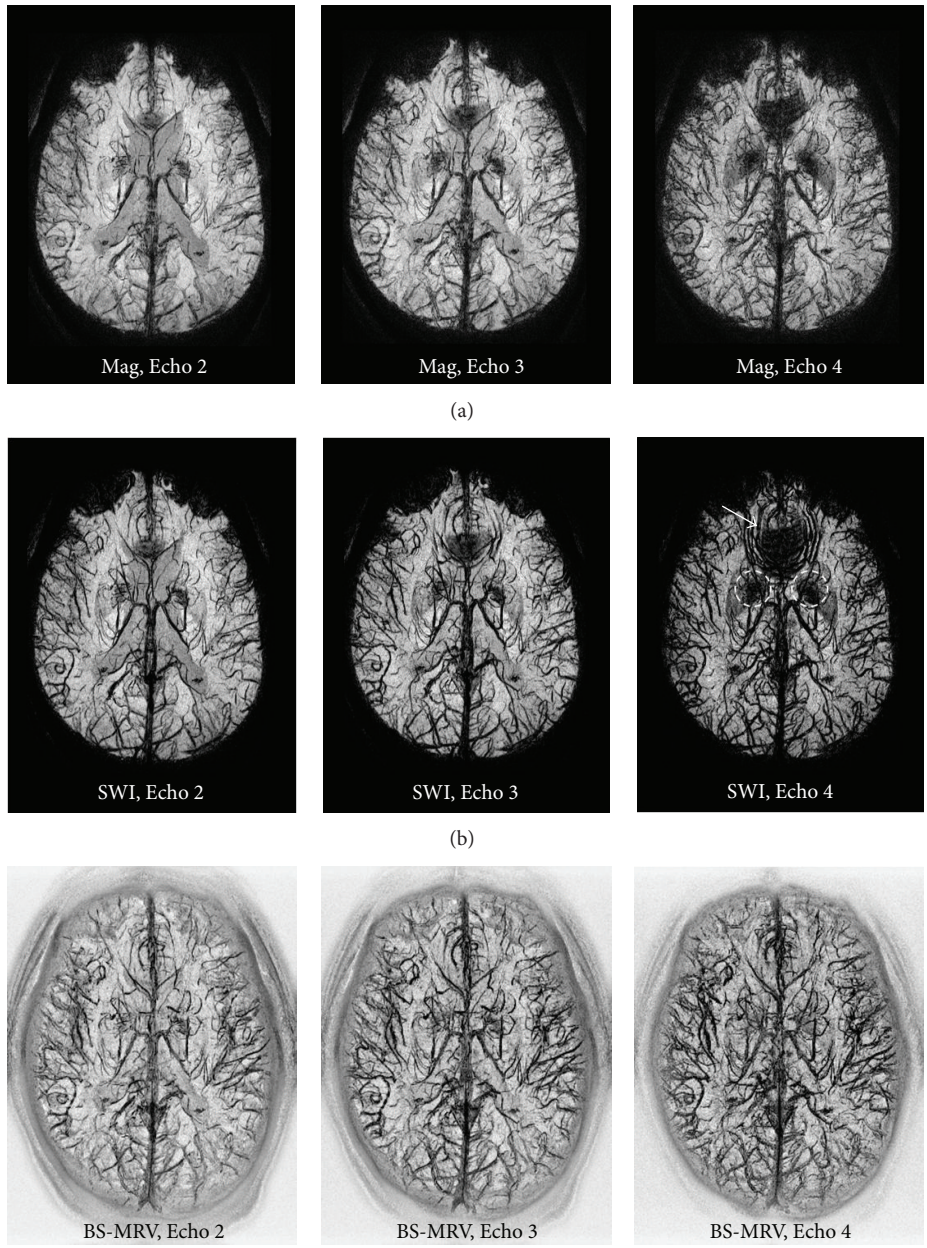


Figure 4: This figure shows the comparison of the magnitude data, SWI data, and BS-MRV data acquired with a 4-echo acquisition technique at Echo 2 (TE = 19.7 ms, left column), Echo 3 (TE = 24.8 ms, middle column), and Echo 4 (TE = 39.0 ms, right column). This figure also shows the comparison of the original 3D magnitude (top row), the conventional SWI (middle row), and the BS-MRV (bottom row) using the same datasets. The off-resonance artifact in the frontal region with severe field inhomogeneity in SWI, as indicated by an arrow, was absent in BS-MRV. The regions with high level of iron deposits, such as the globus pallidus regions indicated by the dashed circles, did not obscure the visibility of veins in the BS-MRV.

Conversely, only the magnitude data are used in the BS-MRV technique. It is noteworthy that the difference of susceptibility phase between venous blood and parenchyma within a voxel causes partial-volume signal cancellation in the magnitude data. This, therefore, has a substantial effect on the contrast of veins in BS-MRV, despite not using the phase of the acquired data in the technique. The results of this study demonstrate that MRV with good visibility of small veins can be obtained by suppressing the background signal in the magnitude data without using the phase data. Suppressing the background signal in postprocessing is an effective approach for the proper display of the venous contrast available in the magnitude data.

The use of magnitude images alone for MRV has a major practical advantage. On most of the clinical MRI scanners, phase data are not readily accessible to users. The BR-MRV

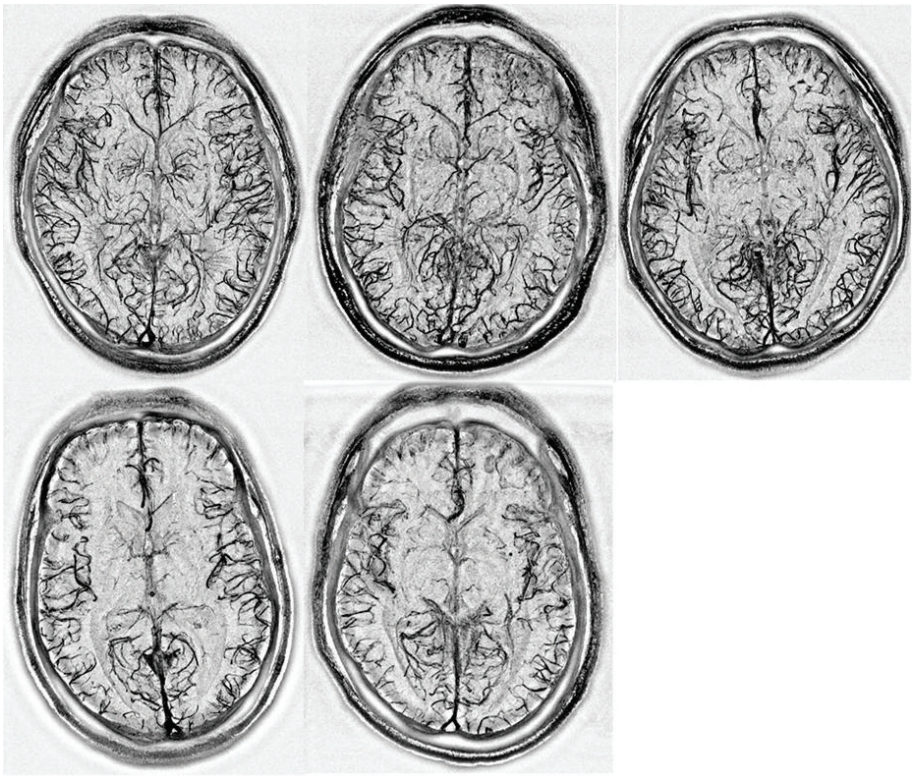


Figure 5: This figure shows the mIP of BS-MRV data of 5 subjects in a subvolume of 18 slices acquired with 8-channel phased-array RF coils.

technique can be directly applied to the magnitude images reconstructed on the scanners, including the images previously acquired. This simplicity of the BS-MRV technique overcomes a major obstacle for routine MRV applications in the clinical environment.

The off-resonance effect causes major artifacts in SWI by introducing phase wraps in regions with severe field inhomogeneity. The off-resonance artifact, commonly appearing as dark bands in the mIP display of SWI, can obscure the visibility of veins in the frontal and temporal regions of the brain. The off-resonance artifact can also appear as vein-like dark lines and be mistaken for veins. However, BS-MRV is not subject to the off-resonance artifact because it does not use the phase data.

Signal loss observed in the peripheral regions of the brain in the mIP display of SWI was substantially reduced in BS-MRV. The signal reduction in the magnitude images in iron-rich regions, such as globus pallidus, can obscure the visibility of the adjacent veins in the mIP display of the magnitude or SWI data. In addition, CSF often has a lower intensity than brain tissue, due to more severe T1 saturation in CSF. The third ventricle often appears darker than the other brain regions and, therefore, reduces the contrast of the veins along the path of its projection. Using the background suppression approach, the adverse effects of such a decrease of background signal on the visibility of veins in the mIP projection can be effectively reduced.

The contrast of veins and the severity of the off-resonance artifacts are strongly dependent on the filter size in SWI [17]. With the use of STD scaling, the contrast of veins in the BS-MRV data remains more or less the same with a wide range of filter sizes. Because of this, the adequate display of the BS-MRV data becomes more feasible with minimal operator interaction. It was further noted that both iH and iF filters provide comparable results at the same filter size.

The BS-MRV technique can be directly applied to the composite magnitude images acquired with phased-array coils, resulting in a substantial simplification of the data processing compared to the conventional SWI technique. In the conventional approach for combining data acquired with multiple coils, only magnitude data from these coils are used to form composite images with a squared-sum algorithm. This approach is not adequate in SWI, because of the loss of phase information in the composite images. For SWI, phase shift in the individual dataset acquired at each coil must be preserved in the composite images [26]. Any coil-dependent phase offset has to be properly addressed. These technical issues in multicoil acquisitions of SWI are no longer problematic in BS-MRV.

There are a few weaknesses and potential pitfalls with the BS-MRV technique. Unlike the SWI technique, the scalp signal appears dark in the BS-MRV data and can obscure the visibility of veins on the surface of the brain. In addition, the signal of major veins, such as the sagittal sinus, can be suppressed because of the use of the HP filters. For clinical diagnoses of vascular diseases of the major veins, it would be preferable to use the original magnitude images.

5. Conclusions

Good visualization of the cerebral venous vasculature can be obtained by applying proper background suppression to the magnitude of a 3D flow-compensated SPGR dataset, acquired with a relatively wide range of TE, through postprocessing. The proposed BS-MRV technique can provide good visualization of the cerebral venous vasculature even in regions with severe field inhomogeneity. With the effective background suppression, the veins adjacent to the regions of low signal intensity, such as globus pallidus, can be well depicted. Without using the phase data, the implementation of the proposed postprocessing technique on clinical scanners can be substantially simplified.

Conflict of Interests

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

Acknowledgments

This work was partially supported by the National Key Basic Research Program of China (2013CB329501), the National Natural Science Foundation of China (81371518, 81261120411, 61372024, 81271530, and 81171310), and the Key Projects in the National Science & Technology Pillar Program during the Twelfth Five-Year Plan Period (2011BAI12B01, 2012BAI10B04). The authors are grateful to Dr. Lynn Ho for editing the paper.

References

[1] H. P. Mattle, K. U. Wentz, R. R. Edelman et al., "Cerebral venography with MR," *Radiology*, vol. 178, no. 2, pp. 453–458, 1991.

[2] L. Liauw, M. A. van Buchem, A. Spilt et al., "MR angiography of the intracranial venous system," *Radiology*, vol. 214, no. 3, pp. 678–682, 2000.

[3] M. H. Pui, "Cerebral MR venography," *Clinical Imaging*, vol. 28, no. 2, pp. 85–89, 2004.

[4] N. Rollins, C. Ison, T. Reyes, and J. Chia, "Cerebral MR venography in children: comparison of 2D time-of-flight and gadolinium-enhanced 3D gradient-echo techniques," *Radiology*, vol. 235, no. 3, pp. 1011–1017, 2005.

[5] K.-O. Lövblad, J. Schneider, C. Bassetti et al., "Fast contrast-enhanced MR whole-brain venography," *Neuroradiology*, vol. 44, no. 8, pp. 681–688, 2002.

[6] K. Kirchhof, T. Welzel, O. Jansen, and K. Sartor, "More reliable noninvasive visualization of the cerebral veins and dural sinuses: comparison of three MR angiographic techniques," *Radiology*, vol. 224, no. 3, pp. 804–810, 2002.

[7] A. Tomasian, N. Salamon, M. S. Krishnam, J. P. Finn, and J. P. Villablanca, "3D high-spatial-resolution cerebral MR venography at 3T: a contrast-dose-reduction study," *The American Journal of Neuroradiology*, vol. 30, no. 2, pp. 349–355, 2009.

[8] J. R. Reichenbach, R. Venkatesan, D. J. Schillinger, D. K. Kido, and E. M. Haacke, "Small vessels in the human brain: MR venography with deoxyhemoglobin as an intrinsic contrast agent," *Radiology*, vol. 204, no. 1, pp. 272–277, 1997.

[9] J. R. Reichenbach, L. Jonetz-Mentzel, C. Fitzek et al., "High-resolution blood oxygen-level dependent MR venography (HRBV): a new technique," *Neuroradiology*, vol. 43, no. 5, pp. 364–369, 2001.

[10] E. M. Haacke, Y. Xu, Y.-C. N. Cheng, and J. R. Reichenbach, "Susceptibility weighted imaging (SWI)," *Magnetic Resonance in Medicine*, vol. 52, no. 3, pp. 612–618, 2004.

[11] V. Sehgal, Z. Delproposto, E. M. Haacke et al., "Clinical applications of neuroimaging with susceptibility-weighted imaging," *Journal of Magnetic Resonance Imaging*, vol. 22, no. 4, pp. 439–450, 2005.

[12] A. Deistung, A. Rauscher, J. Sedlacik, J. Stadler, S. Witoszynskyj, and J. R. Reichenbach, "Susceptibility weighted imaging at ultra high magnetic field strengths: theoretical considerations and experimental results," *Magnetic Resonance in Medicine*, vol. 60, no. 5, pp. 1155–1168, 2008.

[13] A. Rauscher, J. Sedlacik, M. Barth, E. M. Haacke, and J. R. Reichenbach, "Nonnvasive assessment of vascular architecture and function during modulated blood oxygenation using susceptibility weighted magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 54, no. 1, pp. 87–95, 2005.

[14] A. Rauscher, M. Barth, J. R. Reichenbach, R. Stollberger, and E. Moser, "Automated unwrapping of MR phase images applied to BOLD MR-venography at 3 Tesla," *Journal of Magnetic Resonance Imaging*, vol. 18, no. 2, pp. 175–180, 2003.

[15] A. Rauscher, M. Barth, K.-H. Herrmann, S. Witoszynskyj, A. Deistung, and J. R. Reichenbach, "Improved elimination of phase effects from background field inhomogeneities for susceptibility weighted imaging at high magnetic field strengths," *Magnetic Resonance Imaging*, vol. 26, no. 8, pp. 1145–1151, 2008.

[16] C. Denk and A. Rauscher, "Susceptibility weighted imaging with multiple echoes," *Journal of Magnetic Resonance Imaging*, vol. 31, no. 1, pp. 185–191, 2010.

[17] Z. Jin, L. Xia, and Y. P. Du, "Reduction of artifacts in susceptibility-weighted MR venography of the brain," *Journal of Magnetic Resonance Imaging*, vol. 28, no. 2, pp. 327–333, 2008.

[18] J. Neelavalli, Y.-C. N. Cheng, J. Jiang, and E. M. Haacke, "Removing background phase variations in susceptibility-weighted imaging using a fast, forward-field calculation," *Journal of Magnetic Resonance Imaging*, vol. 29, no. 4, pp. 937–948, 2009.

[19] D. S. J. Pandian, C. Ciulla, E. M. Haacke, J. Jiang, and M. Ayaz, "Complex threshold method for identifying pixels that contain predominantly noise in magnetic resonance images," *Journal of Magnetic Resonance Imaging*, vol. 28, no. 3, pp. 727–735, 2008.

[20] Y. P. Du and Z. Jin, "Robust tissue-air volume segmentation of MR images based on the statistics of phase and magnitude: its applications in the display of susceptibility-weighted imaging of the brain," *Journal of Magnetic Resonance Imaging*, vol. 30, no. 4, pp. 722–731, 2009.

[21] Z. Jin, L. Xia, and Y. P. Du, "Semi-automated volume segmentation of the brain based on a level set approach," in *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM '09)*, p. 2870, Honolulu, Hawaii, USA, 2009.

[22] M. J. Lowe and J. A. Sorenson, "Spatially filtering functional magnetic resonance imaging data," *Magnetic Resonance in Medicine*, vol. 37, no. 5, pp. 723–729, 1997.

[23] Y. P. Du, Z. Jin, Y. Hu, and J. Tanabe, "Multi-echo acquisition of MR angiography and venography of the brain at 3 Tesla," *Journal of Magnetic Resonance Imaging*, vol. 30, no. 2, pp. 449–454, 2009.

[24] E. M. Haacke, E. D. Lindskogj, and W. Lin, "A fast, iterative, partial-fourier technique capable of local phase recovery," *Journal of Magnetic Resonance*, vol. 92, no. 1, pp. 126–145, 1991.

[25] A. L. Alexander, B. E. Chapman, J. S. Tsuruda, and D. L. Parker, "A median filter for 3D FAST spin echo black blood images of cerebral vessels," *Magnetic Resonance in Medicine*, vol. 43, no. 2, pp. 310–313, 2000.

[26] E. G. Kholmovski and D. L. Parker, "High resolution magnetic resonance venography at 3 Tesla: optimized acquisition, reconstruction, and post-processing," in *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM '06)*, p. 810, Seattle, Wash, USA, 2006.

[27] S. D. Wolff, J. Eng, and R. S. Balaban, "Magnetization transfer contrast: method for improving contrast in gradient-recalled-echo images," *Radiology*, vol. 179, no. 1, pp. 133–137, 1991.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 842037, 12 pages
http://dx.doi.org/10.1155/2014/842037

Review Article

Screening for Cervical Cancer Using Automated Analysis of PAP-Smears

Ewert Bengtsson and Patrik Malm

Division of Visual Information and Interaction, Department of Information Technology, Centre for Image Analysis, Uppsala University, Box 337, 751 05 Uppsala, Sweden

Correspondence should be addressed to Ewert Bengtsson; ewert@cb.uu.se

Received 20 December 2013; Accepted 16 February 2014; Published 20 March 2014

Academic Editor: Rong Chen

Cervical cancer is one of the most deadly and common forms of cancer among women if no action is taken to prevent it, yet it is preventable through a simple screening test, the so-called PAP-smear. This is the most effective cancer prevention measure developed so far. But the visual examination of the smears is time consuming and expensive and there have been numerous attempts at automating the analysis ever since the test was introduced more than 60 years ago. The first commercial systems for automated analysis of the cell samples appeared around the turn of the millennium but they have had limited impact on the screening costs. In this paper we examine the key issues that need to be addressed when an automated analysis system is developed and discuss how these challenges have been met over the years. The lessons learned may be useful in the efforts to create a cost-effective screening system that could make affordable screening for cervical cancer available for all women globally, thus preventing most of the quarter million annual unnecessary deaths still caused by this disease.

1. Cervical Cancer Screening

Cancer of the cervix uteri is the second most common cancer among women worldwide, with more than half a million new cases each year and about half as many deaths. The variation in incidence rate between countries is striking. In many countries it is the most common cancer among women while in some countries it is down at 10th place. About 86% of the cases occur in developing countries. In Africa the age-standardized incidence rate is 25 per 100,000 per year; in some countries on that continent it is more than double that rate. In India the rate is 27 while it is 5.7 in USA and 3.7 in Finland [1]. We thus see more than a factor of ten variations in cervical cancer incidence rates between the lowest and highest countries.

While a part of this variation may be attributed to general variations in living conditions and the spread of the Human Papillomavirus, HPV, in the population the major part is attributed to the success of screening using the Papanicolaou test (PAP-test). If detected early, cervical cancer is curable and the 5-year survival rate is as high as 92% [2]. The idea behind the PAP-test is that cellular changes that may develop into cancer are detected at such an early stage that they can be removed through a simple operation, thus preventing the cancer. Evidence for the importance of the PAP-test can be found in statistics from many countries where the PAP-test is used in systematic, comprehensive screening programs. In Sweden, for example, the overall incidence of cervical cancer declined by 67% over a 40-year period, from 20 cases per 100 000 in 1965 to 6.6 cases per 100 000 women in 2005. Detailed studies of the cancer statistics confirm this [3, 4].

1.1. The PAP-Smear. The original PAP-smear is produced in a very simple and straightforward way; a brush or spatula is used to gently scrape cellular material from the squamo-columnar junction in the cervix and this is smeared onto a glass slide of about 25 × 50 mm. The cells are stained, fixated, and then visually examined under a microscope. The test was first suggested by Papanicolaou in 1928 but it took

almost 15 years before it was generally accepted by the medical community [5, 6]. A monograph in 1943 [7] gave a detailed account of how the screening should be conducted and this procedure has since been widely adopted, leading to the remarkable reduction in cervical cancer incidence mentioned in the previous paragraphs.

The screening is conducted by cytotechnologists, cytotechs for short, who through a light microscope examine the cell sample for signs of malignancy. Through this procedure they can not only find proof of invasive cancer but also detect certain cancer precursors, allowing for early and effective treatment. The cytotechs are laboratory technologists who go through a specialized training, typically of about one year. When they find something that looks suspicious for malignancy on a specimen it is reported. In many labs the finding is then confirmed by a cytopathologist, a medical doctor specializing in cellular pathology, who makes the final decision whether it is a (pre-)malignant lesion or not and thus takes the medical responsibility for the diagnosis. A detected high grade premalignant lesion typically leads to the woman being offered a colposcopy and, if a lesion is confirmed, an operation to remove it. The detection of a low grade lesion may lead to a follow-up smear being taken after a shorter time interval than the normal 2-3 years.

In principle, the screening task is straightforward. The morphological changes that a cell undergoes when it is being transformed into a malignant cell are quite apparent and easy to describe. The nucleus becomes larger and more irregularly shaped, the cytoplasm becomes smaller so that the nuclear/cytoplasm size ratio changes, and the chromatin distribution in the nucleus changes to become more coarse and irregularly distributed (see Figure 2).

To visually detect these changes we need to see details close to the optical resolution limit. A nucleus is around 10 microns in diameter and the chromatin structures and shape variations are at the micron or submicron level. Therefore a high power lens is used, typically 40x. The precancerous lesion may be quite small and local and the number of diagnostic (pre-)malignant cells on a specimen may be low. It is desirable to detect a precancerous lesion even if there are only a few diagnostic cells present on the specimen. This creates a demanding search problem. A smear covers about 25 × 50 mm and typically contains a few hundred thousand cells, sometimes even more. The screening is initially done at low resolution using a 10x lens, and when something suspicious is seen the screener switches to 40x. At 10x around 1,000 fields of view need to be scrutinized to cover the whole sample. The time required for this varies depending on how difficult the sample is, but on average it only takes 5–10 minutes. There are recommendations saying that, due to the hazards of fatigue, a cytotech should not work more than 7 hours a day and analyse no more than 70 samples [8]. Even when following this recommendation, the cytotech has to inspect three image fields per second on the average. Furthermore, since the visible precancerous changes may be quite local, the cytotech needs to maintain full concentration all the time in order not to risk missing some diagnostic cells.

2. Historical Development of Automated Screening Systems

Based on the fact that the changes in cell morphology are quite obvious and the fact that the visual screening is very demanding, tedious, and expensive in terms of labour requirements, there were very early, only a decade after the PAP-test became generally accepted, proposals for automating the screening through some kind of scanning and image analysis mechanism [9]. The hope was that an automated system would be able to do the screening both at a lower cost and with higher accuracy.

Since then a large number of projects have attempted to develop screening systems. The problem turned out to be a lot harder than anticipated. It took more than 40 years before the first successful commercial systems appeared. And still automated screening is not sufficiently cost-effective to completely replace the visual screening judging from the relatively limited penetration of automated screening systems in the screening operations worldwide. In this section, we briefly outline this development and try to see for each new generation of systems in what ways they improved on earlier systems, what were the main problems, and what was learned. We also discuss the underlying technical aspects and try to understand what makes the problem so hard and how one can go about solving it.

2.1. First Generation Systems. The Cytoanalyzer project in the US was the first attempt at building an automated screening device for PAP-smears [10]. The system was based on the concept that cancer cells could be distinguished from normal cells on the basis of nuclear size and optical density. The system included automatic slide feed and autofocus circuits. The image analysis was based on hard-wired analogue video processing circuits that generated two-dimensional histograms of nuclear size versus nuclear optical density. The spatial resolution was 5 micrometers. Preliminary experiments had shown that it was possible to detect the difference in size between normal and malignant cells at this resolution. This was the first fully automated microscope and as such a quite expensive project. Unfortunately, tests with the Cytoanalyzer revealed that the special purpose fixed logic pattern recognition produced too many false alarms on the cell level [11]. There were numerous objects of a size similar to malignant cells present also on normal specimens, for example, clumps of blood cells, strands of tissue and mucus, overlapping epithelial cells, and so forth. Every sample, including the normal ones, was thus found to be suspicious for abnormality. The project failed in the early sixties, mainly because of this artefact rejection problem.

Due to the bad reputation for cytology automation caused by this early and expensive failure in the US, the attempts at automation over the next couple of decades were shifted to Europe and Japan. In Britain a one-parameter (nuclear size) automatic screener was developed in the late sixties [12]. It failed for the same reason as the Cytoanalyzer.

In Japan, Watanabe and coworkers at Toshiba developed CYBEST [13]. Their first version used special-purpose

electronic circuits while later versions were based on general purpose digital computers, thus bridging the gap between old analogue and new digital technology. The pixel size was around one micron. They extracted four different features from the cell images: nuclear area, nuclear density, cytoplasmic area, and nuclear/cytoplasmic ratio. They also realized that nuclear shape and chromatin pattern were useful parameters but were not able to reliably measure these features automatically mainly because the automatic focusing was unable to reliably produce images with all the cell nuclei in sufficiently good focus. The chromatin pattern measure that was proposed by this group was the number of blobs within the nuclear region. Four generations of prototype systems were developed over a 15-year period. The last one used strobed illumination and nonstop scanning motion to reach high scanning speeds. The prototypes were used in large field trials in the Japanese screening program and showed promising results but none of them became a product [14].

2.2. New Generations of Systems. When the first generation systems were developed there were no interactive computers and no display units capable of showing digital images available. This, of course, made development much harder. However, during the seventies it became possible to develop interactive image analysis systems, albeit with very limited capacity, typically with a memory size of a few hundred kB and a monochrome or binary display. These systems were used to explore new image segmentation, feature extraction, and classification designs which led to a new generation of systems in the early 1980-ies such as BioPEPR [15], FAZYTAN [16], Cerviscan [17], LEYTAS [18], and at the authors' laboratory the Diascanner [19].

Typical cellular features used in these systems were similar to those used by CYBEST, although there were many variations in exactly how the features were extracted. The most important factor was found to be that the cells were digitized at sufficiently high resolution and in better focus. In order to be able to scan a whole specimen sufficiently rapidly while still being able to do the crucial analysis at high resolution, some of these designs, for example, the Diascanner, used a dual resolution approach, an initial low resolution search scan followed by high resolution scans of fields of interest. Most of these systems reached an operational prototype stage in the mideighties. Some of the systems reported classification accuracies that were well within the range of what is achieved by the conventional visual screening. But none reached the market, and an important reason for this was lack of cost effectiveness; automated microscopes and computers with sufficient processing power were still too expensive.

The progress in computer display technology, that had been important in making it possible to create interactive systems that could be used for developing new automated screenings systems, eventually also led to the possibility of developing interactive screening systems. For the early systems the only option was full automation, or possibly stopping the automated microscope to physically show an operator the cell that was suspected as being abnormal. The concept was to create a "prescreening" system; that is, a system that for a reasonably large fraction of specimens would be able to say that they are perfectly normal and could be classified as such without any human inspection. All other specimens, on which the system found something that indicated that they might not be normal, would have to be screened in the conventional fully manual way. In the late eighties, computer displays and memories had reached sufficient capacity to make it feasible to save images of suspicious cells that were good enough for a human to judge whether the object could be a malignant cell or something else. The PAPNET system from Neuromedical Systems was the first to introduce interaction into automated screening [20]. After an initial low resolution object search, high resolution fields were processed, first by an algorithmic classifier and then by a neural network classifier. The output of the classifiers was a ranking of the abnormality of the detected "cells," so that images of the 64 most abnormal ones could be stored on a magnetic tape and later shown to the cytotech at a review station. There the decision whether the specimen should be classified as normal or suspicious was taken. For the suspicious cases a cytopathologist would do the final analysis and make the decision whether the woman should be called for follow-up or not.

In the late eighties there was a great increase in interest in cytology automation in the US for economic/legal reasons and many new projects were started [21]. One new aspect that appeared at this time was new ways of preparing the samples. The Cytyc Corporation had developed their own automated specimen preparation technique, ThinPrep, which based on liquid cytology made much cleaner specimens than the conventional smears, at the expense of significantly more complex preparation technique [22]. Another similar preparation method was developed by AutoCyte [23].

The AutoPap 300 from NeoPath was similar to PAPNET in that it used conventional Pap-smears and neural network classifiers [24]. It increased the image acquisition rate by utilizing strobed illumination similar to the CYBEST system. This was used at two resolution levels, an initial low resolution mapping of the specimen, followed by a high resolution field by field analysis of the most "interesting looking" parts of the specimen in a way similar to the earlier generation Diascanner. The image processing was carried out in custom designed processing boards. Most of the processing was based on mathematical morphology operations resulting in as many as 68 different features being sent to the classifiers. The final result was a "normal" versus "requires visual inspection" decision on the specimen level; that is, no interactive confirmation was used of the machine decision for the negative cases.

2.3. The First Commercially Available Systems. During the nineties there was strong competition between the American companies developing screening technology as well as struggles to get the various solutions approved by the powerful Food and Drug Administration, FDA. Screening systems were classified in a category of medical devices requiring premarket approval, meaning that no system can be sold in the US without FDA approval. Hundreds of millions were

spent on developments and field trials and there was a shake-out; the companies merged and were acquired by larger companies. The first company with a screening product to finally receive FDA approval was Tripath in 1998. It was the merger of NeoPath, Neuromedical, and AutoCyte. The Tripath Company was in turn acquired by BD in 2006 and the system renamed BDFocalPoint Slide Profiler [25]. It is to a large extent based on the AutoPap 300 system. A new liquid based specimen preparation technique called SurePath has been added to further improve the system performance although it can also analyse conventional smears. According to the FDA approval, the system can be used to recognize about 25% of the slides as normal for no further review; the other 75% are ranked into five categories at risk for abnormality. There is also a possibility of visually reviewing fields of particular interest at a special review station. The system can also be used for quality control and claims increased sensitivity in detecting abnormalities [26].

Cytyc was quite successful with their improved liquid based preparation technique and could demonstrate better performance for that technique as compared to conventional smears. They also developed an interactive system with a computer prescreen that selected the most abnormal looking objects on each specimen for human inspection. In 2003 they received FDA approval for their ThinPrep Imaging System [27], and in 2007 they became part of the Hologic Company. The system is marketed for increasing detection of abnormalities by improved specimen preparation and screening both visually and by machine [28].

3. The Technical Challenges

In the quick review of the historical development above we have briefly mentioned some of the key features of the different generations of systems. We will now return to the different crucial aspects of the technologies behind a screening system and discuss what needs to be achieved in order to screen a sample in a time comparable to that of a human screener that is less than 10 minutes.

3.1. Specimen Preparation. In the original PAP-smear the cellular material is manually spread over the glass slide. It is important that both the endocervical and ectocervical regions (see Figure 3) are represented in the sample and through the smearing there may actually be a mapping between the source region and the location on the slide. [29]. The samples are fixed and stained in a rather straightforward procedure, which can be done fully manually or in staining machines with varying degrees of sophistication. The material cost for the whole preparation is quite low, on the order of 1-2 US dollars. In Figure 1, an image of a high resolution field from a PAP-smear is shown.

The manual smearing and staining do unfortunately lead to big variations in specimen quality. Sometimes the cellular material may be unevenly distributed leading to dense clumps which light cannot penetrate while other parts of the slide may be empty. Even when the smear is done well there will still be regions which are too dense and have too many overlapping cells for reliable interpretation. An experienced cytotech can cope with great variations in specimen quality and still make a rather reliable assessment of the specimen, but the smears are very challenging to analyse automatically.

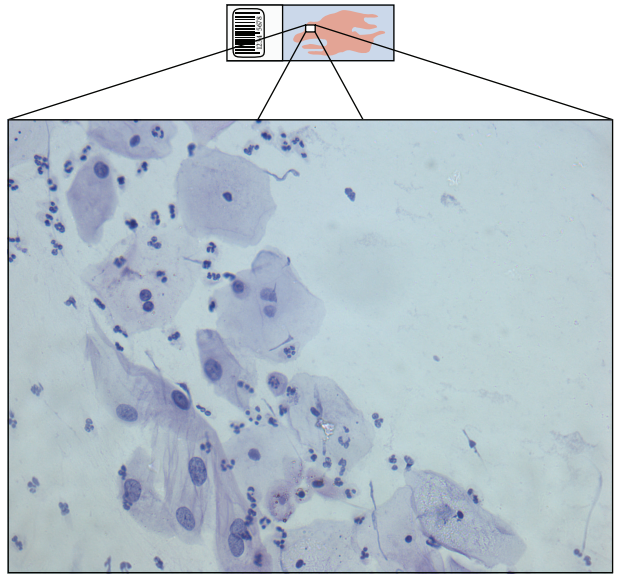
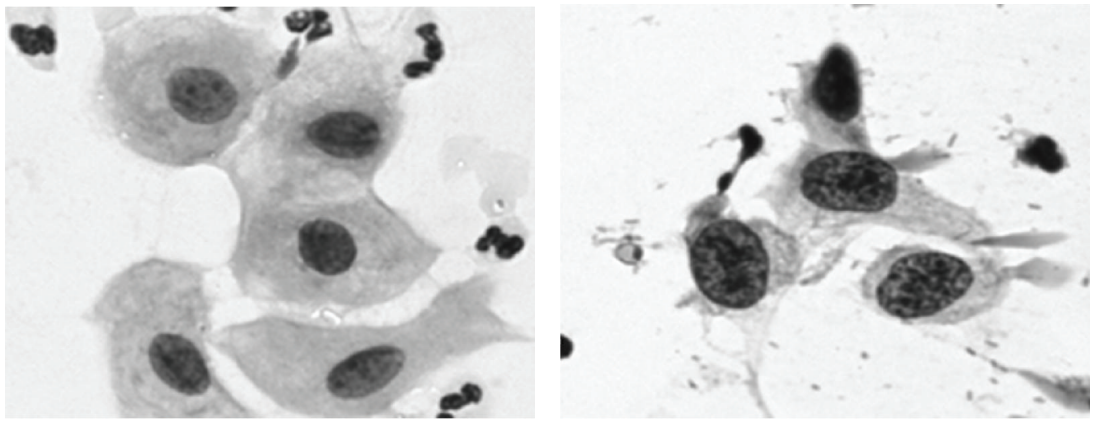


Figure 1: A typical PAP-smear and a high resolution field of view through a 40x lens. Approximately 10.000 such fields of view are needed to cover the whole slide.

To make specimens that are better, both for visual and machine analysis, various liquid based cytology (LBC) preparation techniques have been developed. The common strategy here is to submerge the brush or spatula with all the cellular materials collected from the cervix in a liquid, which then is treated in various ways before it is deposited onto a glass slide, fixed, and stained. The result is ideally a cellular sample that is spread in a monolayer with optimal density over a well-defined part of the glass slide. The goal of this procedure is that the resulting samples should be easier to interpret reliably visually and in particular by machines. Several different techniques for liquid based preparations have been developed over the years, the two leading techniques are Surepath [25] and Thinprep [27] mentioned above. There have been numerous studies comparing the liquid based preparations to the conventional smears and most of them come to the conclusion that they are at least as good or better when it comes to reliability of detecting abnormalities [28, 30–32]. All currently marketed machine screening systems work with liquid based preparations.

The great disadvantage of the liquid based preparations is the associated operational costs. They require significantly more materials to be used for preparing a specimen, for example, vials, liquids, filters, and also more complex equipment, for example, centrifuges. The procedures are proprietary and the necessary equipment is sold as kits which increases the cost of preparing a slide to at least 10 US dollars. This causes significant economic problems in regions with limited resources. Still there are studies indicating that liquid based



(a) (b)

Figure 2: To the left a few normal cells and to the right some clearly atypical, premalignant cells.

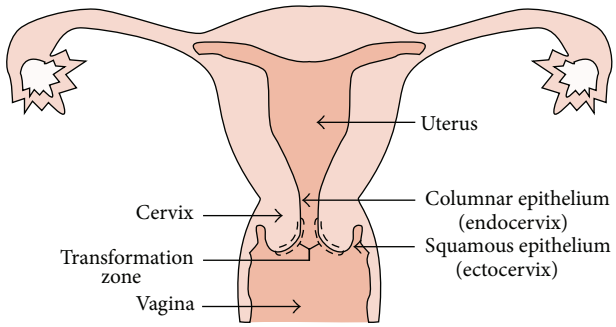


Figure 3: Illustration showing the anatomy of a uterus. For specimen acquisition it is important that cells are acquired from both the endocervical and ectocervical regions, that is, both above and below the region known as the transformation zone.

preparations are more effective [33, 34] while a large metastudy concluded that they could see no significant differences [35]. There are also alternative liquid based preparations that have been developed and are competing with lower costs [36, 37] although those have so far not been tested as extensively as the leading techniques.

3.2. Scanning. In order to analyse a cell sample in a computer, it should be scanned at sufficiently high resolution to reliably extract the features that can determine whether it is normal or indicating a precancerous change. This is very challenging. At a pixel size of 0.2 microns, a smear of 25×50 mm will give 31 billion pixels. Just transferring this amount of data from the camera to the computer will take minutes, even with the latest high-speed transfer techniques. Since there are no lenses that can resolve the whole specimen area at once and no image sensors with 31 gigapixels, we have the problem of repositioning the lens over a large number of image fields that together cover the specimen. A high resolution microscope lens gives a field of view with a diameter of around 0,5 mm and with a matching 6 megapixel sensor we will get 5000 image fields. Repositioning and capturing an image at each of these will take at least 10 minutes. This can be reduced by using nonstop motion and flash illumination to freeze the images. The CYBEST4 system was the first screening system to use this idea [14] and later it was used in the AutoPap [24]. An alternative is to use a 1D sensor with a length of, for example, 2000 pixels and smoothly move the microscope stage in the orthogonal direction. The Cerviscan [17] and Diascanner [19] systems used this idea. It is also used in the currently popular slide scanners by Aperio [38] although at a lower resolution.

Another serious issue is focusing. In order to reliably extract the texture information from the cell image they must be in very good focus which requires high quality autofocus, which also is time consuming. An alternative is to scan the specimen at several focus levels and choose the best for each cell, which reduces the need for autofocus but increases the amount of data even more. So in summary it is quite demanding to scan a whole smear in a sufficiently short time at sufficiently high image quality. In Figure 4, an illustration of the two different scanning approaches mentioned above is seen.

One way to decrease the demands is to use a smaller part of the slide surface for the specimen. With a smear this cannot be done without decreasing specimen sampling quality. With liquid based preparation the area of the sample is around 1/10 of that of a smear, a great advantage when it comes to scanning. For smears we can instead use a dual resolution approach mimicking the way cytotechs switch between 10x and 40x lenses. If we scan with 1 micron pixel size we can cover the smear in 200 fields which can be done in about 20 seconds. This will produce a map of where cellular material with a suitable density is distributed, which can then be used to control where a number of scans at high resolution are acquired. A variant of this approach is to not only look for areas with suitable density of cells but also for areas with cells that look suspicious for abnormality. This dual resolution screening approach was first proposed by Poulsen [39] and used in some of the early systems, for example, the Diascanner [19]. A potential risk with this design approach is that, if the low resolution scan systematically

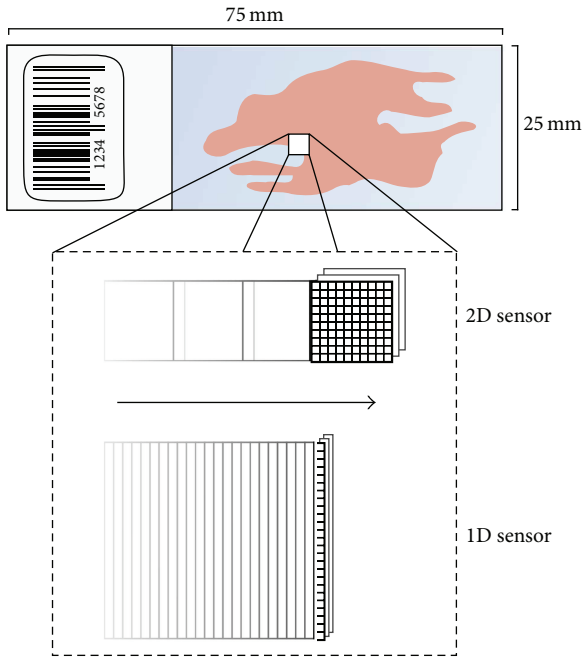


FIGURE 4: Illustration schematically showing image scanning using one- or two-dimensional sensor arrays. With a 2D array an image or a stack of images at different focus levels are read before the microscope stage moves a few hundred micrometers to a new position where this process is repeated as soon as the vibrations caused by the move have died out. With a 1D array the microscope stage is moving continuously and single lines in the direction orthogonal to the move are read into the computer creating a continuous flow of image data.

misses some type of abnormalities, those abnormalities never become subject to the high resolution analysis.

3.3. Segmenting Cells and Nuclei. In order to extract the features describing the cells we must find and delineate each cell and/or cell nucleus in the specimen image. This is called image segmentation and is a crucial step in almost all image analysis based systems. Segmenting nuclei in PAP-smears is made very difficult by the same complications that make the smears hard for humans to analyse, that is, variable smear thickness and staining intensity, obscuring elements, and so forth. The earliest systems used thresholding based on greyscale for the segmentation in the very first systems using a fixed threshold value but later on with a value determined by histogram analysis as originally suggested by Prewitt and Mendelsohn [40]. More recent projects have used more complicated approaches. Bergmeir et al. [41] use mean shift and morphological filtering and later try Canny edge detection followed by the randomized Hough transform [42]. Bamford and Lovell [43] use a dual active contour algorithm. Malm and Brun [44] use Canny edge detection followed by anisotropic curve closing. In a recent review [45], five different classes of approaches to cell segmentation are identified and it is demonstrated how they have appeared and gained popularity over the years. There is still a need of developing new methods, since none of the existing ones are as flexible and robust as the human visual system in really identifying where the nuclear or cytoplasmic border is located in difficult cases.

The main requirement for a good cell nucleus segmentation method is that it accurately can detect and delineate the cell nucleus under different staining conditions and in the presence of disturbing object in the direct vicinity. A second important requirement is that this segmentation can be done quickly. We cannot spend more than a few milliseconds per cell if we are to accomplish the analysis in an acceptable time. The increasing computer power has made it possible to do this even with somewhat complex algorithms. It does, however, require the algorithms to be implemented in an efficient way, for instance, taking advantage of the possibilities of parallelism possible in modern computers.

3.4. Artefact Rejection. The goal for the segmentation algorithms is to find and accurately delineate cell nuclei (and sometimes cytoplasms) that are sufficiently well preserved and imaged to allow accurate extraction of features for the subsequent classification. But it will fail sometimes either because the image of the nucleus is corrupted by overlaying objects or other artefacts or when the cell is so poorly preserved or presented in the image that the extracted outline of the object will be wrong. It is then very important that we can detect this failure and discard the data from the object. Otherwise it will lead to unreliable classification performance on the specimen level. The process of analysing the segmentation results in order to remove erroneous results is called artefact rejection.

Artefact rejection is a difficult topic because there are an infinite variety of ways in which blood cells, inflammatory cells, folded and distorted cells, overlapping objects, mucus, staining mistakes, and so forth, influence the image of a cell (see Figure 5). But it is an absolutely essential step in a screening system. The motivation for this can be found in the statistics we have to deal with. A standard PAP-smear may typically contain 100,000–200,000 cells of the relevant cell types and we should be able to call it positive if we find 10–20 diagnostic, premalignant, or malignant cells (ideally a single clearly malignant cell should be enough). A classifier that only makes one percent false positive error will call at least 1000 cells positive even on a healthy sample, making every sample called positive and the system thus useless. One approach to deal with this problem is to make the classifier highly asymmetrical between false positive and false negative, that is, allowing it to miss-classify a large fraction of the actually malignant cells as normal. This may seem to defeat the purpose of the system which is to detect (pre-)malignancy. But the highly unbalanced numbers work that way. If the system has a false negative rate of 80% it will still detect 2–4 of the diagnostic malignant cells if we have 10–20 available. This is acceptable as long as the false positive rate is virtually zero, less than 0.001%. Creating such a classifier is hard but possible if we can work with perfectly imaged cells with accurate segmentation and carefully extracted features. To make sure this is the case we need very effective artefact rejection.

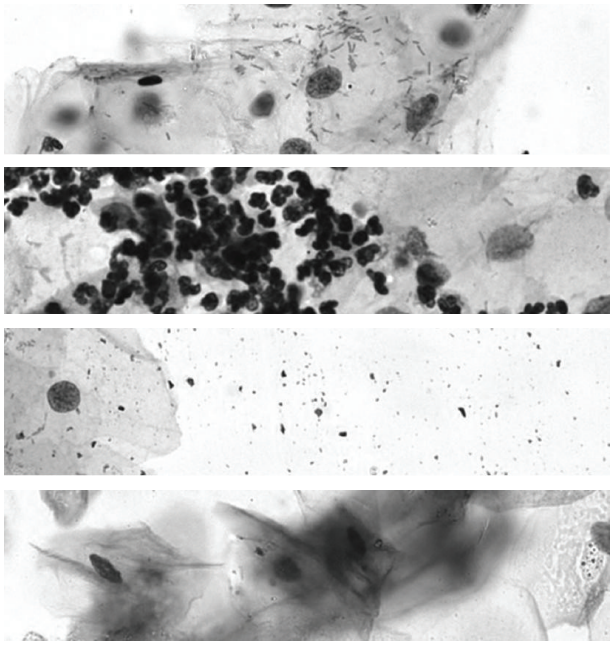


Figure 5: Images displaying four common types of artefacts found in PAP-smears. From top to bottom: bacteria, leucocytes, stain residues, and overlapping and folded material.

There is very little explicit research done on artefact rejection for cervical screening. Some research papers ignore the problem by working on visually selected or verified images of nuclei thus relying on manual artefact rejection, which of course cannot be done for a real screening system. Other papers include the artefact rejection in the segmentation or classification steps. Still, analysing the artefact rejection problem on its own makes it easier to see what performance can be achieved and to relate that to what is needed. Malm et al. recently presented such a study where they demonstrated a specificity of 99.38% on smears and 99.83% on LBC specimens, while maintaining a sensitivity of around 98% based on a material of around 12,000 automatically detected and segmented images of objects visually classified into cell nuclei and artefacts [46]. With that kind of performance we would still have a few hundred artefacts corrupting the data if we analyse 100,000 objects, so it may be hard to achieve the sensitivity of detecting a few abnormal cells without getting too many false positive samples. Still it points in the direction of what is necessary to achieve for a useful system.

3.5. Feature Extraction. When we have an accurate segmentation of cell nuclei, we can extract features describing the size, shape, and texture of the object. The most obvious features are those representing the greater size and more irregular overall shape of the malignant nuclei. Those features can be extracted at relatively low resolution and even without having the cell in perfect focus. Assuming perfect artefact rejection those features may be useful in detecting a large proportion of the clearly malignant cells and specimens. They were used in the first generation systems, which failed because of the lack of adequate artefact rejection. Over the years, many more features have been invented and tested. In [47] the different kinds of features that have been proposed were reviewed and systematically categorized.

The most important information about whether the nucleus is normal or (pre-)malignant is found in the chromatin pattern or texture of the nucleus. The DNA in the nucleus is distributed in a different way when the cell is influenced by a malignant process. This effect can be seen and measured even with PAP-stain which is not stoichiometric for DNA. Measuring the chromatin distribution is, however, difficult. The most common approaches are based on a statistical description of neighbouring grey levels typically measured through so-called transition probability matrices as originally proposed in [48]. Another approach is to see the individual chromatin granules as objects that are segmented, and then the spatial relations between these objects are described, for example, through graph analysis methods [49]. Different ways of measuring chromatin features are discussed in [50] and also in [47]. A very important aspect of the chromatin analysis is that you need perfect focus and very high quality images to reliably represent this pattern which is at or beyond the optical resolution limit.

4. Classification Strategies

The ultimate goal of the PAP-smear screening process is to find women with precancerous lesions, so that they can be treated before the malignancy develops into potentially lethal invasive cancer. When running the conventional visual screening process, the cytotechs can classify most specimens as clearly normal and needing no further review. It is important to realize that we typically are screening a general population so that the great majority of samples, perhaps 96%, are normal. Still some specimens look more suspicious and are referred to a cytopathologist for review; in some cases the malignancy may be so obvious that the cytotech can be sure about it; still the confirmation by a pathologist is required. A positive sample will then lead to the women being called in for additional investigation, possibly involving colposcopy and a biopsy, and if the lesion is confirmed, a simple operation with a loop electrosurgical excision procedure or similar to remove it. There are different levels of changes in cell appearance that can be detected and there is a consensus standard for how to classify these called the Bethesda system [51]. According to the Bethesda system there is also a range of abnormalities from the perfectly normal slide via slight abnormalities ASC-US, low grade lesions, LSIL, high grade lesions, HSIL, and finally cancer. It is of course particularly important to pick up the higher grades and not clear whether it really is necessary to detect the slight changes. Not all low grade lesions will progress to cancer even when left untreated and when they do it may take a decade or even more. With a regular recurring screening program taking a new sample with a few years interval, the probability of detecting a lesion before it progresses to cancer is therefore high even if the risk of missing it at a single screening occasion is rather high, perhaps 20–30%.

When adding an automated screening device to the overall screening setup, it can be used in various ways. The original concept was to do an automated prescreening, which would be able to say that a substantial fraction of the specimens were normal while having very few false negatives, that is, not missing any, or very few true positive specimens. To reach a low false negative rate a relatively high false positive rate could be accepted since those specimens were screened visually. Even if only 50% of the specimens could be dismissed as clearly normal the system would remove half the visual screening workload and could be cost-effective if it did not add too much to the overall screening cost. The SurePath system is typically used in this mode and set to only remove 25% of the specimens, while also ranking the positives into different categories of likelihood of being truly positive.

Another way of using an automated system is to run it in parallel to visual screening. Since humans and machine most likely will make different errors, the combined system will increase the sensitivity of the overall screening process, that is, reduce the false negative rate. But the downside is that the overall workload and cost are increased rather than decreased. The Thinprep system is mainly marketed to be used in this mode.

4.1. The Rare Event Approach. An image analysis based automated screening system analyzes cells one by one and can, based on the features extracted from the cell image, classify it as being normal or abnormal. The simplest way of using the information from the cell classifier is to simply count how many abnormal cells we have found on the specimen and if it is over a low threshold we call the specimen suspicious. The problem is to set the threshold so that we do not miss true positives in particular not high grade ones, while avoiding too many false positives [52].

This approach disregards the information about how certain the cell classifier is about its decision. It may be that one cell is found to be clearly malignant while another one is very close to the threshold for being normal. If we retain this information, we can make a specimen level decision about whether we have found too much abnormality to call the specimen normal, either based on just a few clearly malignant cells, or a larger number of cells slightly over the threshold [53].

If we can do the image analysis and feature extraction online as we scan the specimen, we may stop the analysis as soon as we have found sufficient evidence that the specimen is not clearly normal. This may save time by making it unnecessary to scan and process the rest of the specimen. This strategy is not controversial since it does not increase the risk of false negatives. But since the great majority of specimens are normal in a typical screening situation we would need to be able to stop early also when we have found sufficient evidence that a specimen is normal in order to really save time. And this is controversial; it is generally required that a cytotech looks at the entire specimen before calling it normal. Still only a small part of the cells scraped from the cervix really makes it onto the glass so we are not analysing all possible cells even when we look at the whole slide. However, with a conventional PAP-smear there is a kind of mapping between areas in the cervix onto the slide so we may systematically miss some important region by stopping early. For a liquid based preparation, much fewer cells are available for analysis on the specimen, but there is a mixing step involved so we can assume that we have a random sample and can stop as soon as we have sufficiently many cells for a required statistical significance in the decision function.

4.2. Malignancy Associated Changes (MAC). In the approach to the screening problem described so far the systems have been mimicking the way humans do it, that is, searching for potentially rare (pre-)malignant cells. Achieving a low false negative rate even for specimens with a low number of diagnostic cells is challenging and requires analysing very many cells. There is, however, an alternative approach based on so-called malignancy associated changes (MAC). It was discovered already in 1967 that cells in the vicinity of a malignancy are influenced so that they undergo small, often subvisual changes in the chromatin texture [54]. These discoveries were confirmed in the early research on automated cervical screening [55, 56]. Even though these shifts were not strong enough to be useful on the individual cell level, it made it possible to detect abnormal specimens through a statistical analysis of the feature distributions of a small population, a few hundred cells, provided that these features were extracted very accurately. Since these changes are present in all cells in a large neighbourhood of a malignant process we can have a different approach to the screening. We need to be able to reliably detect these subtle changes in cell populations from a specimen. But we will not need to search through the whole specimen; only data from sufficiently many cells to characterize the chromatin distribution of the cell population is needed, typically around 500 cells. The group that has been pursuing this idea most systematically is the one at the British Colombia Cancer Research Centre [57, 58]. Also in this case we need very accurate artefact removal; we do not want to extract texture data from artefacts. And we need perfect focus for each cell. Even a small deviation from perfect focus causes significant changes in the chromatin image. It has not yet been convincingly demonstrated that MAC alone can detect early premalignant changes with sufficient sensitivity.

4.3. The DNA Ploidy Approach. The malignant process not only modifies the distribution of DNA in the nucleus, it also increases the amount of DNA. With a stoichiometric stain, a histogram over the integrated optical density of all the nuclei will show a diploid distribution for normal cells and a different aneuploid distribution for malignant cells. Therefore modified, stoichiometric PAP-like stains have been developed and used for automated screening studies [59–61], showing quite promising results. This method is currently being used in China in a study involving several hundred thousand women [62]. Since this approach is based on densitometric measurements, there are rather strong requirements of consistent staining and control of the illumination and calibration of the imaging. The artefact rejection is also very important; we must be sure that the DNA measurements

are only from single, free-lying, well-preserved nuclei. A significant problem with these modified stains is to get them accepted by the wider community, since new appearance of the samples may require expensive retraining. This concept can also be used with the conventional PAP-stain, but due to the lack of stoichiometric staining the ploidy measurements will be less reliable.

4.4. Field Test Statistics and Performance Requirements: Technical and Ethical Issues. The statistics needed for developing a successful screening system is difficult on several levels as described in the previous paragraphs. But we also have difficult statistics on the highest population screening level. A screening machine that systematically misses a significant proportion of the positive samples will decrease the confidence in the screening programs and put the women at risk of developing invasive cancer before the problem is detected. Since a screening system is typically used mainly to analyse normal specimens, at most a few percent of the samples are truly positive. To prove the detection capabilities of the system with high confidence we need it to analyse hundreds of positive cases. In a development phase we can achieve high numbers of positive cases by selectively running positive cases in the machine. But for the final evaluation we should run it on the typical mix of routine specimens. We thus need a volume of tens of thousands of specimens to properly test the system. And the situation is made even more complicated by the fact that there are various kinds of rare abnormalities that are detected by the visual screening. We need to verify that the machine does an acceptable job also for those. During this testing phase the machine will have to be used in parallel to visual screening, causing double operational costs plus extra work for doing the comparisons and statistical evaluations. The final system verification stages are thus a difficult and expensive threshold to get over before a new system is ready for widespread use.

What performance requirements a screening machine must meet is an issue that has caused controversy over the years. The perfect machine should have zero percent false negatives and zero percent false positives. In practice this is impossible so we have to consider what the realistic requirements are.

The false negative rate is primarily an ethical issue. If we miss to detect high grade lesions, it may lead to the women getting cancer. The requirement should then be that the machine is at least as good as the current visual screening process. But there are two aspects also of this requirement. It should on average not miss more specimens than what is missed by a good cytotech. Additionally it should not systematically miss any relevant kind of lesion.

The false positive rate of the initial automated specimen inspection on the other hand is an economic issue. False positives from the whole screening setup, after inspection by a cytologist are expensive and cause anxiety for the woman who is called for a new investigation which may be interpreted as a message that she may have a cancer. But no screening system is set up so that the machine positive samples lead to a call for the woman to come to a new examination. The samples classified as potentially positive by the machine are screened visually by a cytotech and, if still found to be positive, by a cytologist. This rescreening costs money and reduces the gain of having the machine screening. Still very high levels of "machine false positives" can be accepted, for example, 75% for one of the commercial systems. The machine is then set up so that only clearly normal specimens are classified as normal, and everything else should be inspected also by a human.

4.5. Performance Requirements: Legal Issues. The deployment of an automated screening process is made significantly more complicated by the legal aspects. There are in most countries, for good reasons, strong regulations for how to test a screening machine before it is approved for routine use. In the USA a screening machine needs premarket approval by FDA before it can be sold for clinical screening use. Obtaining such approval involves detailed documentation of all aspects of the machine as well as extensive testing of its performance in large, well-documented studies. But the legal aspect is not limited to obtaining approval from the appropriate authorities. If a machine misses to detect a high grade lesion present in a sample and this leads to a woman getting cancer, the manufacturer of that machine can be sued for the damages caused. This can be extremely expensive and is a risk no manufacturer can take. Therefore the procedures for how to use the screening machines are designed to minimize this risk. One way of doing this is to require a human screener to visually inspect some data from every specimen, for example, a set of images of objects the machine determined to be the most "malignant looking" on the specimen or a number of selected image fields optically through the eye-piece of the microscope. Thus the responsibility of calling the specimen "normal" is transferred from the manufacturer to the user. Another approach has been to run the machine screening in parallel to conventional visual screening with the rationale that the errors done by machine and by the human are different and thus the overall sensitivity for detecting malignancy is increased. By setting the threshold for when to call a specimen "normal" without further human inspection very conservatively some manufacturers have decreased the risk of missing a positive specimen to the point where it has been deemed acceptable. In the USA that threshold currently seems to be at the 25% level; that is, 75% of all specimens will have to be screened both by machine and humans. No fixed such thresholds are set in other parts of the world.

All these legal precautions are meant to protect the women from unnecessary risks of obtaining cancer in spite of having been screened, which of course is a good thing. But they have also significantly contributed to the fact that automated screening so far has failed to have a real impact on the screening costs. Still the majority of women in the world are not offered regular screening because of the high associated costs. An automated screening system that is almost as good as the best visual screening systems at a significantly lower cost could save millions of women from dying in cervical cancer. But will such a device ever be

accepted by the legal systems? It will not probably be accepted in the USA but perhaps in some other parts of the world.

5. Conclusion

We have in this paper outlined the 60-year history of efforts to automate the screening for cervical cancer and pointed out how the different generations of systems have tried to meet the challenges of this difficult task. We have also discussed the different aspects of these challenges and how they can be met. Now in conclusion let us discuss where we stand today.

The purpose of an automated inspection system is to decrease the cost and/or false negative rate of a screening program. To achieve the first goal it is necessary that the cost of operating the system including capital and maintenance costs in addition to the direct operational costs is less than what it costs to do the same work as the system does with conventional manual methods. It is doubtful if the present generation commercial systems meet these goals. They have been more focused on the second goal. By running machine screening in parallel to visual screening it is likely that the machine misses other abnormalities than the human screener, thus reducing the overall false negative rate. The overall operational cost will however be higher. The currently available commercial systems may thus marginally increase the quality of the screening but they will not significantly decrease the cost.

It is today known that cervical cancer is caused by human papillomavirus infection. An alternative or supplementary screening method is to test for such infections. There are studies showing that combining both kinds of analysis adds sensitivity of detecting precancerous lesions [63]. Still there seems to be limited advantage of replacing the PAP-test by only a virus test. The knowledge that the cancer is caused by a virus infection has also opened up the possibility of vaccination against the HPV virus. If such vaccination programs became globally comprehensive, the prevalence of the cancer could be decreased to the level where screening would be no longer necessary. But unfortunately it is not likely that that will happen any time soon and it will take decades even after everyone is offered vaccination before the effects reach all age groups.

The historical developments of the screening field have taken place in parallel to the fantastic development of computer technology. We now have millions of times more computer power available per dollar than we had when the first digital screening systems were built. Similarly the image sensor technology has developed dramatically. Even since the time the first version of the current commercial systems was developed some 15 years ago there has been significant progress in the underlying technologies. There have also been significant developments on the algorithmic side, although perhaps not as dramatic as for the hardware. All this sets the stage for a good opportunity today to develop a really cost-effective screening system. It is most likely that a fully automated screening system today can be built at a cost of at least an order of magnitude lower than the cost of the currently available commercial systems. Such a system could make it economically feasible to implement comprehensive screening systems also in the poorer parts of the world and eventually have an impact on the high incidence of cervical cancer there. There are of course major challenges in organizing an effective screening program in such countries, but a compact, robust automated screening system could make a big difference.

Conflict of Interests

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

References

[1] WHO/ICO information centre on hpv and cancer, 2014, http://www.hpvcentre.net/statistics.php/.

[2] D. Saslow, D. Solomon, H. W. Lawson et al., "American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer," *American Journal of Clinical Pathology*, vol. 137, no. 4, pp. 516–542, 2012.

[3] B. Andrae, L. Kemetli, P. Sparén et al., "Screening-preventable cervical cancer risks: Evidence from a nationwide audit in Sweden," *Journal of the National Cancer Institute*, vol. 100, no. 9, pp. 622–629, 2008.

[4] B. Andrae, T. M.-L. Andersson, P. C. Lambert et al., "Screening and cervical cancer cure: Population based cohort study," *British Medical Journal*, vol. 344, no. 7849, article e900, 2012.

[5] G. Papanicolaou, "New cancer diagnosis," in *Proceedings of the 3rd Race Betterment Conference*, pp. 528–534, 1928.

[6] G. Papanicolaou and H. Traut, "The diagnostic value of vaginal smears in carcinoma of the uterus," *The American Journal of Obstetrics and Gynecology*, vol. 42, pp. 193–206, 1941.

[7] G. N. Papanicolaou and H. F. Traut, *Diagnosis of Uterine Cancer by the Vaginal Smear*, Oxford University Press, Oxford, UK, 1943.

[8] T. M. Elsheikh, R. M. Austin, D. F. Chhieng, F. S. Miller, A. T. Moriarty, and A. A. Renshaw, "American society of cytopathology workload recommendations for automated pap test screening: Developed by the productivity and quality assurance in the era of automated screening task force," *Diagnostic Cytopathology*, vol. 41, no. 2, pp. 174–178, 2013.

[9] R. C. Mellors and R. Silver, "A microfluorometric scanner for the differential detection of cells: Application to exfoliative cytology," *Science*, vol. 114, no. 2962, pp. 356–360, 1951.

[10] W. E. Tolles and R. C. Bostrom, "Automatic screening of cytological smears for cancer: the instrumentation," *Annals of the New York Academy of Sciences*, vol. 63, no. 6, pp. 1211–1218, 1956.

[11] C. C. Spencer and R. C. Bostrom, "Performance of the cytoanalyzer in recent clinical trials," *Journal of the National Cancer Institute*, vol. 29, pp. 267–276, 1962.

[12] A. I. Spriggs, R. A. Diamond, and E. W. Meyer, "Automated screening for cervical smears?" *The Lancet*, vol. 1, no. 7538, pp. 359–360, 1968.

[13] S. Watanabe, "An automated apparatus for cancer prescreening: cybest," *CGIP*, vol. 3, no. 4, pp. 350–358, 1974.

[14] N. Tanaka, T. Ueno, H. Ikeda et al., "CYBEST Model 4. Automated cytologic screening system for uterine cancer utilizing image analysis processing," *Analytical and Quantitative Cytology and Histology*, vol. 9, no. 5, pp. 449–454, 1987.

[15] D. J. Zahniser, P. S. Oud, M. C. Raaijmakers, G. P. Vooys, and R. T. van de Walle, "BioPEPR: a system for the automatic prescreening of cervical smears," *Journal of Histochemistry and Cytochemistry*, vol. 27, no. 1, pp. 635–641, 1979.

[16] R. Erhardt, E. R. Reinhardt, W. Schlipf, and W. H. Bloss, "FAZYTAN: a system for fast automated cell segmentation, cell image analysis and feature analysis and feature extraction based on TV-image pickup and parallel processing," *Analytical and Quantitative Cytology*, vol. 2, no. 1, pp. 25–40, 1980.

[17] J. H. Tucker and O. A. Husain, "Trials with the Cerviscan experimental prescreening device on polylysine-prepared slides," *Analytical and Quantitative Cytology*, vol. 3, no. 2, pp. 117–120, 1981.

[18] V. N. Ploem JS and A. M. J. van Driel-Kulker, "Leytas—a cytology screening system using the new modular image analysis computer (miac) from leitz," in *Clinical Cytometry and Histometry*, G. Burger, J. S. Ploem, and K. Goerttler, Eds., Academic Press, London, UK, 1987.

[19] B. Nordin, *The development of an automatic prescreener for the early detection of cervical cancer: algorithms and implementation [Ph.D. thesis]*, Uppsala University, Uppsala, Sweden, 1989.

[20] L. J. Mango, "Computer-assisted cervical cancer screening using neural networks," *Cancer Letters*, vol. 77, no. 2-3, pp. 155–162, 1994.

[21] L. G. Koss, "The Papanicolaou test for cervical cancer detection. A triumph and a tragedy," *Journal of the American Medical Association*, vol. 261, no. 5, pp. 737–743, 1989.

[22] M. L. Hutchinson, C. M. Cassin, and H. G. Ball III, "The efficacy of an automated preparation device for cervical cytology," *The American Journal of Clinical Pathology*, vol. 96, no. 3, pp. 300–305, 1991.

[23] L. P. Howell, R. L. Davis, T. I. Belk, R. Agdigos, and J. Lowe, "The AutoCyte preparation system for gynecologic cytology," *Acta Cytologica*, vol. 42, no. 1, pp. 171–177, 1998.

[24] S. J. Bernstein, L. Sanchez-Ramos, and B. Ndubisi, "Liquid-based cervical cytologic smear study and conventional papanicolaou smears: a metaanalysis of prospective studies comparing cytologic diagnosis and sample adequacy," *The American Journal of Obstetrics and Gynecology*, vol. 185, no. 2, pp. 308–317, 2001.

[25] BD Focal Point, 2014, http://www.bd.com/tripath/labs/fp_screening.asp.

[26] T. F. Kardos, "The FocalPoint system: FocalPoint slide profiler and FocalPoint GS," *Cancer*, vol. 102, no. 6, pp. 334–339, 2004.

[27] Thinprep, 2014, http://www.thinprep.com/hcp/about_imaging_system.html/.

[28] M. R. Quddus, T. Neves, M. Reilly, M. Steinhoff, and C. Sung, "Does the ThinPrep Imaging System increase the detection of high-risk HPV-positive ASC-US and AGUS the Women and Infants Hospital experience with over 200,000 cervical cytology cases," *CytoJournal*, vol. 6, article 15, 2009.

[29] H. K. Grohs and O. A. N. Husain, *Automated Cervical Cancer Screening*, Igaku-Shoin, New York, NY, USA, 1994.

[30] K. R. Lee, R. Ashfaq, G. G. Birdsong, M. E. Corkill, K. M. McIntosh, and S. L. Inhorn, "Comparison of conventional papanicolaou smears and a fluid-based, thin- layer system for cervical cancer screening," *Obstetrics and Gynecology*, vol. 90, no. 2, pp. 278–284, 1997.

[31] J. Linder and D. Zahniser, "The ThinPrep pap test: a review of clinical studies," *Acta Cytologica*, vol. 41, no. 1, pp. 30–38, 1997.

[32] J. Monsonego, A. Autillo-Touati, C. Bergeron et al., "Liquid-based cytology for primary cervical cancer screening: a multi-centre study," *British Journal of Cancer*, vol. 84, no. 3, pp. 360–366, 2001.

[33] J. Karnon, J. Peters, J. Platt, J. Chilcott, E. McGoogan, and N. Brewer, "Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic anlysis," *Health Technology Assessment*, vol. 8, no. 20, pp. 1–78, 2004.

[34] B. Strander, A. Andersson-Ellström, I. Milsom, T. Rådberg, and W. Ryd, "Liquid-based cytology versus conventional Papanicolaou smear in an organized screening program : a prospective randomized study," *Cancer Cytopathology*, vol. 111, no. 5, pp. 285–291, 2007.

[35] M. Arbyn, C. Bergeron, P. Klinkhamer, P. Martin-Hirsch, A. G. Siebers, and J. Bulten, "Liquid compared with conventional cervical cytology: a systematic review and meta-analysis," *Obstetrics and Gynecology*, vol. 111, no. 1, pp. 167–177, 2008.

[36] H. Kuramoto, Y. Iwami, N. Sugimoto, C. Kato, T. Sugahara, and M. Iida, "Application of a new liquid-based procedure (TACAS) for the screening of cervical cancer: a preliminary study," *Acta Cytologica*, vol. 56, no. 1, pp. 74–79, 2012.

[37] Cell Solutions, 2014, http://www.cellsols.com/.

[38] "Aperio slide scanner," 2014, http://www.leicabiosystems.com/.

[39] R. Poulsen, *Automated prescreening of cervical cytology specimens [Ph.D. thesis]*, McGill University, Montreal, Canada, 1973.

[40] J. M. Prewitt and M. L. Mendelsohn, "The analysis of cell images," *Annals of the New York Academy of Sciences*, vol. 128, no. 3, pp. 1035–1053, 1966.

[41] C. Bergmeir, M. García Silvente, J. Esquivias López-Cuervo, and J. M. Benítez, "Segmentation of cervical cell images using mean-shift filtering and morphological operators," in *Medical Imaging 2010: Image Processing*, San Diego, Calif, USA, February 2010, 76234C.

[42] C. Bergmeir, M. García Silvente, and J. M. Benítez, "Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 497–512, 2012.

[43] P. Bamford and B. Lovell, "Unsupervised cell nucleus segmentation with active contours," *Signal Processing*, vol. 71, no. 2, pp. 203–213, 1998.

[44] P. Malm and A. Brun, "Closing curves with Riemannian dilation: application to segmentation in automated cervical cancer screening," in *Advances in Visual Computing*, vol. 5875 of *Lecture Notes in Computer Science*, pp. 337–346, 2009.

[45] E. Meijering, "Cell segmentation: 50 years down the road [life sciences]," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 140–145, 2012.

[46] P. Malm, *Image analysis in support of computer-assisted cervical cancer screening [Ph.D. thesis]*, Division of Visual Information and Interaction, Uppsala University, Uppsala, Sweden, 2013.

[47] K. Rodenacker and E. Bengtsson, "A feature set for cytometry on digitized microscopic images," *Analytical Cellular Pathology*, vol. 25, no. 1, pp. 1–36, 2003.

[48] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[49] B. E. Bejnordi, R. Moshavegh, K. Sujathan, P. Malm, E. Bengtsson, and A. Mehnert, "Novel chromatin texture features for

the classification of pap smears," in *SPIE Medical Imaging 2013*, February 2013.

[50] A. Doudkine, C. Macaulay, N. Poulin, and B. Palcic, "Nuclear texture measurements in image cytometry," *Pathologica*, vol. 87, no. 3, pp. 286–299, 1995.

[51] B. S. Apgar, L. Zoschnick, and T. C. Wright Jr., "The 2001 Bethesda System Terminology," *The American Family Physician*, vol. 68, no. 10, pp. 1992–1998, 2003.

[52] K. R. Castleman and B. S. White, "Optimizing cervical cell classifiers," *Analytical and Quantitative Cytology*, vol. 2, no. 2, pp. 117–122, 1980.

[53] P. H. Bartels, D. Thompson, and J. E. Weber, "Diagnostic decision support by inference networks," *In Vivo*, vol. 7, no. 4, pp. 379–386, 1993.

[54] H. E. Nieburgs, A. F. Goldberg, B. Bertini, J. Silagi, B. Pacheco, and H. Reisman, "Malignancy associated changes (MAC) in blood and bone marrow cells of patients with malignant tumors," *Acta Cytologica*, vol. 11, no. 5, pp. 415–423, 1967.

[55] G. Burger, U. Jutting, and K. Rodenacker, "Changes in benign cell populations in cases of cervical cancer and its precursors," *Analytical and Quantitative Cytology*, vol. 3, no. 4, pp. 261–271, 1981.

[56] G. L. Wied, P. H. Bartels, M. Bibbo, and J. J. Sychra, "Cytomorphometric markers for uterine cancer in intermediate cells," *Analytical and Quantitative Cytology*, vol. 2, no. 4, pp. 257–263, 1980.

[57] M. Guillaud, A. Doudkine, D. Garner, C. MacAulay, and B. Palcic, "Malignancy associated changes in cervical smears: Systematic changes in cytometric features with the grade of dysplasia," *Analytical Cellular Pathology*, vol. 9, no. 3, pp. 191–204, 1995.

[58] M. Guillaud, D. Cox, K. Adler-Storthz et al., "Exploratory analysis of quantitative histopathology of cervical intraepithelial neoplasia: Objectivity reproducibility, malignancy-associated changes, and human papillomavirus," *Cytometry A*, vol. 60, no. 1, pp. 81–89, 2004.

[59] G. Anderson, C. Macaulay, J. Matisic, D. Garner, and B. Palcic, "The use of an automated image cytometer for screening and quantitative assessment of cervical lesions in the British Columbia Cervical Smear Screening Programme," *Cytopathology*, vol. 8, no. 5, pp. 298–312, 1997.

[60] X. R. Sun, J. Wang, D. Garner, and B. Palcic, "Detection of cervical cancer and high grade neoplastic lesions by a combination of liquid-based sampling preparation and DNA measurements using automated image cytometry," *Cellular Oncology*, vol. 27, no. 1, pp. 33–41, 2005.

[61] M. Guillaud, J. L. Benedet, S. B. Cantor, G. Staerkel, M. Follen, and C. MacAulay, "DNA ploidy compared with human papilloma virus testing (Hybrid Capture II) and conventional cervical cytology as a primary screening test for cervical high-grade lesions and cancer in 1555 patients with biopsy confirmation," *Cancer*, vol. 107, no. 2, pp. 309–318, 2006.

[62] "Innovative design and development of cancer screening solutions and public services," 2014, http://www.med.motic.com/.

[63] P. Naucler, W. Ryd, S. Törnberg et al., "Human papillomavirus and Papanicolaou tests to screen for cervical cancer," *The New England Journal of Medicine*, vol. 357, no. 16, pp. 1589–1597, 2007.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 120182, 4 pages
http://dx.doi.org/10.1155/2014/120182

Research Article

Tract-Based Bayesian Multivariate Analysis of Mild Traumatic Brain Injury

Yongkang Liu,[1] Tianyao Wang,[2] Xiao Chen,[2] Jianhua Zhang,[2] Guoxing Zhou,[2] Zhongqiu Wang,[1] and Rong Chen[3]

[1] *Department of Radiology, Affiliated Hospital of Nanjing University of Traditional Chinese Medicine, 155 Hanzhong Road, Nanjing, Jiangsu 210029, China*
[2] *Department of Radiology, East Hospital, Tongji University School of Medicine, Shanghai 200120, China*
[3] *Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA*

Correspondence should be addressed to Zhongqiu Wang; zhq2001us@163.com and Rong Chen; rchen@umm.edu

Received 20 November 2013; Accepted 20 January 2014; Published 10 March 2014

Academic Editor: Yuanjie Zheng

Purpose. Detecting brain regions characterizing mild traumatic brain injury (mTBI) by combining Tract-Based Spatial Statistics (TBSS) and Graphical-model-based Multivariate Analysis (GAMMA). *Materials and Methods*. This study included 39 mTBI patients and 28 normal controls. Local research ethics committee approved this prospective study. Diffusion-tensor imaging was performed in mTBI patients within 7 days of injury. Skeletonized fractional anisotropy (FA) maps were generated by using TBSS. Brain regions characterizing mTBI were detected by GAMMA. *Results*. Two clusters of lower frontal white matter FA were present in mTBI patients. We constructed classifiers based on FA values in these two clusters to differentiate mTBI and controls. The mean accuracy, sensitivity, and specificity, across five different classifiers, were 0.80, 0.94, and 0.61, respectively. *Conclusions*. Combining TBSS and GAMMA can detect neuroimaging biomarkers characterizing mTBI.

1. Introduction

More than 1.125 million people experience a mild traumatic brain injury (mTBI) each year in the United States [1]. 7-8% of mTBI patients suffer from chronic symptoms [2]. In a one-year follow-up study, Van der Naalt et al. found that 84% still displayed mTBI symptoms including headaches, irritability, memory problems, poor concentration, and fatigue [3].

Computed tomography (CT) and conventional magnetic resonance (MR) imaging results of mTBI are typically normal. Diffusion-tensor imaging (DTI) examines the molecular diffusion of water and can measure white matter microstructural integrity noninvasively. Water diffuses more readily along the direction of axonal fibers. The diffusion profile in each voxel can be measured by DTI. One of the most commonly used DTI-based feature maps is the Fractional Anisotropy (FA) map, which describes the degree of directionality of diffusion. In mTBI, DTI has been used to identify microstructural changes that cannot be detected by CT or conventional MR [4].

Tract-Based Spatial Statistics (TBSS) is an automated whole-brain analysis method which aims to address two problems in voxel-based analysis of DTI data, the alignment and smoothing issue. TBSS projects a subject's FA map to a common space, creates the mean FA image and its skeleton, and projects each subject's FA onto the skeleton. This results in a skeletonized FA image for each subject. TBSS achieves alignment between the FA skeleton and a subject's FA map without requiring perfect nonlinear registration and does not require smoothing. Therefore, TBSS could improve the sensitivity, objectivity, and interpretability of the group-level analysis of DTI data. Several studies used TBSS to examine white matter changes in mTBI [5–7].

Graphical-model-based Multivariate Analysis (GAMMA) [8, 9] is a group-level analysis method to detect linear/nonlinear interactions among brain regions and a clinical

variable. Let C denote the clinical variable. C could be a group membership variable which represents presence or absence of a disease, or a demographic variable. The input to GAMMA is image-derived feature maps which are defined in the same stereotaxic space and contains potential biomarkers of C. GAMMA detects a set of brain regions which are jointly predictive of C. GAMMA is fully automatic and does not rely on any assumption about the structure form of such interactions. It has been used in brain morphometry [10], functional MR data analysis [11, 12], and lesion-deficit analysis [13].

Combining TBSS and GAMMA, we are able to detect interactions among brain regions and the clinical variable in the FA skeleton space. We propose a novel analytic method, which combines TBSS and GAMMA, for the detection of brain regions characterizing mTBI.

2. Methods

2.1. Subjects. Local research ethics committee approved this prospective study. 39 mTBI patients and 28 normal controls were recruited from the emergency department in Shanghai Dongfang Hospital, Shanghai, China, between February 2013 and August 2013.

The diagnosis of mTBI was established by using the criteria of the American Congress of Rehabilitative Medicine for mild brain injury [14]. The exclusion criteria were (1) history of significant ear surgery, (2) penetrating head injury, (3) pregnancy, (4) history of dementia or mental disorder, (5) uremia, liver cirrhosis, heart failure, pulmonary edema, coagulopathy, and renal dysfunction, (6) ischemic and hemorrhagic stroke, (7) in vivo magnetic implants (such as iron, or with cochlear implants, vascular clips, etc.) or with pacemaker, and (8) the patient being either dead or having already received cardiopulmonary resuscitation before arrival at hospital.

The control group included healthy subjects who had no neurological or psychiatric illness and no prior TBI.

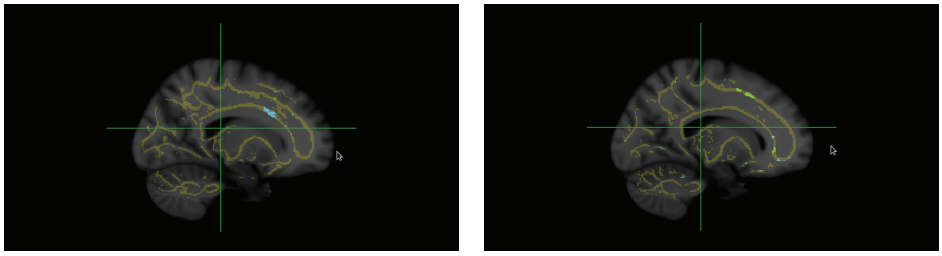
2.2. Data Acquisition and Imaging Parameters. All MR images were acquired with a Philips Achieva 3.0T TX MRI scanner (Royal Philips, Amsterdam, Netherlands). Diffusion-tensor images were acquired with a single-shot echo-planar sequence (TR/TE = 9,000 ms/90 ms, slice thickness = 2 mm, voxel size = 2 mm $*$ 2 mm, and field of view = 256 $*$ 256 mm). Diffusion gradients were set in 32 noncollinear directions by using two b values (b = 0 and 1,000 s/mm^2). Diffusion-tensor imaging was performed in mTBI patients within 7 days of injury.

2.3. DTI Data Processing. The diffusion-weighted images were preprocessed by using FMRIB Software Library [15]. The diffusion-weighted data were registered to the b0 image using an affine registration algorithm in order to minimize distortion due to motion and eddy currents. Brain Extraction Tool [16] was used to remove nonbrain tissues in the T1- and diffusion-weighted data. Skull-stripped images were visually inspected and, if necessary, manually corrected for skull-stripping error. FA images were generated by using the Diffusion Toolbox [15].

The procedure to generate skeletonized FA images was as follows. First, all FA maps were normalized to the widely used FMRIB58 FA template using the nonlinear registration algorithm in FSL [15]. Then the mean of all FA maps was created by averaging normalized FA maps. The mean FA map was the input to the tract skeleton generation step, which aims to represent all tracts common to all subjects. The skeleton of a tract is a single line (or surface) running down the center of the tract. The FA skeleton was thresholded with FA > 0.2 to exclude voxels which are primarily gray-matter or cerebrospinal fluid. The last step was to project individual subject's FA onto the skeleton. At each point of a skeleton, the maximum FA value in the perpendicular tract direction was the value of this point.

2.4. Graphical-Model-Based Multivariate Analysis. GAMMA is a machine learning method for biomarker detection. It is based on two principles of brain functional organization: functional segregation and integration. GAMMA models the associations among a set of brain regions and a clinical variable C as a Bayesian network. In this study, the clinical variable represents whether a participant has mTBI (C = 1) or is a normal control (C = 0). GAMMA is a voxel-based method. There are two main tasks in GAMMA: voxel-space partition and Markov blanket detection. In voxel-space partition, GAMMA groups voxels into functional equivalent regions. A Markov blanket of C is variables which are jointly most predictive of C. Given the Markov blanket of C, knowing the states of other variables provides no additional information about C. Therefore, variables in the Markov blanket of C are biomarkers of C. GAMMA uses a specific type of Bayesian network called Bayesian network with inverse-tree structure. The output of GAMMA is a label field which defines a set of brain regions and a Bayesian network which describes the associations among these brain regions and the clinical variable. Each region of interest (ROI) in the label field can be represented by a single variable which represents the regional state. Then we can predict C using these regional states.

The input to GAMMA is the skeletonized FA maps. These skeletonized FA maps are defined in the Montreal Neurological Institute (MNI) space. For each skeletonized FA map, we aimed to generate a binary effect map, in which voxels with value 1 represent FA reduction, and voxels with value 0 represent no FA reduction. The procedures to generate the binary effect map were as follows. First, we calculated voxelwise 40th percentile values, based on the skeletonized FA maps of all subjects in the normal control group, and generated a threshold map $\mathbf{T}$. Second, we generated a binary effect map for each subject. Let $V_i(j)$ and T_i denote the signal intensity of voxel i in the skeletonized FA map j and the threshold map, respectively. If $V_i(j) < T_i$, the value of generated binary map at voxel i is 1; otherwise, it is 0. We used the GAMMA suite v1.2 (http://www.nitrc.org/projects/gamma_suite/) to perform GAMMA analysis.



(a) (b)

Figure 1: Voxels characterizing mTBI are shown in blue (ROI 1) and green (ROI 2). The white matter skeleton generated by TBSS is shown in yellow. ROIs are shown in the sagittal view and overlaid on the MNI152 template. (a) MNI coordinates $X = 18$; (b) MNI coordinates $X = -16$.

3. Results

This study included 67 subjects (39 mTBI and 28 normal controls). The mean ages of the mTBI and control group were 31 (standard deviation (SD) 7.4) and 33 (SD 9.4), respectively. There was no significant difference in the mean baseline age (P value = 0.319 based on two-sample t-test). The number of female participants of mTBI participants was 11 (total number of subjects in the group = 39) and that of normal controls was 13 (total number of subjects in the group = 28). There was no significant difference in the proportion of female participants (P value = 0.125).

GAMMA detected two ROIs characterizing mTBI. These two ROIs are depicted in Figure 1. ROI 1 is centered on the right frontal lobe, and ROI 2 is centered on the left frontal lobe. Relative to normal controls, participants in the mTBI group demonstrated reduced FA values in these two ROIs.

Each ROI had a set of voxels. We used the regional state inference (RSI) algorithm in [17] to infer the regional states of a ROI. RSI infers the regional state using a latent-variable model, with an online Gibbs sampling algorithm. The regional state of a ROI is a biomarker characterizing mTBI. Then we constructed predictive models to differentiate mTBI and controls based on two biomarkers, regional states of ROI 1 and 2. We constructed different kinds of classifiers [18] in order to avoid the bias associated with a specific type of classifier. Classification performance was evaluated using 10-fold cross-validation. Table 1 lists accuracies, sensitivities, and specificities of different kinds of classifiers. We found that these two biomarkers can predict C with mean accuracy = 0.80, sensitivity = 0.94, and specificity = 0.61.

Table 1: Accuracies, sensitivities, and specificities of different kinds of classifiers.

Classifier	Accuracy	Sensitivity	Specificity
Logistic model trees [21]	82	97	61
AdaBoost [22]	77	90	61
Bagging [23]	82	97	61
Naïve BN [24]	79	92	61
Support vector machine [25]	79	92	61
Mean	80	94	61

4. Conclusion and Discussion

We found that biomarkers detected by combining GAMMA and TBSS accurately differentiated mTBI patients and controls. The mean accuracy, sensitivity, and specificity, across five different classifiers, were 0.80, 0.94, and 0.61, respectively.

We found that mTBI participants have decreased FA in two ROIs mainly in the frontal lobe than that in controls. White matter injury in the frontal lobe is consistently reported in TBI studies [19]. In a TBSS study of 51 mTBI patients and 50 controls, Wada et al. reported that patients with mTBI in the chronic stage had decreased FA in the superior frontal gyrus, superior longitudinal fasciculus, insula, and fornix [6]. In [20], Kraus et al. analyzed DTI data of 20 mTBI patients and 18 controls and found decreased FA in the corticospinal tract, sagittal stratum, and superior longitudinal fasciculus for the mTBI group. Our finding also suggested that DTI is sensitive to detect white matter injury in the frontal lobe in mTBI patients.

Identifying neuroimaging biomarkers for diagnosis or prognosis is of great importance. Such neuroimaging biomarkers can be identified based on expert knowledge or machine learning algorithms. The advantage of using machine learning algorithms for biomarker detection is that it can detect biomarkers in an automated, unbiased manner. GAMMA is a Bayesian machine learning method for biomarker detection. Our results suggested that GAMMA can automatically detect biomarkers in the skeletonized FA space.

One limitation of this preliminary study is the small sample size. The generated predictive model was accurate (accuracy = 0.80, sensitivity = 0.94, and specificity = 0.61). The classification performance is evaluated using tenfold cross-validation. However, we did not validate this model using an independent test data set because this study had a limited number of subjects. In future, we plan to evaluate the predictive model generated in this study in a cohort with larger sample size. We will split the data set into training and testing data sets. This could provide more reliable estimation of the generalizability of the predictive model. The predictive model generated using data from a cohort with larger sample size may have higher specificity than the current one.

In conclusion, we found that combining TBSS and GAMMA can detect neuroimaging biomarkers characterizing mTBI.

Conflict of Interests

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

Authors' Contribution

Yongkang Liu and Tianyao Wang contribute equally to this work and are considered co-first authors.

References

[1] CDC, *Report to Congress on Mild Traumatic Brain Injury in the United States: Steps to Prevent a Serious Public Health Problem*, 2003.

[2] P. Nolin and L. Heroux, "Relations among sociodemographic, neurologic, clinical, and neuropsychologic variables, and vocational status following mild traumatic brain injury: a follow-up study," *Journal of Head Trauma Rehabilitation*, vol. 21, no. 6, pp. 514–526, 2006.

[3] J. van der Naalt, A. H. van Zomeren, W. J. Sluiter, and J. M. Minderhoud, "One year outcome in mild to moderate head injury: the predictive value of acute injury characteristics related to complaints and return to work," *Journal of Neurology Neurosurgery and Psychiatry*, vol. 66, no. 2, pp. 207–213, 1999.

[4] S. N. Niogi and P. Mukherjee, "Diffusion tensor imaging of mild traumatic brain injury," *Journal of Head Trauma Rehabilitation*, vol. 25, no. 4, pp. 241–255, 2010.

[5] K. M. Kinnunen, R. Greenwood, J. H. Powell et al., "White matter damage and cognitive impairment after traumatic brain injury," *Brain*, vol. 134, part 2, pp. 449–463, 2011.

[6] T. Wada, Y. Asano, and J. Shinoda, "Decreased fractional anisotropy evaluated using tract-based spatial statistics and correlated with cognitive dysfunction in patients with mild traumatic brain injury in the chronic stage," *The American Journal of Neuroradiology*, vol. 33, no. 11, pp. 2117–2122, 2012.

[7] S. Fakhran, K. Yaeger, and L. Alhilali, "Symptomatic white matter changes in mild traumatic brain injury resemble pathologic features of early Alzheimer dementia," *Radiology*, vol. 269, no. 1, pp. 249–257, 2013.

[8] R. Chen and E. H. Herskovits, "Graphical-model-based morphometric analysis," *IEEE Transactions on Medical Imaging*, vol. 24, no. 10, pp. 1237–1248, 2005.

[9] R. Chen and E. H. Herskovits, "Graphical Model Based Multivariate Analysis (GAMMA): an open-source, cross-platform neuroimaging data analysis software package," *Neuroinformatics*, vol. 10, no. 2, pp. 119–127, 2012.

[10] R. Chen, M. A. Pawlak, T. B. Flynn, J. Krejza, E. H. Herskovits, and E. R. Melhem, "Brain morphometry and intelligence quotient measurements in children with sickle cell disease," *Journal of Developmental and Behavioral Pediatrics*, vol. 30, no. 6, pp. 509–517, 2009.

[11] R. Chen and E. H. Herskovits, "Graphical-model-based multivariate analysis of functional magnetic-resonance data," *NeuroImage*, vol. 35, no. 2, pp. 635–647, 2007.

[12] R. Chen and E. H. Herskovits, "Clinical diagnosis based on bayesian classification of functional magnetic-resonance data," *Neuroinformatics*, vol. 5, no. 3, pp. 178–188, 2007.

[13] R. Chen, A. E. Hillis, M. Pawlak, and E. H. Herskovits, "Voxelwise Bayesian lesion-deficit analysis," *NeuroImage*, vol. 40, no. 4, pp. 1633–1642, 2008.

[14] ACRM, "Definition of mild traumatic brain injury," *Journal of Head Trauma Rehabilitation*, vol. 8, no. 3, pp. 86–87, 1993.

[15] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, pp. 782–790, 2012.

[16] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002.

[17] R. Chen and E. H. Herskovits, "A bayesian network classifier with inverse tree structure for voxelwise magnetic resonance image analysis," in *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '05)*, pp. 4–12, August 2005.

[18] R. Duda, R. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2001.

[19] G. Zappalà, M. Thiebaut de Schotten, and P. J. Eslinger, "Traumatic brain injury and the frontal lobes: what can we gain with diffusion tensor imaging?" *Cortex*, vol. 48, no. 2, pp. 156–165, 2012.

[20] M. F. Kraus, T. Susmaras, B. P. Caughlin, C. J. Walker, J. A. Sweeney, and D. M. Little, "White matter integrity and cognition in chronic traumatic brain injury: a diffusion tensor imaging study," *Brain*, vol. 130, part 10, no. 10, pp. 2508–2519, 2007.

[21] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Machine Learning*, vol. 59, no. 1-2, pp. 161–205, 2005.

[22] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," in *Proceedings of the 30th International Conference on Machine Learning*, pp. 148–156, 1996.

[23] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[24] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2-3, pp. 131–163, 1997.

[25] V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, NY, USA, 1998.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 973972, 9 pages
http://dx.doi.org/10.1155/2014/973972

Research Article

Enhancing the Detection of BOLD Signal in fMRI by Reducing the Partial Volume Effect

Yiping P. Du,[1,2,3] Renxin Chu,[3] and Jason R. Tregellas[3,4]

[1] *Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China*
[2] *Key Laboratory for Biomedical Engineering of Education Ministry of China, China*
[3] *Department of Psychiatry, Brain Imaging Center, University of Colorado School of Medicine, Aurora, CO, USA*
[4] *Research Service, Denver VA Medical Center, Denver, CO 80220, USA*

Correspondence should be addressed to Yiping P. Du; yipingdu@zju.edu.cn

Received 26 October 2013; Accepted 1 February 2014; Published 9 March 2014

Academic Editor: Yuanjie Zheng

Purpose. To investigate the advantages of reducing the partial volume effect (PVE) to enhance the detection of the BOLD signal in fMRI. *Methods*. A linear phase term was added in k-space to obtain half-voxel shifting of 64×64 T_2^*-weighted echo-planar images. Three sets of image data shifted in the x, y, and diagonal direction, respectively, are combined with the original 64×64 data to form the 128×128 voxel-shifted interpolated data. *Results*. A simulation of a synthetic fMRI dataset shows that the voxel-shifted interpolation (VSI) can increase the t-score up to 50% in single-voxel activations. An fMRI study ($n = 7$) demonstrates that 20.4% of the interpolated voxels have higher t-scores than their nearest neighboring voxels in the original maps. The average increase of the t-score in these interpolated voxels is 13.3%. *Conclusion*. VSI yields increased sensitivity in detecting voxel-size BOLD activations, improved spatial accuracy of activated regions, and improved detection of the peak BOLD signal of an activated region. VSI can potentially be used as an alternative to the high-resolution fMRI studies in which reduction in SNR and increase in imaging time become prohibitive.

1. Introduction

Functional MRI (fMRI) studies with the blood oxygenation level-dependent (BOLD) contrast [1, 2] typically use a spatial resolution of a few millimeters (e.g., with an in-plane resolution of 2–4 mm and through-plane resolution of 3–6 mm). It is desirable to increase the spatial resolution in fMRI for at least two reasons. First, the spatial localization of neuronal activation can be improved with a higher spatial resolution [3, 4]. The gray matter (GM) typically has a thickness of 2-3 mm in human brains. Even within a cortical layer of GM, the distribution of venous vessels is heterogeneous and, therefore, the BOLD contrast can be increased by using a voxel size as small as $1\,\text{mm}^3$ [5–7]. Secondly, the BOLD response originating from sub-voxel activation can be increased by reducing the voxel size. For example, it was demonstrated that BOLD activation in the hippocampal formation is often undetectable at a resolution of $4 \times 4 \times 5\,\text{mm}^3$ but can be reliably detected at a higher resolution of $2 \times 2 \times 3\,\text{mm}^3$ [8]. In fMRI studies designed to examine midbrain nuclei [9], single-voxel BOLD activation is not unusual because of the small size of these nuclei.

The ability to detect small region (e.g., single-voxel size) BOLD activation is highly dependent on the location of the activation. The detected BOLD response is highest when the activation site is located at the center of a voxel and is lowest when the activation is located at the corners of 4 neighboring voxels in two-dimensional data (see Figure 1(a)) or 8 neighboring voxels in three-dimensional data. Similarly, the detected BOLD response at the edge of an activated region is dependent on the relative location of the edge to the voxels. This dependency of detected BOLD response on the relative location of the BOLD activation to the grid of voxels is determined by the voxel sensitivity function [10] and is referred to as the partial volume effect (PVE). PVE occurs in MRI when a voxel is only partially filled with the source of imaging contrast.

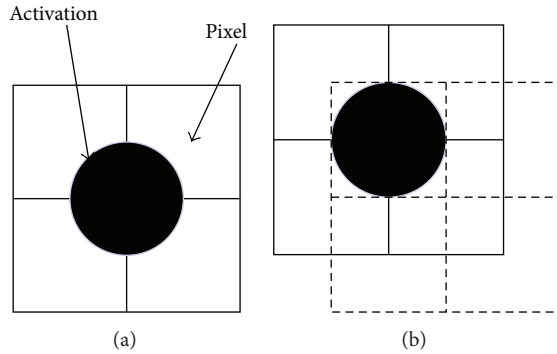


Figure 1: A scheme illustrates that the detection of BOLD signal is greatly affected by the location of the activated region relative to the voxels sampled when the region of activation, represented as a filled circle, is comparable to the size of a voxel, represented by the squares. When the distance of an activated region from its nearest voxel is close to half of the voxel size, the measured BOLD signal will be substantially reduced (a). Diagonally shifting the location of voxels by half of the voxel size, as shown in the dashed squares in (b), can greatly reduce the PVE and improve the detection of BOLD signal.

PVE can be reduced by acquiring images with high spatial resolution. High spatial resolution is commonly achieved by increasing the acquisition matrix of fMRI data, because any reduction of field-of-view (FOV) is limited by the dimension of the imaging objects or region of interest. Several studies have demonstrated the advantages of higher spatial resolution fMRI, including the increased spatial specificity of BOLD activations [4–6] and suppression of the BOLD signal originating from large vessels [3]. The acquisition of high-resolution fMRI data, however, inevitably reduces spatial SNR and prolongs acquisition time. With the current echo-planar imaging (EPI) or spiral imaging technology, a low resolution (e.g., with a 64×64 matrix) whole-brain dataset can be acquired with a repetition time (TR) on the order of 2 seconds [11]. Acquiring whole-brain fMRI with a higher spatial resolution (e.g., with a matrix size larger than 128×128) will considerably reduce the temporal resolution and spatial SNR, resulting in a greatly reduced statistical power in fMRI data analysis. This reduced statistical power can make the high-resolution approach unfavorable or even prohibitive to most fMRI studies. Furthermore, event-related fMRI studies require a TR of 2 seconds or shorter in order to adequately sample the hemodynamic response, making high-resolution acquisition even less useful for these studies. Additionally, image distortion induced by field inhomogeneity becomes more problematic in single-shot high-resolution echo-planar images [12].

In fMRI, PVE can be the result of three conditions: (1) when the size of the activated region is smaller than the voxel size, (2) when the peak of the true BOLD activation is mismatched from the location of any voxel, and (3) when the voxels at the boundary of activated regions are only partially occupied by the activation. When the dimension of an activation site is similar to the size of a voxel and the location of the activation is not centered on any one voxel (Figure 1(a)), the second and third conditions above can become the primary sources of PVE. In this case, the averaging of the BOLD signal with nonactivated partial volume greatly dilutes the BOLD contrast. When the BOLD contrast is diluted below certain threshold, the activated region becomes undetectable.

In this study, we hypothesize that spatial resampling with a half-voxel shift in the image space can greatly reduce PVE (Figure 1(b)) and enhance the ability to detect the activation through a better spatial match. In this paper, both simulation and fMRI experiments demonstrate that reducing PVE with voxel-shifted interpolation can substantially improve the detection of BOLD signal in fMRI.

2. Materials and Methods

2.1. Voxel-Shifted Interpolation. In the voxel-shift interpolation (VSI), three image sets with a 64×64 matrix were reconstructed with a half-voxel shift along the x, y, and diagonal direction, respectively, in addition to the original 64×64 image. According to the Fourier shift theorem, these shifts were implemented by applying a linear phase term in the complex k-space data in the corresponding directions:

$$\Phi(\kappa) = \frac{2\pi\kappa}{N}, \tag{1}$$

where half-integer $\kappa = -N/2 + 1/2, \ldots, N/2 - 1/2$ and N is the matrix size of the original image. Finally, the voxels in the three shifted image matrices were interspersed with those in the original image, based on the relative shift that was applied to each of the shifted image, to form an interpolated image with a 128×128 matrix size (see Figure 2).

VSI can also be implemented by filling zeroes in the k-space and is thus also referred to as zero-filled interpolation [13] or by sinc interpolation with complex images. Although the names of VSI, zero-filled interpolation, and sinc interpolation suggest different approaches of implementation, they are nevertheless mathematically equivalent. We use the term VSI in this paper, as it provides a pictorial description of the interpolation that can better illustrate the reduction of PVE.

2.2. Simulation. A synthetic dataset mimicking fMRI activation was generated for simulation. The data set had a matrix

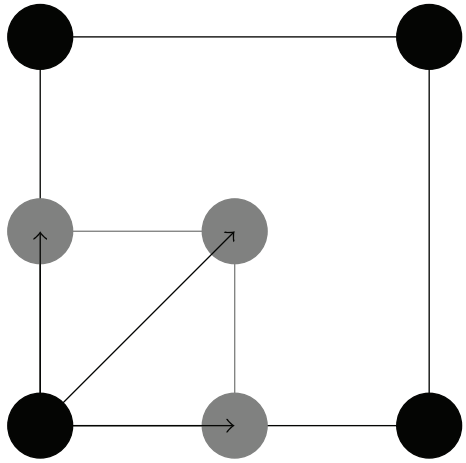


Figure 2: The scheme illustrates the voxel-shifting approach. The dark dots represent the centers of four original voxels. The gray dots represent the centers of three interpolated voxels shifted from the original voxel at the lower-left corner.

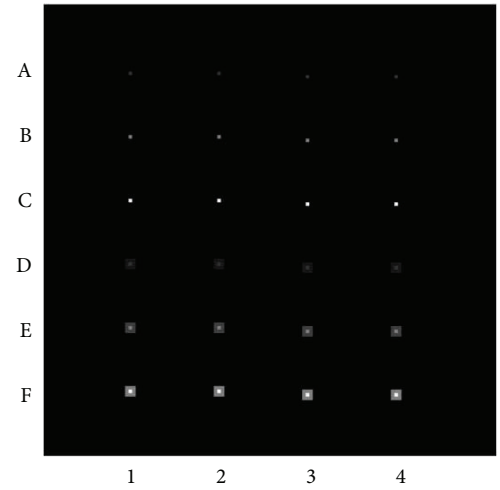


Figure 3: A synthetic data set (128 × 128) simulating fMRI activations of various sizes. Single-voxel "activations," with signal changes of 1%, 2.5%, and 5% were placed at rows A, B, and C, respectively. The 2D coordinates of activation at column 1 are odd, odd; column 2 are even, odd; column 3 are odd, even; and column 4 are even, even, respectively. Rows D to F show multivoxel "activations": the same single-voxel "activations" in rows A to C were duplicated, together with 8 surrounding "activated" voxels added to these single-voxel "activations." The signal of the surrounding voxels is half of that in the central voxel.

size of 128 × 128, with 240 time points, alternating "ON" states for 30 time points and "OFF" states for 30 time points. The background intensity was set at 10,000 in the real component and zero in the imaginary component. The data set had an SNR of 500 with Gaussian noise introduced in both real and imaginary components. The SNR of the synthetic data was intentionally set higher than that typically used in fMRI scans in order to better demonstrate the improvement in detecting "activation" through reducing PVE. In the "ON" state, a single-voxel "activation," with a signal of 1%, 2.5%, and 5%, respectively, was placed at 4 locations. The 2D coordinates of these 4 locations are in the combination of odd, odd; even, odd; odd, even; and even, even, respectively, as shown in the upper part of Figure 3. The same "activated" voxels were duplicated at the lower part of the figure, with surrounding "activated" voxels being added to these single-voxel "activations." The "activation" signal of each of the surrounding voxels was half of that in the central voxel. The simulation of the single-voxel "activation" was to mimic the activations in mid-brain nuclei, while the simulation of the multivoxel "activation" was to demonstrate the effect of PVE in cortical activations.

A two-dimensional Fourier transform was applied to the dataset to generate the k-space data. The central portion of the k-space data with a size of 64 × 64 was used to reconstruct the 64 × 64 *original* images. The same data set was also used to reconstruct the *shifted* images using VSI to form the 128 × 128 *interpolated* images. T-scores of each voxel in the original and the interpolated images were calculated using the general linear model in SPM2b software package (Wellcome Department of Cognitive Neurology, University College London, UK). A t-score of 3.18 was used to threshold the "activation" maps.

2.3. fMRI Experiments. Functional MRI data were acquired using T_2^*-weighted gradient-echo EPI on a 3T MRI scanner (General Electric, Waukesha, WI), while subjects (n = 7, male/female = 3/4, age = 36.3 ± 10.5 years old) were performing a smooth pursuit eye movement (SPEM) task [14–16], under a protocol approved by a local institutional review board. The imaging parameters were TE/TR/flip angle = 32 ms/2500 ms/77°, 24 slices, field-of-view = 22 cm, slice thickness = 4 mm, and matrix size = 64 × 64. The task was run twice; each consisted of task (30 s)/rest (30 s) for 4 cycles.

Functional images were reconstructed offline from the k-space data using a software package developed by our lab. The original image set with a 64 × 64 matrix was reconstructed directly from the acquired data. VSI was applied to the k-space data to obtain the 128 × 128 interpolated images.

VSI was also applied to the original *magnitude* images for comparison. The magnitude VSI is an approximation of the k-space VSI because the magnitude fMRI data are more accessible in scanners. The magnitude VSI can also be used to retrospectively interpolate the fMRI data acquired without saving the complex k-space data. In this approach, an original 64 × 64 magnitude image was used as the real component of the complex image and the imaginary component of the complex image was set to zero. Fourier transform was applied to the complex images to generate the 64 × 64 k-space data prior to VSI to a 128 × 128 matrix.

2.4. fMRI Data Analysis. fMRI data analysis was performed using the SPM2b software package. A t-score was calculated for each individual voxel. The same procedure was used in the data analysis to ensure the comparability of results between the original images with a 64 × 64 matrix and the interpolated images. The procedure is not involved in

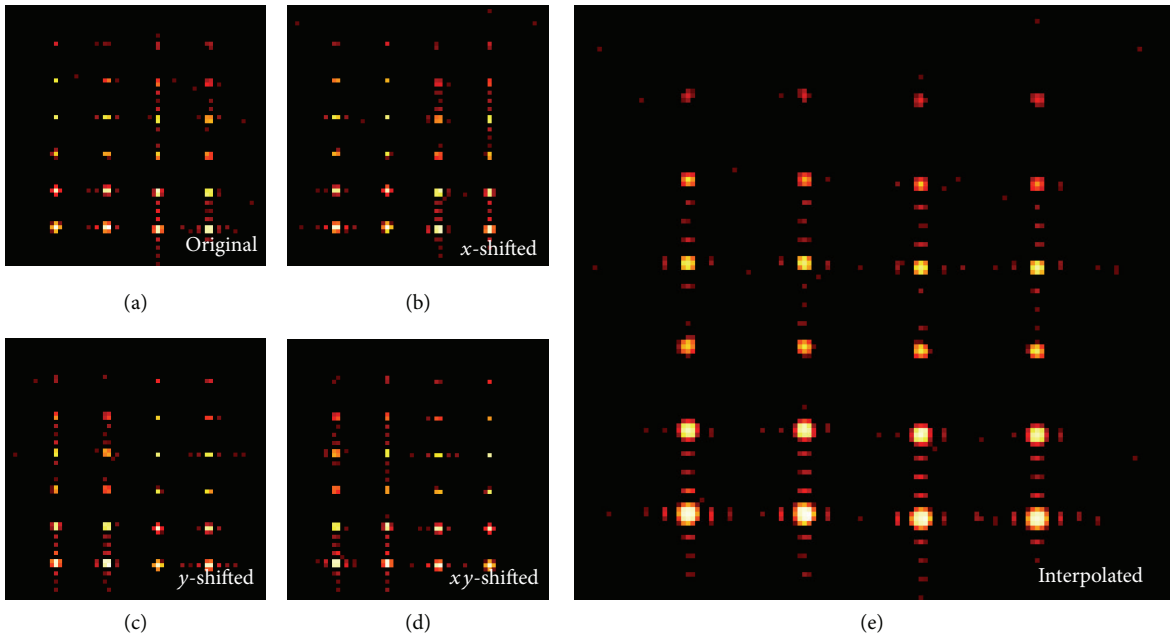


Figure 4: Simulation results of VSI represented as t-score maps of the "activations" using the template shown in Figure 3. (a) is the original 64 × 64 map (without VSI) obtained from the central portion (64 × 64) of the k-space of the synthetic data set shown in Figure 3, (b) is the 64 × 64 map shifted half-voxel along the horizontal direction, (c) is the 64 × 64 map shifted half-voxel along the vertical direction, and (d) is the 64 × 64 map shifted half-voxel along the diagonal direction. It is observed that the left column is blurred horizontally in (b), vertically in (c), and both horizontally and vertically in (d). The left column has the least blurring and typically has the highest t-scores in (a). The same column has the most severe burring and typically has the lowest t-scores in (d). The voxel-shifted interpolated 128 × 128 map (e) is the interspersed combination of (a), (b), (c), and (d). The location-dependent blurring and reduction of t-score seen in (a)–(d) was largely eliminated in (e). The SNR of the images is approximately 500 with Gaussian noise introduced to the complex k-space.

any spatial interpolation or smoothing. In order to focus on the comparison between the interpolated and original fMRI data and avoid unnecessary variables in the comparison, no other preprocessing procedures, such as motion correction or normalization to standard space, were performed prior to the fMRI data analysis. These steps involve additional spatial interpolation and can further complicate the comparisons of interest in this study. The data from one of the subjects was discarded due to excessive motion during data acquisition. A t-score of 3.18 ($P \leq 0.000022$) was used as the threshold of activation and the same scale of t-score (i.e., from 3.18 to 22.83) was used in the display of the statistical parametric mapping results.

3. Results

Simulation results show that the activated regions have different degrees of blurring in the *original* 64 × 64 images (see Figure 4(a)), depending on whether their x- and y-coordinates are even or odd numbers. The activations located at odd, odd coordinates (left column) have the least blurring and more likely have the highest t-scores. The activations located at even, even coordinates (right column) have the most severe blurring and more likely have the lowest t-scores. The activations located at even, odd coordinates are blurred horizontally and the activations located at odd, even coordinates are blurred vertically (see the middle two columns). These activations likely have intermediate t-scores. In the 64 × 64 images shifted along x-, y-, and xy-directions, similar spatial dependency of the t-scores on the relative location of the activations to the voxels is observed, except that the peak t-scores are at the second column from the left in the x-shifted image (Figure 4(b)), the third column from the left in the y-shifted image (Figure 4(c)), and the right column in the xy-shifted image (Figure 4(d)), respectively. In the interpolated 128 × 128 image (Figure 4(e)), this spatial dependency of t-scores is largely removed. The residual variation of the t-scores in different columns is caused by the added noise. As expected, all the activations were spatially blurred due to the 64 × 64 truncation of the k-space data prior to the interpolation.

Figure 5 is the scatter plot demonstrating the increase of peak t-scores in the interpolated 128 × 128 images (y-axis) compared to the peak t-score in the original and shifted 64 × 64 images (x-axis) for each of the 24 activations as in Figure 3. Up to a 50% increase in the maximum t-score of the 1% single-voxel activation was observed (see the lower-left corner of the plot). Substantial increase of the peak t-score can also occur with VSI in the single-voxel activations with higher BOLD signal and multivoxel activation with lower BOLD signal, as shown in the middle part of the plot. This increase of peak t-score with VSI becomes modest in multivoxel activations with high BOLD signal, as shown in the upper-right corner of the plot. These simulation results indicate that VSI can substantially improve the detection of small activated regions

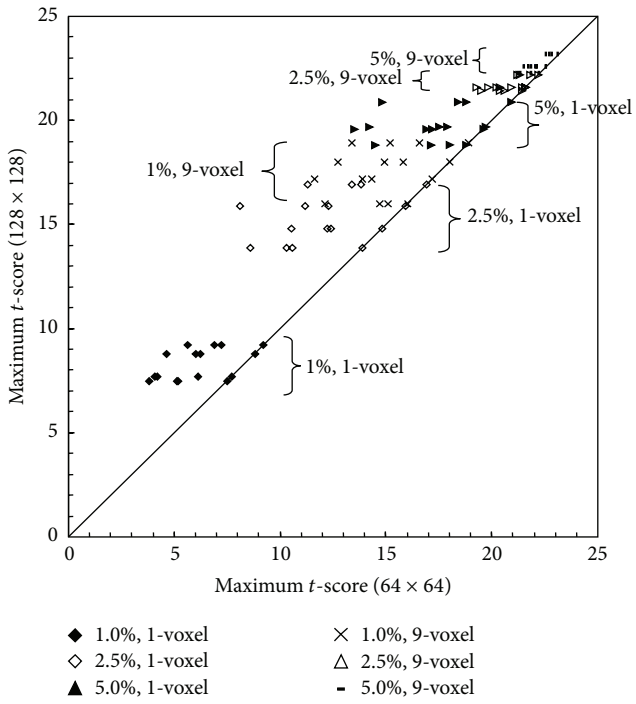


FIGURE 5: The measurement of peak t-score in each of the activations in the simulated data is shown in this plot. For each of the 24 activated regions, the y-axis represents the maximum t-score in the interpolated 128 × 128 map. The x-axis represents the maximum t-score in each of the four 64 × 64 maps. The highest of the four t-scores in each region is the same as the maximum t-score in that region in the interpolated map. This plot shows that the VSI can increase the t-score by up to 50% in single-voxel activations with a signal change of 1% (solid diamonds). The increase of t-score with VSI is substantial for single-voxel activations with a signal of up to 5% and 9-voxel activations with a signal change of 1%.

with a low BOLD signal. This feature is quite relevant in real fMRI studies, because small activated regions usually have a lower BOLD signal and are more likely undetectable in fMRI analysis.

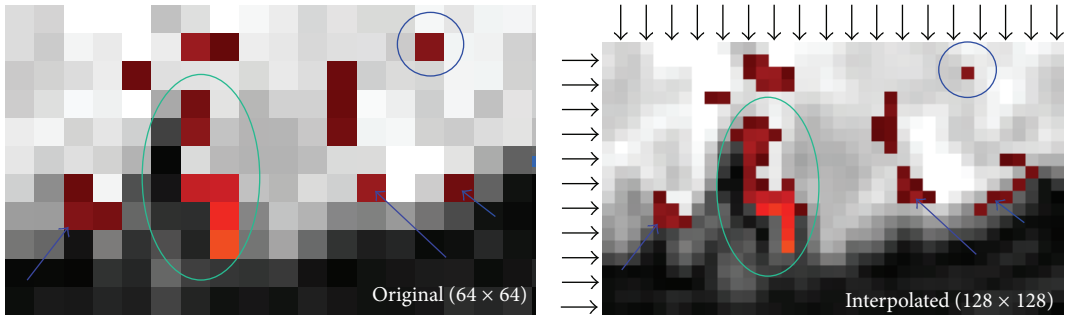
Human fMRI experiments also show substantial improvement using VSI to detect BOLD signal response. An example of the original (a) and interpolated (b) activation map at the visual cortex is shown in Figure 6. In this example, a single voxel activation in the original map can be a single-voxel (circle), a group of 6 voxels (short arrow), or a strip of 7 voxels (long arrow) in the interpolated map. Two apparently separated activated areas in the original map are connected in the interpolated map (ellipse). On the other hand, two similar 6-voxel activated regions in the interpolated map (long and medium arrows) appear as single-voxel and 3-voxel activated regions in the original map. In Figure 6(b), the voxels in the rows and columns indicated by the black arrows are obtained through VSI. The other voxels, one out of four in any 2 × 2 cluster, are from the original image.

The activation maps at the frontal eye fields (FEF) in 6 of the 7 subjects are shown in Figure 7. In this figure, the original maps are at the top, the maps interpolated with k-space VSI are in the second row, and the maps interpolated with magnitude VSI are at the third row. In subject 1, two apparently connected activated voxels in the original map are separated in the interpolated map (large blue circle). A single voxel activation in the original maps can be a single voxel activation (small blue circle in subject 1) or a group of 6 activated voxels (small green circle in subject 2) in the interpolated maps. In subject 3, the activated voxels apparently form an enclosed loop in the original map (large green circle) but not in the interpolated map. In general, the activation with VSI confirms better to the known gyral anatomical location of the FEF acquired with a high-resolution SPEM study [17].

Figure 8 shows the evaluation of the t-scores in the *k-space* interpolated and original activation maps in the FEF region in six subjects. There were 304 voxels that had t-scores above the threshold and thus considered as activated voxels. An interpolated voxel with a t-score higher than any of its two nearest horizontal or vertical original neighbors (labeled as Os in the inset drawing) is referred to as an A-type voxel (labeled as A), while an interpolated voxel with a higher t-score than any of its four nearest diagonal original neighbors (labeled as B) is referred to as a B-type voxel. Each circle in this plot shows an interpolated voxel that has a t-score (y-axis) higher than the highest t-score of its nearest original neighbors (x-axis). In the 6 interpolated maps, 20.4% of the voxels (i.e., 62 out of 304 voxels) are either A- (47) or B-type (15) voxels, with an average of 13.3% increase in t-score. In each of the 6 A-type interpolated voxels shown as dark solid circles, none of its nearest original neighbors has a t-score above the threshold of 3.18. These 6 interpolated voxels have t-scores above the threshold and become part of the multivoxel activated regions. Single-voxel activations in the interpolated maps are excluded in the statistics. Similar evaluation of the t-scores was applied to the *magnitude* interpolated activation maps (see Figure 9), with 311 activated voxels. Note that the signal in each voxel is slightly different between the *k-space* interpolated data and *magnitude* interpolated data. This evaluation demonstrates that 15.8% of the voxels (i.e., 49 out of 311 voxels) are either A- (38) or B-type (11) voxels, with an average of 14.4% increase of the t-scores. The location of 76% A-type (31) and B-type voxels (6) in the magnitude interpolated maps are overlapped with that in the k-space interpolated maps.

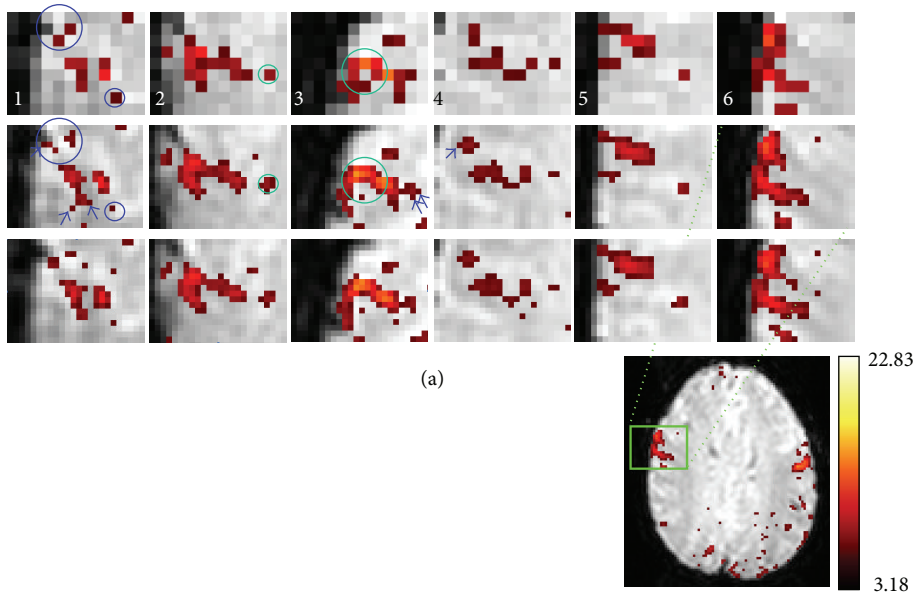
4. Discussion

4.1. VSI versus Other Spatial Interpolation Approaches. The shifted and original images are independently reconstructed from the same k-space data using the same algorithm. The only difference in the reconstruction of the shifted images is the added linear phase in k-space prior to the inverse Fourier transform. The shifted images can be considered to carry the same amount of information and as "true" as the original images. Therefore, the interpolated fMRI data is not expected to increase false positive detection when voxel-by-voxel fMRI analysis is performed. Also, the shifted images have the same SNR and physical voxel dimension as the original images. In comparison, other spatial interpolation methods use



(a) (b)

Figure 6: An example of the original (a) and k-space interpolated (b) activation maps at the visual cortex. The rows and columns indicated by the black arrows in the interpolated map consist of the voxels obtained through voxel-shifting. The interpolated map would be reduced to the original map if these arrow-indicated rows and columns were removed. In this example, a single voxel activation in the original map could be a single voxel (circle), a group of 6 voxels (long arrow), or a strip of 7 voxels (short arrow) in the interpolated map. Two apparently separated activated areas in the original map are connected in the interpolated map (ellipse). Two 6-voxel activated regions with similar shape in the interpolated map are represented as a single-voxel (long arrow) activation or a 3-voxel activation (medium arrow) in the original map.



(a) (b)

Figure 7: The activation maps at the frontal eye fields (FEF) in 6 subjects. In (a), the original maps are shown at the top, the interpolated maps (k-space) are shown at the middle, and the interpolated maps (magnitude) are shown at the bottom. Two apparently connected activated voxels in the original map are separated in the interpolated map (large blue circle). A single-voxel activation in the original maps can be a single-voxel activation (small blue circle) or a group of 6 activated voxels (small green circle) in subjects 1 and 2 in the interpolated map. In subject 3, the activated voxels form an enclosed loop in the original map but not in the interpolated map (large green circle). Similar results were also seen in the interpolated maps obtained with magnitude VSI. (b) shows the interpolated map in a slice from subject 6, at which the region of FEF is indicated by a green box.

the intensity of the neighboring voxels to generate the point data at locations not originally sampled. As such, the intensities of these spatially interpolated voxels are different than the true intensities if they were actually acquired. It is noteworthy that k-space VSI does not introduce additional blurring, because it does not increase the physical voxel dimension, although the displayed interpolated images appear much less pixelated. In comparison, other interpolation methods usually introduce additional spatial blurring.

Complex k-space VSI offers several advantages compared to the truncated sinc interpolation in the image space. First, the sinc interpolation applied to the magnitude image is only an approximation of the complex k-space VSI, due to the lack of phase information presented in the image data. Secondly, the kernel size of the sinc interpolation is usually severely truncated in order to perform the interpolation with reasonable computation time. Because the equivalent kernel size of VSI is equal to the matrix size of the image, both

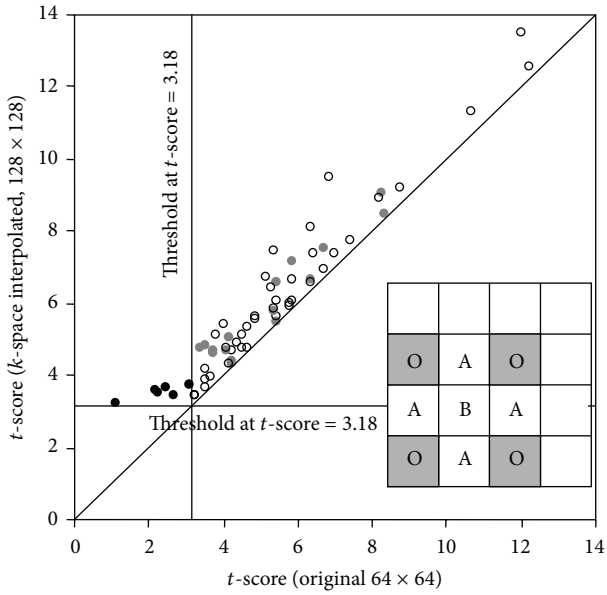


Figure 8: The t-scores are evaluated in the original and k-space interpolated activation maps in the FEF region in six subjects. The inset drawing shows the spatial relationship of the A- and B-type voxels in the interpolated image to the location of the original voxels, labeled as Os. Each circle in this plot shows an interpolated voxel that has a higher t-score (y-axis) than the highest t-score of its nearest original neighbors (x-axis), either an A-type (open circle) or B-type (gray solid circle) voxel. In the 6 interpolated maps, the t-scores in 62 of a total 304 voxels are higher than the highest t-scores in the nearest neighboring original voxels, with an average of 13.3% increase of the t-score. In each of the 6 interpolated voxels shown as dark solid circles, the highest t-score of its nearest original neighbors is below the threshold of 3.18 in the original maps. These 6 interpolated voxels, indicated by the blue arrows in Figure 7(a), have t-scores above the threshold and are located at the edge of the multivoxel activated regions. The diagonal line represents the situation when the t-scores in the original maps are identical to that in the interpolated maps.

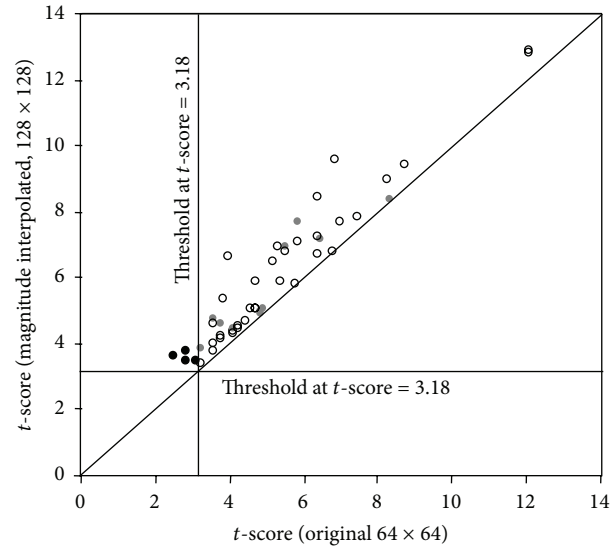


Figure 9: The t-scores are evaluated in the original and magnitude interpolated activation maps in the FEF region in six subjects. Each circle in this plot shows an interpolated voxel that has a higher t-score (y-axis) than the highest t-score of its nearest original neighbors (x-axis), either an A-type (open circle) or B-type (gray solid circle) voxel. In the 6 interpolated maps, the t-scores in 49 of a total 311 voxels are higher than the highest t-scores in the nearest neighboring original voxels, with an average of 14.4% increase of the t-score. In each of the 4 interpolated voxels shown as dark solid circles, the highest t-score of its nearest original neighbors is below the threshold of 3.18 in the original maps.

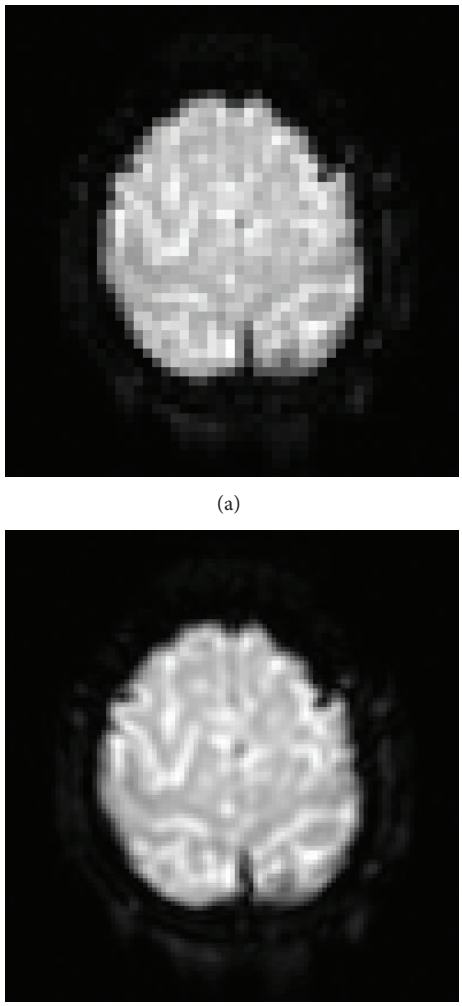
complex k-space VSI and magnitude VSI are superior to severely truncated sinc interpolation.

4.2. PVE and Its Reduction in BOLD fMRI. VSI provides a unique alternative for the detection of highly localized activation with a weak BOLD response. A voxel-size BOLD activation usually has a weak BOLD signal. For example, a previous study of the point spread function of the BOLD activation in human V1 at 4 Tesla demonstrated that both the spatial extent and the amplitude of BOLD signal decrease when the stimuli with a constant amplitude decrease in size [18]. PVE can severely hamper the detection of a voxel-size BOLD activation, due to the small size and low BOLD signal change of the activation. Figure 4 shows that PVE can thus severely reduce the spatial accuracy in depicting the activated regions in the original map, especially at the boundaries of activated regions or when the size of the activated region is in the order of voxel size. Although PVE can be reduced by increasing the spatial resolution, the reduction of SNR in high-resolution fMRI can outweigh the benefits of reduced PVE or even become prohibitive in detecting the weak BOLD signal response in some studies.

The shape of the activated region is better depicted and the apparent location of activated region, which can be defined by the center of mass, can be more precisely determined with VSI. The increased precision in locating the activated regions can potentially improve the fMRI mapping of visual cortices [19], tonotopy [20–22], and somatotopy [23]. VSI may be particularly useful in presurgical mapping studies, where the precision of anatomical location is of paramount importance.

In the simulation, the spatial dependency of the shape and maximum t-score on the relative location of an activation site on the grid of voxels in the low resolution images (i.e., 64 × 64) clearly demonstrate the effect of PVE on the ability to detect BOLD responses in fMRI. In comparing the spatially dependent variation of the maximum t-score for each of 24 activated regions in simulation, only the second condition of PVE discussed in the introduction (i.e., the peak of the BOLD activation is mismatched from the location of any voxel) was involved. The results of this simulation show that VSI is very effective in reducing PVE arising from this condition, especially when the activation is weak and highly localized.

The increased t-score in the A- and B-type voxels in the human fMRI experiment is the result of reducing PVE through better matching the voxel sensitivity function with the spatial profile of the BOLD signal. In six A-type voxels shown as dark solid circles in Figure 8, the t-scores of their



(a)

(b)

Figure 10: A comparison of the original (a) and VSI interpolated (b) echo-planar images shows that VSI improves delineation between the gray and white matter. In order to show these two images in the same size, each voxel in the original image was duplicated by a factor of 4 to form a 128 × 128 image in (a). The edge of the brain also appears much smoother in the interpolated image than in the original image.

nearest original neighbors are below the threshold. These A-type voxels, as indicated by the blue arrows in Figure 7, demonstrates the reduction of PVE arising from the third condition (i.e., the voxels at the boundary of activated regions are partially occupied by the activation) and improvement in delineating the boundary of activated regions.

In the magnitude interpolated maps, the A- and B-type voxels consist of 15.8% of the total activated voxels, comparable to 20.4% in *k*-space interpolated maps. Furthermore, 76% of these A- and B-type voxels are at the same location as the A- or B-type voxels in the *k*-space interpolated maps. These comparisons suggest that the magnitude interpolation is a good approximation of the *k*-space interpolation and thus puts forward this VSI method as a viable postprocessing option for conventional fMRI image data.

4.3. Other Considerations. VSI also improves the delineation of GM and white matter (WM) in the functional images, as shown in Figure 10. The thickness of GM in human brains is on the order of 2 mm, smaller than the typical voxel size used in fMRI. The shape of the GM regions is less continuous in the original EPI data due to PVE or the lack of sampling density in the image space. After VSI, the GM regions in the EPI data became clearer and more continuous. This improved delineation between GM and WM can lead to (1) more accurate coregistration between the EPI volumes at different time points during motion correction, (2) more accurate coregistration between the EPI data and the high-resolution anatomical data, and (3) improved warping of the EPI data onto standard brain template, such as the Montreal Neurological Institute (MNI) template.

Further increasing the matrix size with an interpolation factor of four (e.g., from 64 × 64 to 256 × 256) or higher may slightly reduce PVE and improve the detection of fMRI signal. However, the additional improvement is expected to be much less substantial than the improvement obtained with an interpolation factor of two (e.g., from 64 × 64 to 128 × 128) [13]. The penalty in computation time and the increased size of a dataset at a higher interpolation factors can easily outweigh the benefit of a more reduced PVE.

A direct comparison between the VSI activation maps and the ones obtained from high-resolution acquisition would be helpful for the validation of the proposed technique. Since high-resolution fMRI scans were not performed in the current study, such a comparison should be considered in future studies.

5. Conclusions

VSI improves the spatial accuracy of activation in fMRI through reducing PVE. Additionally, VSI can improve the sensitivity in detecting BOLD signal, especially when the size of activated regions is on the order of voxel size. VSI achieves these advantages of higher resolution fMRI without penalties in SNR, temporal resolution, and image artifacts.

Conflict of Interests

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

Acknowledgments

This work was partially supported by the National Key Basic Research Program of China (2013CB329501), the National Natural Science Foundation of China (81371518, 81261120411), and the Ministry of Science and Technology (2011BAI12B01). The Brain and Behavior Research Foundation, and the Blowitz-Ridgeway Foundation.

References

[1] S. Ogawa, T.-M. Lee, A. S. Nayak, and P. Glynn, "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain

at high magnetic fields," *Magnetic Resonance in Medicine*, vol. 14, no. 1, pp. 68–78, 1990.

[2] P. A. Bandettini, E. C. Wong, R. S. Hinks, R. S. Tikofsky, and J. S. Hyde, "Time course EPI of human brain function during task activation," *Magnetic Resonance in Medicine*, vol. 25, no. 2, pp. 390–397, 1992.

[3] R. S. Menon, "Postacquisition suppression of large-vessel BOLD signals in high-resolution fMRI," *Magnetic Resonance in Medicine*, vol. 47, no. 1, pp. 1–9, 2002.

[4] E. Yacoub, T. Q. Duong, P.-F. van de Moortele et al., "Spin-echo fMRI in humans using high spatial resolutions and high magnetic fields," *Magnetic Resonance in Medicine*, vol. 49, no. 4, pp. 655–664, 2003.

[5] J. S. Hyde, B. B. Biswal, and A. Jesmanowicz, "High-resolution fMRI using multislice partial k-space GR-EPI with cubic voxels," *Magnetic Resonance in Medicine*, vol. 46, no. 1, pp. 114–125, 2001.

[6] F. Zhao, P. Wang, and S.-G. Kim, "Cortical depth-dependent gradient-echo and spin-echo BOLD fMRI at 9.4T," *Magnetic Resonance in Medicine*, vol. 51, no. 3, pp. 518–524, 2004.

[7] R. S. Menon and B. G. Goodyear, "Submillimeter functional localization in human striate cortex using BOLD contrast at 4 Tesla: implications for the vascular point-spread function," *Magnetic Resonance in Medicine*, vol. 41, no. 2, pp. 230–235, 1999.

[8] J. P. Szaflarski, S. K. Holland, V. J. Schmithorst, R. S. Dunn, and M. D. Privitera, "High-resolution functional MRI at 3 T in healthy and epilepsy subjects: Hippocampal activation with picture encoding task," *Epilepsy and Behavior*, vol. 5, no. 2, pp. 244–252, 2004.

[9] P. F. Rodriguez, A. R. Aron, and R. A. Poldrack, "Ventral-striatal/nucleus-accumbens sensitivity to prediction errors during classification learning," *Human Brain Mapping*, vol. 27, no. 4, pp. 306–313, 2006.

[10] D. L. Parker, Y. P. Du, and W. L. Davis, "The voxel sensitivity function in Fourier transform imaging: applications to magnetic resonance angiography," *Magnetic Resonance in Medicine*, vol. 33, no. 2, pp. 156–162, 1995.

[11] A. R. Preston, M. E. Thomason, K. N. Ochsner, J. C. Cooper, and G. H. Glover, "Comparison of spiral-in/out and spiral-out BOLD fMRI at 1.5 and 3 T," *NeuroImage*, vol. 21, no. 1, pp. 291–301, 2004.

[12] P. Jezzard and R. S. Balaban, "Correction for geometric distortion in echo planar images from B(o) field variations," *Magnetic Resonance in Medicine*, vol. 34, no. 1, pp. 65–73, 1995.

[13] Y. P. Du, D. L. Parker, W. L. Davis, and G. Cao, "Reduction of partial-volume artifacts with zero-filled interpolation in three-dimensional MR angiography," *Journal of Magnetic Resonance Imaging*, vol. 4, no. 5, pp. 733–741, 1994.

[14] A. D. Radant and D. W. Hommer, "A quantitative analysis of saccades and smooth pursuit during visual pursuit tracking. A comparison of schizophrenics with normals and substance abusing controls," *Schizophrenia Research*, vol. 6, no. 3, pp. 225–235, 1992.

[15] B. Luna and J. A. Sweeney, "Cognitive functional magnetic resonance imaging at very-high-field: eye movement control," *Topics in Magnetic Resonance Imaging*, vol. 10, no. 1, pp. 3–15, 1999.

[16] J. Tanabe, J. Tregellas, D. Miller, R. G. Ross, and R. Freedman, "Brain activation during smooth-pursuit eye movements," *NeuroImage*, vol. 17, no. 3, pp. 1315–1324, 2002.

[17] C. Rosano, C. M. Krisky, J. S. Welling et al., "Pursuit and saccadic eye movement subregions in human frontal eye field: a high-resolution fMRI investigation," *Cerebral Cortex*, vol. 12, no. 2, pp. 107–115, 2002.

[18] K. Ueno, R. A. Waggoner, K. Tanaka, and K. Cheng, "Spatial precision of BOLD-fMRI in human V1: point spread function measured at 4T with spatially localized and size-varied stimuli," in *Proceedings of 11th Annual Meeting of the Organization for Human Brain Mapping (OHBM '05)*, p. 124, Toronto, Canada, 2005.

[19] M. I. Sereno, A. M. Dale, J. B. Reppas et al., "Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging," *Science*, vol. 268, no. 5212, pp. 889–893, 1995.

[20] M. Schönwiesner, D. Y. Von Cramon, and R. Rübsamen, "Is it tonotopy after all?" *NeuroImage*, vol. 17, no. 3, pp. 1144–1161, 2002.

[21] T. M. Talavage, P. J. Ledden, R. R. Benson, B. R. Rosen, and J. R. Melcher, "Frequency-dependent responses exhibited by multiple regions in human auditory cortex," *Hearing Research*, vol. 150, no. 1-2, pp. 225–244, 2000.

[22] D. Bilecen, K. Scheffler, N. Schmid, K. Tschopp, and J. Seelig, "Tonotopic organization of the human auditory cortex as detected by BOLD- FMRI," *Hearing Research*, vol. 126, no. 1-2, pp. 19–27, 1998.

[23] P. Hluštík, A. Solodkin, R. P. Gullapalli, D. C. Noll, and S. L. Small, "Somatotopy in human primary motor and somatosensory hand representations revisited," *Cerebral Cortex*, vol. 11, no. 4, pp. 312–321, 2001.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 747549, 13 pages
http://dx.doi.org/10.1155/2014/747549

Research Article

A New Multistage Medical Segmentation Method Based on Superpixel and Fuzzy Clustering

Shiyong Ji,[1] Benzheng Wei,[2] Zhen Yu,[1] Gongping Yang,[1] and Yilong Yin[1]

[1] *School of Computer Science and Technology, Shandong University, Jinan 250101, China*
[2] *College of Science and Technology, Shandong University of Traditional Chinese Medicine, Jinan 250355, China*

Correspondence should be addressed to Yilong Yin; ylyin@sdu.edu.cn

Received 22 November 2013; Accepted 9 January 2014; Published 9 March 2014

Academic Editor: Yuanjie Zheng

The medical image segmentation is the key approach of image processing for brain MRI images. However, due to the visual complex appearance of image structures and the imaging characteristic, it is still challenging to automatically segment brain MRI image. A new multi-stage segmentation method based on superpixel and fuzzy clustering (MSFCM) is proposed to achieve the good brain MRI segmentation results. The MSFCM utilizes the superpixels as the clustering objects instead of pixels, and it can increase the clustering granularity and overcome the influence of noise and bias effectively. In the first stage, the MRI image is parsed into several atomic areas, namely, superpixels, and a further parsing step is adopted for the areas with bigger gray variance over setting threshold. Subsequently, designed fuzzy clustering is carried out to the fuzzy membership of each superpixel, and an iterative broadcast method based on the Butterworth function is used to redefine their classifications. Finally, the segmented image is achieved by merging the superpixels which have the same classification label. The simulated brain database from BrainWeb site is used in the experiments, and the experimental results demonstrate that MSFCM method outperforms the traditional FCM algorithm in terms of segmentation accuracy and stability for MRI image.

1. Introduction

The medical image is human body image captured by medical imaging equipment, including Computed Tomography (CT), Magnetic Resonance imaging (MRI), and Ultrasonography (US.) Based on the computer graphics technology, the quality and displaying method of medical image have greatly improved. MRI technology has many advantages such as nonradioactive contamination, high resolution, without electricity radiation damage to the human body, so it is widely applied in clinical diagnosis and treatment now. MRI image processing promotes the development of medical research and has important applications value. According to the above process, the accurate medical image segmentation by computational techniques plays an essential role [1].

In MRI medical image segmentation, the image is parsed into a number of meaningful regions based on the consistency principle. These regions usually do not cross each other and satisfy the consistency principle. If merging any two adjacent regions, it will break this principle. Hence, the medical image segmentation can be seen as the classification of image pixels in this viewpoint.

As is shown in Figure 1, the region marked by blue circle represents gray matter and region marked by red circle represents white matter. Generally, white matter has a larger gray value than gray matter. Nevertheless, with the influence of intensity inhomogeneity, gray matter in blue circle has a larger gray value than white matter in red circle, which makes an overlap in the image.

The influence is that a slowly varying shading artifact over the MRI image can produce errors with conventional intensity-based classification. Consequently, several methods for MRI intensity inhomogeneity correction are applied before the image segmentation. Series methods on intensity inhomogeneity correction/removal have been proposed in the last two decades [2–5]. Nevertheless, intensity inhomogeneity correction is still incompletely solved problem. Because of this and the evolving MRI technology and associated applications, the problem of intensity inhomogeneity

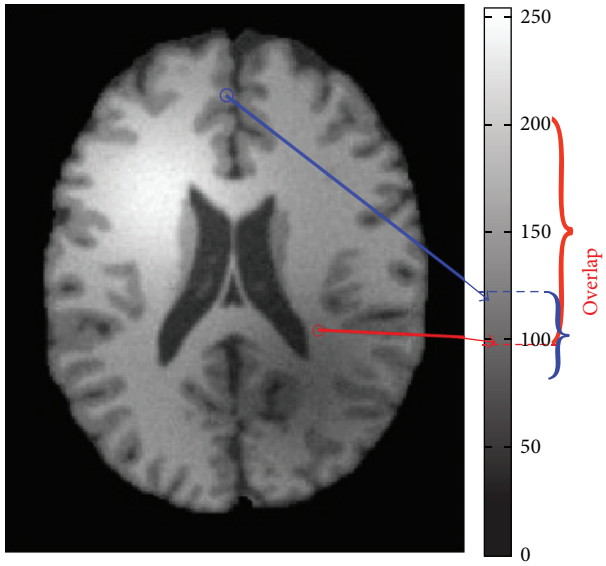


Figure 1: The illustration of bias field in MRI image.

correction will certainly continue to be paid more research attention in the future [6].

In previous works of MRI image segmentation, as the thorough application of statistical theory, fuzzy set theory and machine learning theory deserve paying much more attention [3, 7–11]. In particular, for the fuzziness of the medical image, the fuzzy theory is introduced into the medical image processing, which generates lots of new segmentation methods and achieves good segmentation results. The most representative method is the fuzzy C-means clustering algorithm proposed by Bezdek et al. in [12]. Pham and Prince proposed the adaptive fuzzy segmentation method to segment the MRI images [11]. In order to solve the bias field estimation and segmentation problem, Ahmed et al. in [5] proposed the modified fuzzy C-means algorithm for the MRI image segmentation. Amini et al. in [13] made use of the FCM (fuzzy C-means) to solve the segmentation problem of the thalamus in the brain MRI image. Shen et al. in [14] modified the fuzzy C-means clustering method to segment the brain MRI image. Awate et al. in [15] proposed a segmentation framework to deal with the DT and MRI image based on the fuzzy and the nonparametric estimation. Halt et al. in [16] proposed the Bayesian segmentation method based on the local adaptive fuzzy, which was used to solve the volume measurement problem in the PET image. The improved FCM method based on the histogram was proposed in [17] by Zhang et al. for the medical image segmentation.

The FCM algorithm has many advantages such as being without supervision, simple realization, and fast processing speed, which can carry out the accurate segmentation for the image with high contrast and signal-to-noise ratio. But there are also lots of obvious disadvantages. In the process of fuzzy clustering, the gray value distance between single pixel and cluster center can be considered only, while the influence of the adjacent pixels is neglected. That is to say, the spatial information cannot be well used in image segmentation. So the large deviation will be produced when the FCM algorithm is used to *segment* the brain MRI image with noises and the low signal-to-noise ratio [18]. Moreover, in order to segment the image with intensity inhomogeneity, the result by the FCM algorithm will be unsatisfactory because of the low contrast in the whole image.

Superpixels, also known as regions in an over segmentation of the image, would be more natural and presumably lead to more efficient processing [19]. The superpixel method had been increasingly used in image processing field, which can group pixels using the degree of feature similarity between pixels and acquire the redundant information of the image. Hence, it can greatly reduce the complexity of image post-processing tasks. The method that combines the superpixel and the clustering has successfully solved the arterial segmentation and tracking problem in the CT image [20]; Zhang and Ji in [21] proposed an image segmentation framework based on the superpixel and had achieved good result on natural images; Gan et al. applied the superpixel to the multiclass segmentation of the SAR image [22]; Zhou et al. in [23] proposed a superpixel driver method to track the target; Gong and Liu have employed the superpixel method to detect the rock in the image in [24]; the superpixel method has been used to perform segmentation of the background in the image by Jiang in [25]; Li et al. had proved that the particle aggregation information provided by the superpixel is useful for image segmentation in [26].

The superpixel method can make full use of spatial information, this feature makes the algorithm has well anti-noise performance. Besides, it can save the edge information of the original image during the process of enhancing the local consistency. The segmented atomic area has some image characteristics such as shape boundary contour information and area histogram and, in which characteristics are not affiliated with the single pixel. So the superpixel method can improve the segmentation accuracy and the processing time. In addition, the gray value of each pixel within the superpixels is very similar, and the phenomenon of intensity inhomogeneous will not exit in the superpixels.

Due to the fuzziness feature, it is hard to segment MRI images well. FCM algorithm may solve this problem to a certain extent, but it is sensitive to noise and bias field. To take advantages of superpixel method to reduce the effect of these problems, a new multistage segmentation method based on the superpixel and designed FCM is proposed, which does not only make use of the beneficial aspects of the fuzzy clustering algorithm in the medical image, but also use the superpixel method to enhance the space constraint and it effectively solved the inhomogeneity problem.

The structure of this work is as follows. In Section 2, the basics of superpixel method and FCM algorithm are presented. In Section 3, the proposed MCFCM method is introduced, along with detailed multistage segmentation processing. The experimental results are discussed in Section 4, and from the numerical analysis, conclusions are presented in Section 5.

2. Preliminaries

In this section, the fundamentals of the superpixel method and the FCM algorithm are explained in detail.

2.1. Superpixel Level Segmentation Method. The superpixel is to use some algorithms to aggregate some pixels together to form atomic regions that have a certain meaningful perception. Atomic regions are used to replace the regional grids that are segmented rigidly. The superpixel as the basic unit seems inefficient. What leads to this situation is that it needs to carry out a task completely unrelated to the final decision task to converge pixels into different groups. However, redundant information can be got partially from the data and the risk of merging unrelated pixels in the process of aggregating pixels into superpixels can be minimized, which can help to achieve the purpose of decision making. Meanwhile, superpixel helps to obtain some characteristics of the statistical information in a natural adaptive area rather than an artificial divided area. As boundary information is considered when the image is segmented into superpixels, more accurate segmentation results can be got by finding some superpixels belonging to the target.

There are many superpixel segmentation methods in recent years, such as turbo pixel [27, 28], normalized cuts [29], quick shift [30], and SLIC superpixel [19]. The feature of normalized cuts segmentation is that the number of superpixels can be controlled, the shape of superpixel is relatively compact, and the area of superpixel is broadly similar as well. But normalized cuts segmentation has a low running speed, especially for large pictures that need large amount of computation. SLIC is an efficient method that uses color similarity of pixels and spatial information of image to generate compact and uniform superpixels. Due to that superpixels achieved by normalized cuts and SLIC are always compact and with uniform shape, their semantics performance is poor. Superpixels with compact structure cannot cover a complete object, and uniform shapes lead to different semantic levels in segmenting target of different scales. Quick shift is a gradient based pattern search segmentation method. This method achieves image segmentation by promoting data points in feature space move along the Parzen density ascendant direction. Quick shift algorithm cannot limit the shape and size of superpixel, and the compactness of superpixel is also poor. Turbo pixel algorithm can control the number of superpixels and has a high processing speed. What is more, superpixels generated by turbo pixel have approximate sizes, and the boundaries are more close to the real image. The basic idea of turbo pixel is to select a certain number of seed points on the image and devise a flow by which curves evolve to obtain superpixel boundaries.

In this paper, the turbo pixel method is utilized, and the details of this method are presented as follows.

Let C be a vector of curve coordinates parameterized by t, a parameter to denote evolution in time. Let N represent its outward normal and let each point move with speed S. Then let level set curve evolution equation be

$$\frac{\partial C}{\partial t} = SN. \tag{1}$$

This curve evolution equation is implemented by first embedding C as a level set of a smooth and continuous function $\Psi : R^2 \times [0, \tau) \rightarrow R^2$ and then evolving this embedding function according to

$$\psi_t = -S \left\| \nabla \psi \right\|. \tag{2}$$

And the first-order discretization formula of (2) is

$$\psi^{n+1} = \psi^n - S_I S_B \left\| \nabla \psi^n \right\| \Delta t. \tag{3}$$

Each application of formula (2) corresponds to one "time step" Δt in the evolution of the boundary. The key term controlling the evolution is the product of two speeds $S_I S_B$. S_I depends on local image structure and superpixel geometry at each boundary point and S_B depends on the boundary point's proximity to other superpixels. The S_I of formula (3) consists of reaction-diffusion-based shape segmentation model and the geodesic active contour model as

$$\begin{aligned} S_I(x, y) = {} & \left[1 - \alpha\kappa(x, y)\right] \varphi(x, y) \\ & - \beta \left[N(x, y) \cdot \nabla \varphi(x, y)\right]. \end{aligned} \tag{4}$$

The first half of the formula (4) that named reaction-diffusion term ensures that the boundary's evolution slows down when it gets close to a high gradient region in the image. $\varphi(x, y)$ is local affinity function: $\varphi(x, y) = e^{-E(x,y)/\nu}$, $E(x, y) = \|\nabla_I\| / (G_\sigma \cdot \|\nabla_I\| + \gamma)$, computed for every pixel on the image plane, and the $\varphi(x, y)$ has a low value near the edge and has a high value elsewhere. $\kappa = (\Psi_{xx}\Psi_y^2 - 2\Psi_x\Psi_y\Psi_{xy} + \Psi_{yy}\Psi_x^2)/(\Psi_x^2 + \Psi_y^2)^{3/2}$ expresses the curvature of the boundary at point (x, y) and smoothes the evolving boundaryand α is balancing parameter that weighs the contribution of the curvature term. Intuitively, the latter part of the formula (4) that named doublet term ensures that the boundary is attracted to image edges $N = \nabla\Psi / \|\nabla\Psi\|$.

The entire algorithm of turbo pixel is summarized as follows.

Step 1. Initialize seeds, and perturb the seed positions away from high gradient regions.

Step 2. Set all seed pixels to "assigned."

Step 3. Set Ψ^0 to be the signed Euclidean distance from the "assigned" regions,
$\sum_{x,y}[\Psi^0(x, y) \geq 0] \rightarrow$ assigned pixels.

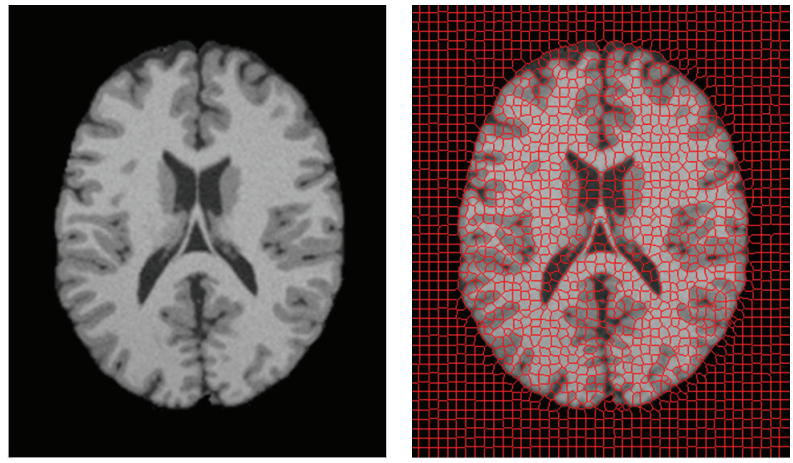
Step 4. Compute $\varphi(x, y)$, $n \rightarrow 0$.

Step 5. While change in assigned pixels is large, do

(i) compute $S_I S_B$

(ii) $S_I S_B \rightarrow S$, extend the speed S in a narrow band near the zero level-set of Ψ^n;

(iii) compute Ψ^{n+1} by evolving Ψ^n within the narrow band, $n = n + 1$;



(a) The original MRI image

(b) The segmentation result by turbo pixel method

Figure 2: An example of the superpixel level segmentation results.

(iv) $\sum_{x,y}[\Psi^n(x,y) \geq 0] \rightarrow$ assigned pixels;

(v) Homotopic skeleton of $\Psi^n \rightarrow B$.

Step 6. Return superpixel boundary B.

The following Figure 2 shows the superpixel segmentation results of an MRI image by turbo pixel method.

2.2. FCM Algorithm. FCM algorithm was proposed by Dunn and later on modified by Bezdek [12]. The basic principle of FCM is the iterative minimization of the following objection function:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{p}\left\|y_k - v_i\right\|^2. \tag{5}$$

Let $\{y_k, k = 1,2,\ldots,N\}$ denote an image with N pixels to be categorized into c clusters, $\{v_i, i = 1,2,\ldots,c\}$ denote every cluster centers, and $C = (c_1, c_2, \ldots, c_c)$ is the cluster center matrix. The vector $U_k = (u_{1k}, u_{2k}, \ldots, u_{ik})^T$ denotes the membership of the kth pixel in i clusters, u_{ik} $(u_{ik} \in [0,1])$ is the membership of the kth pixel in the ith cluster, the $U = (U_1, U_2, \ldots, U_k)$ is the membership matrix, p is the membership function index that controls the fuzziness of resulting partitions, and $\|y_k - v_i\|^2$ is a norm metric which usually uses Euclidean distance.

The algorithm steps are as follows.

Step 1. Set iteration stop threshold ε, initialize the membership matrix U and cluster center matrix C, and let iteration counter q be equal to 0.

Step 2. The membership function is updated as

$$u_{ik} = \frac{\left(1/\left\|y_k - v_i\right\|^2\right)^{1/(p-1)}}{\sum_{j=1}^{c}\left(1/\left\|y_k - v_j\right\|^2\right)^{1/(p-1)}}. \tag{6}$$

Step 3. The cluster centers are updated as

$$v_i = \frac{\sum_{k=1}^{N} u_{ik}^{p} y_k}{\sum_{k=1}^{N} u_{ik}^{p}}. \tag{7}$$

Step 4. When the objective function value changes less than the setting threshold, then stop the algorithm.

3. The Proposed MSFCM Method

The Multistage medical image segmentation method Based on superpixel and Fuzzy clustering (MSFCM) is proposed taken into account the advantages of fuzzy clustering in medical image processing and superpixel's advantages in strengthening space information and effectively processing in intensity inhomogeneity problem. The MSFCM is based on superpixel, which can make up the insufficiencies in noise and bias field processing aspect by only using original FCM. And this method can be divided into three stages.

(i) *Rough Segmentation*. Partition the image into superpixels.

(ii) *Deep Segmentation*. Parsing superpixels which have large variance into smaller atomic regions.

(iii) *Cluster and Label Superpixels*. Cluster superpixels with FCM method and label superpixel to the appropriate class using spatial and gray information and finally obtain the segmentation result by merging superpixels belonging to the same class.

The flowchart is shown in Figure 3 and the detailed process of MSFCM is shown as follows.

3.1. Rough Segmentation. Superpixel segmentation is equivalent to an image over segmentation, and its essence is also described in the form of image segmentation. So it is suitable to preprocess the image with superpixel method to

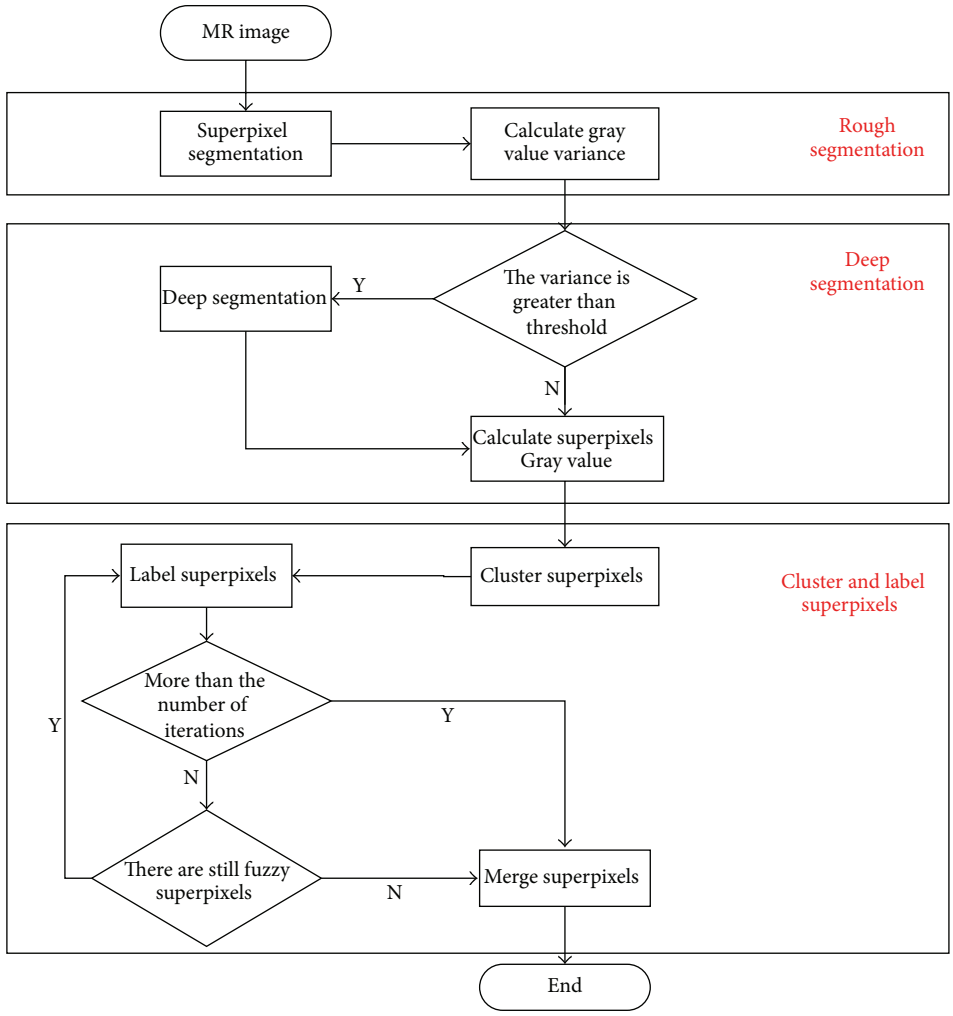
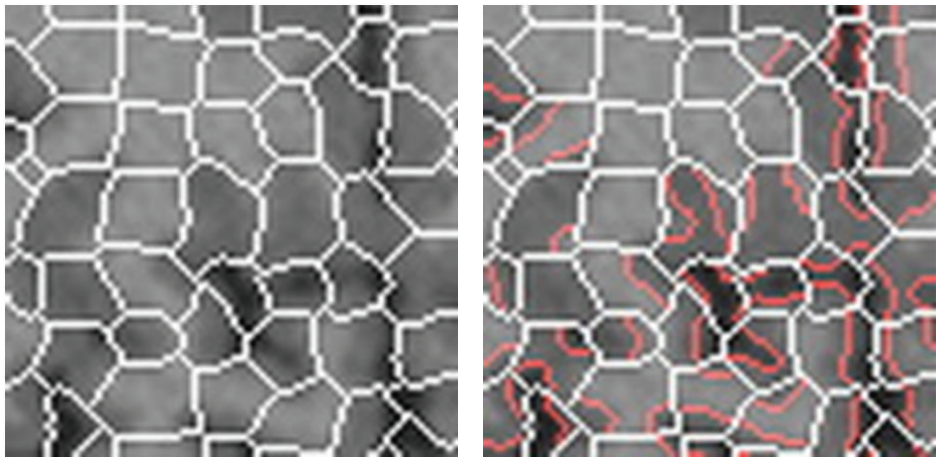


Figure 3: The flowchart of MSFCM.



(a) The result of superpixel level segmentation (b) The result of deep segmentation

Figure 4: An example of the typical deep segmentation result.

get the rough segmentation of the MRI image. For an MRI image which has a size of M^*N $(0 \le x \le M, 0 \le y \le N)$, let $\Lambda(x, y)$ represent the entire image grid. The segmentation for Λ can be considered as dividing it into n nonempty region $(R_1, R_2, \ldots, R_n)$ which must satisfy the following five conditions.

(i) $\cup_{i=1}^{N} R_i = \Lambda$.

(ii) For all i and j, when $i \ne j$, $R_i \cap R_j = \phi$.

(iii) For all R_i, $i = 1, 2, \ldots, N$, $P(R_i) = \text{true}$.

(iv) For all i and j, when $i \ne j$, $P(R_i \cap R_j) = \text{false}$.

(v) For all R_i, $i = 1, 2, \ldots, N$, R_i is a connected region.

$P(R_i)$ is the logical predicate of the elements for every R_i $(i = 1, 2, \ldots, N)$ and ϕ represent the empty set.

MSFCM used turbo pixel method to segment the image. And L superpixels R_i $\{i = 1, 2, \ldots, L\}$ can be got as the rough segmentation result.

3.2. Deep Segmentation. As part of the boundary of some regions in the MRI image is fuzzy, a problem can occur after superpixel level segmentation, which is that the different tissues are wrongly divided into the same superpixel. In order to reduce such errors, it is necessary to deeply segment the image on some specific superpixels.

Because such superpixel's variance is bigger than other superpixels, it is feasible to take the sequence of the front of a certain percentage of superpixels to do the further segmentation. Automatic threshold segmentation method can solve this problem well. Furthermore, a scale parameter t (the value of t can be set according to specific situation) is introduced to eliminate the impact of noise points and those regions in which proportion in original superpixel is greater than t after threshold segmentation is saved to the deep segmentation process.

After this procession, K $(K \ge L)$ superpixels R_i $\{i = 1, 2, \ldots, k\}$ can be got, and these regions are the objects for FCM clustering.

An example of the typical deep segmentation result is shown in Figure 4, the white boundary is generated by superpixel method, and the red boundary is generated by deep segmentation process. It is easy to see that different tissues can be separated clearly after deep segmentation.

3.3. Cluster and Label Superpixels. To obtain the final result, it is essential to cluster and label the superpixels to the right classifications. And this process can be divided into the following three parts.

3.3.1. Part 1: FCM Clustering. As far as MRI image is concerned, it can be divided into three tissues: the gray matter, white matter, and cerebrospinal fluid. So the classification parameter of FCM could be set to three. In this paper, the mean of every superpixel's gray value μ is used as clustering parameter. As the FCM clusters MRI image to K superpixels generated in Section 3.2, the clustering center matrix $C(c_1, c_2, c_3)$ and membership matrix U can be obtained. And each superpixel's classification label is determined by U.

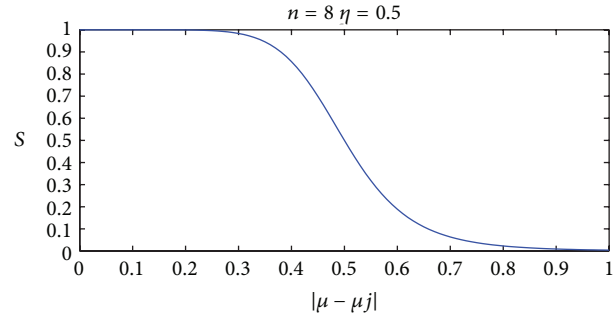


Figure 5: An example of the typical Butterworth function curve.

3.3.2. Part 2: Label the Superpixels. In view of the fuzzy and inhomogeneity property of brain MRI medical image, it is not feasible to label the superpixels to the right classification with clustering results directly. So it is necessary to introduce other information to help label superpixels.

As the organization of brain MRI image has the characteristics of continuity, spatial adjacent information of superpixel is presented to determine which class the superpixel belongs to. Let $S(s_1, s_2, \ldots, s_n)$ indicate the similarity between adjacent superpixels. s_i is the similarity value between current superpixel and its ith adjacent superpixels.

To measure the similarity between superpixels, it is necessary to employ a function which has characteristic as follows.

(i) For superpixels that have small gray-scale difference, they should return a large value in similarity.

(ii) For superpixels that have large gray-scale difference, they should return a small value in similarity.

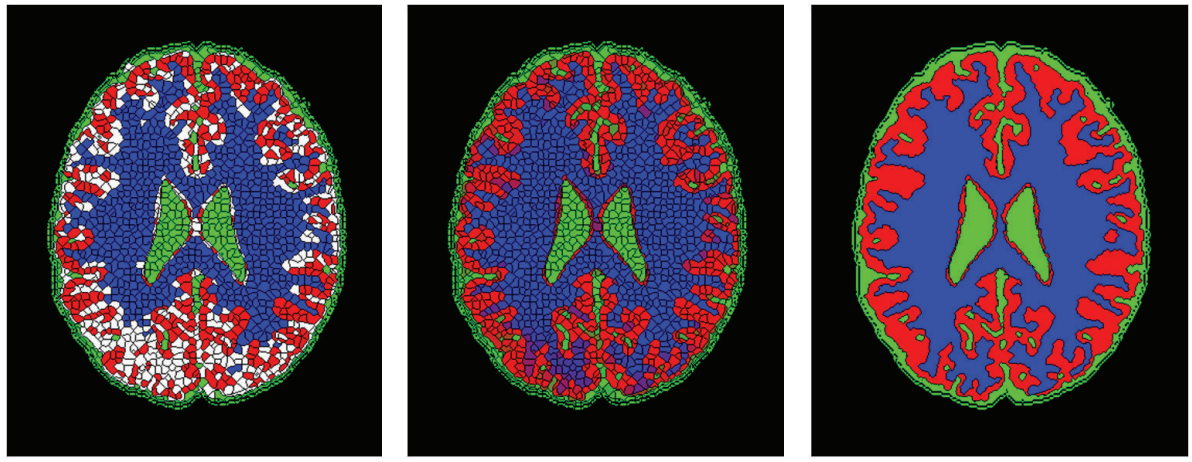
(iii) When the gray-scale difference between two superpixels exceeds a certain threshold, the similarity should decrease rapidly.

Based on the above requirements, Butterworth function is employed in this paper. The Butterworth function form is as follows:

$$S_i = \frac{1}{1 + \left((\mu - \mu_i)/\eta\right)^n}. \tag{8}$$

In (8), η is the tolerance value. With the increase of η, gray value difference would be allowed more larger as judging the similar superpixels; μ is the mean gray value of the superpixel to be judged, and μ_i is the mean gray value of this superpixel's adjacent superpixel I; n is function series; the greater of n value, the faster function declines. The Butterworth function curve is shown in Figure 5.

In order to label the superpixels to the right classification, a broadcast method based on spatial adjacent information is introduced. The detailed steps are presented as follows



(a) The labeled image after Part 1 (b) The labeled image after Part 2 (c) The final result after Part 3

Figure 6: An example of the labeling classification for superpixels.

Step 1. For superpixel R_i, define the membership vector $U_i(u_1, u_2, u_3)$, if there exist $u_i = \max\{u_1, u_2, u_3\} > T_c$ (T_c is confidence threshold), then this superpixel is marked to ith classification, else the superpixel should be marked to fuzzy block which is denoted by F.

Step 2. For $R_j \in F$, assuming that its adjacent superpixels set is $\Omega = \{R_{j1}, R_{j2}, \ldots, R_{jk}\}$, compute $S_j(s_{j1}, s_{j2}, \ldots, s_{jk})$ of R_j with each element of Ω respectively.

Step 3. If there exists a $s_{ij} = \max S_j > T_s$, then the R_j was marked to the same classification with R_{ji}. T_s is the confidence threshold.

Step 4. If the number of iterations is not more than a limited number and there are still fuzzy blocks, then go to Step 2.

Step 5. If the number of iterations is more than a limited number and there are still fuzzy blocks, then for fuzzy block R_i which has a membership vector $U_i = (u_1, u_2, u_3)$, if u_j has the maximum value in U_i, R_i should be marked to jth classification.

After all above steps, each superpixel has a clear classification.

3.3.3. Part 3: Merge Superpixels. Superpixels are going to be merged after processing Part 2. Superpixels which belong to the same classification and adjacent to each other will be merged, and the final segmentation result will be achieved.

Figure 6 shows the change of the superpixels' classification in this stage. In the figure, blue region represents the white matter, red region represents the gray matter, green region represents cerebrospinal fluid, and white region represents the fuzzy block. As is shown in Figure 6(a), there are lots of fuzzy blocks with the influence of inhomogeneity property and noise, while after being adopted by the broadcast label method proposed in this paper, the fuzzy blocks can be labeled to the proper classification and the final result can be obtained by merging the superpixels with the same classification which is shown in Figure 6(c).

4. Experimental Results

To verify the effectiveness of the algorithm, the synthetic MRI images with ground truth from Brain Web [31] are used as experimental data. In the database, noise parameters are settled as 0%, 3%, 5%, 7%, and 9%, and bias field parameters are settled as 0%, 20%, and 40%, and then 30 images are selected for each image sequence. Therefore, a total of 15 experiments need to be done. In these experiments, Jaccard similarity (JS) is applied as the metric to quantitatively evaluate the segmentation accuracy. The JS is defined as

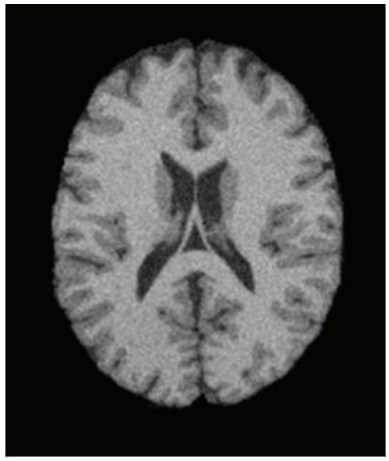
$$J(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}. \quad (9)$$

S_1, S_2 represent segmentation results of different algorithms and ground truth, respectively.

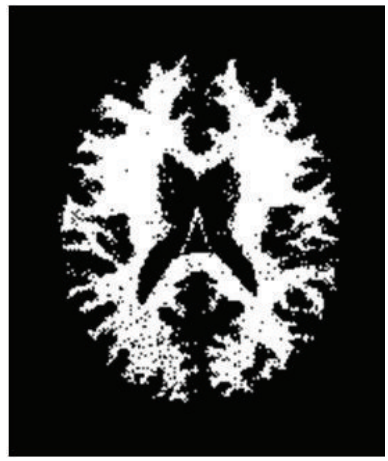
Under the setting 7% noise and 20% bias field, one of the segmentation results of MSFCM and FCM is shown in Figure 7. The MRI segmentation accuracy comparison of the proposed method MSFCM and the FCM is shown in Figures 8, 9, and 10 under the different setting noise and bias field.

Table 1 is the mean accuracy table of 30 images under the different setting noise parameter when bias field parameter is fixed. The mean accuracy table of 30 images under the different setting bias field parameter as the settled noise parameter is shown in Table 2.

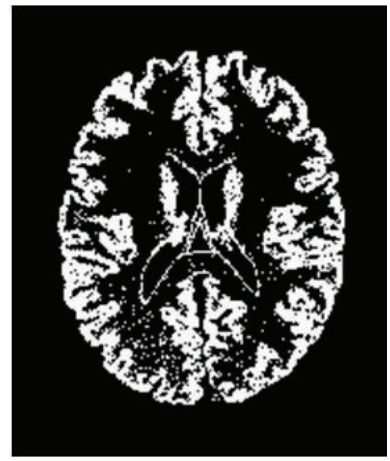
The 15 comparative experiments results in the various conditions show that the segmentation accuracy of MSFCM is much higher than FCM. In addition, in the case of gradually increase of noise and bias field parameter, the MSFCM's accuracy rate of decline is far less than FCM. As shown in Figure 11, with the increase of noise, the accuracy rate of MSFCM decrease more slowly than FCM in white matter and gray matter. The result shows that the proposed algorithm has advantages in accuracy and robustness compared with FCM.



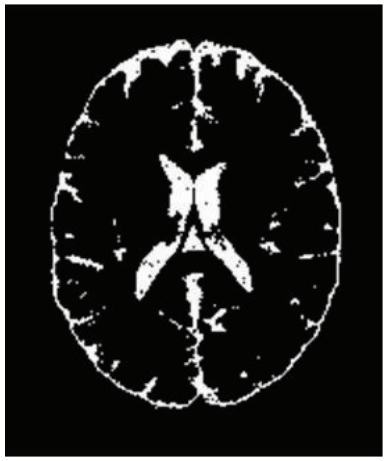
(a) The original image



(b) The white matter segmentation results of FCM

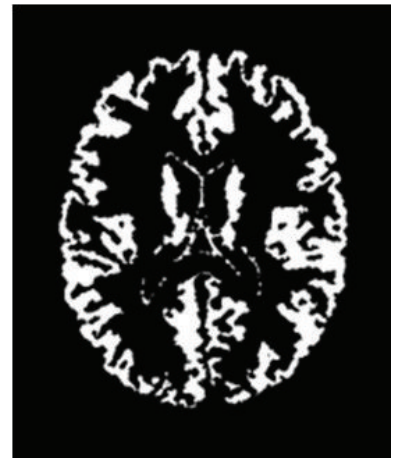


(c) The gray matter segmentation results of FCM

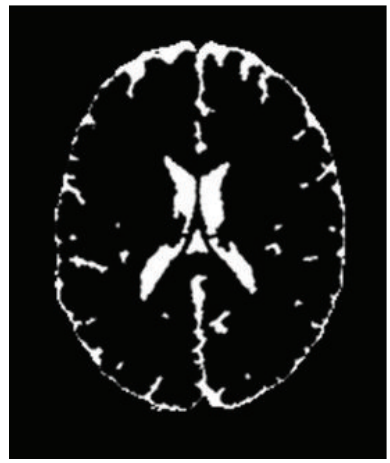


(d) The cerebrospinal fluid segmentation results of FCM

(e) The white matter segmentation results of MSFCM

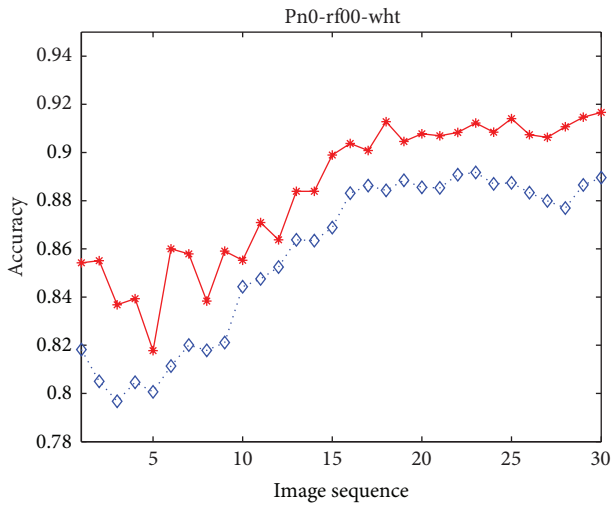


(f) The gray matter segmentation results of MSFCM

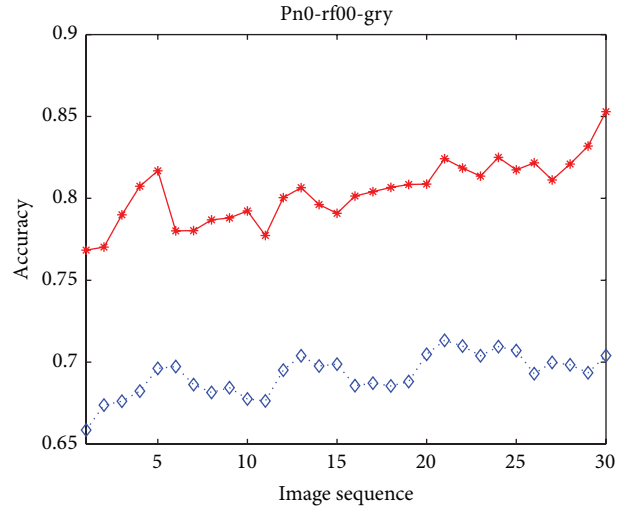


(g) The cerebrospinal fluid segmentation results of FCM

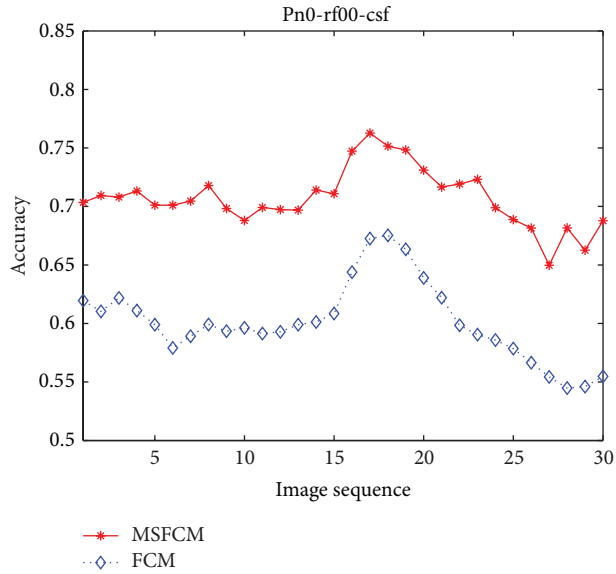
Figure 7: A typical comparison of FCM and MSFCM segmentation results.



(a) The accuracy comparison of FCM and MSFCM method on white matter segmentation



(b) The accuracy comparison of FCM and MSFCM method on gray matter segmentation

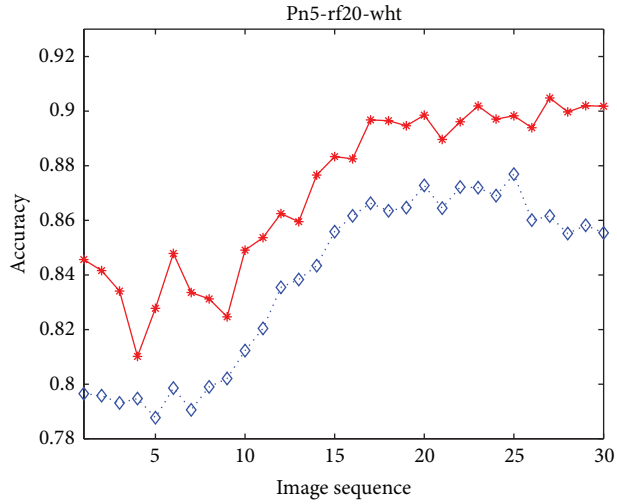


(c) The accuracy comparison of FCM and MSFCM method in cerebrospinal fluid segmentation

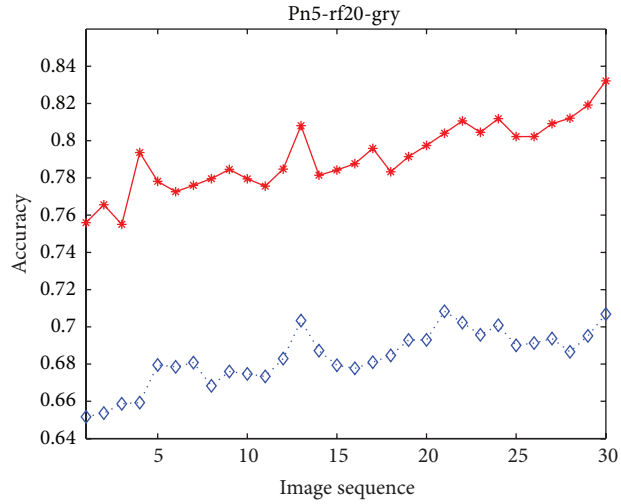
Figure 8: The comparison of segmentation accuracy as the 0% noise and 0% bias field.

Table 1: The mean accuracy of 30 images under different setting noise parameter as 0% bias field.

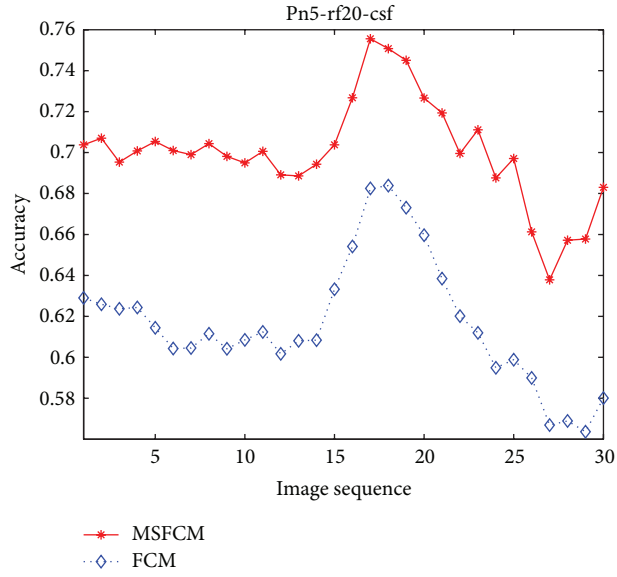
Method	Noise				
	0	3	5	7	9
MSFCM-CSF	0.7071	0.7058	0.6999	0.6918	0.6781
FCM-CSF	0.6016	0.6064	0.6110	0.6134	0.6067
MSFCM-GRY	0.8039	0.8003	0.7930	0.7825	0.7696
FCM-GRY	0.6923	0.6880	0.6789	0.6529	0.6018
MSFCM-WHT	0.8837	0.8793	0.8697	0.8566	0.8390
FCM-WHT	0.8574	0.8586	0.8473	0.8065	0.7404



(a) The accuracy comparison of FCM and MSFCM method on white matter segmentation



(b) The accuracy comparison of FCM and MSFCM method on gray matter segmentation



(c) The accuracy comparison of FCM and MSFCM method on cerebrospinal fluid segmentation

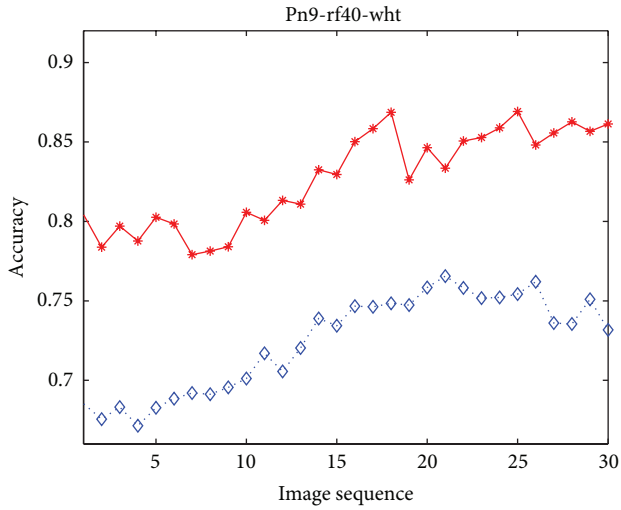
Figure 9: The comparison of segmentation accuracy as 5% noise and 20% bias field.

Due to that the clustering objects of MSFCM algorithm are the superpixels instead of pixels, this transformation increases the granularity of the clustering and is able to make full use of spatial constraint information, so MSFCM method has a good performance in terms of noise immunity. Furthermore, superpixels, which can also be considered as atomic regions, have some perceived significance: superpixel has lower difference on gray value in its internal space, and this feature can reduce the impact of inhomogeneity in the whole image. So in processing an image which has bias fields, the proposed method can effectively avoid the impact of this phenomenon on the segmentation.

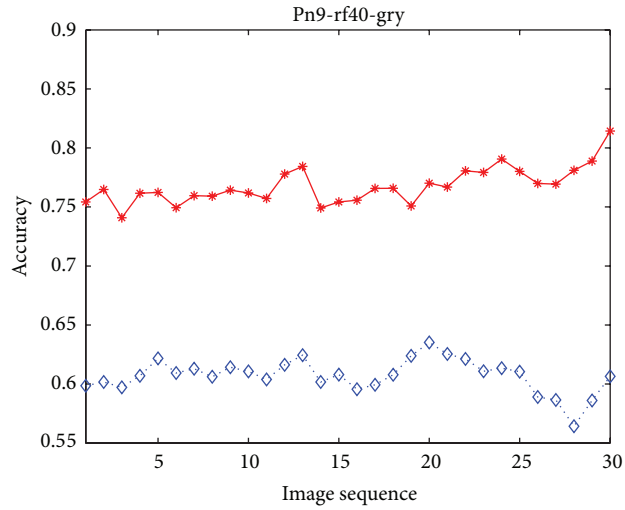
In superpixel clustering processing, MSFCM utilizes Butterworth function to process class discrimination issues for the fuzzy blocks. This way could take advantage of adjacent superpixels information and improved accuracy class determination compared with rigid partition.

Table 2: The mean accuracy of 30 images under different setting bias field parameter as 0% noise.

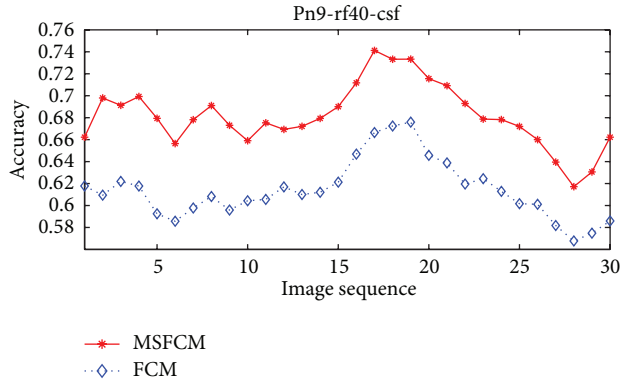
Method	Bias-filed		
	0	20	40
MSFCM-CSF	0.7071	0.7101	0.7073
FCM-CSF	0.6016	0.6075	0.6187
MSFCM-GRY	0.8039	0.8029	0.8016
FCM-GRY	0.6923	0.6963	0.7049
MSFCM-WHT	0.8837	0.8839	0.8787
FCM-WHT	0.8574	0.8517	0.8345



(a) The accuracy comparison of FCM and MSFCM method on white matter segmentation



(b) The accuracy comparison of FCM and MSFCM method on gray matter segmentation



(c) The accuracy comparison of FCM and MSFCM method on cerebrospinal fluid segmentation

FIGURE 10: The comparison of segmentation accuracy as 9% noise and 40% bias field.

MSFCM combined superpixel method and FCM method, effectively used both advantages in image processing, and targeted to overcome FCM's defects in noise and bias field aspect.

Therefore, MSFCM algorithm has higher accuracy and higher robustness in segmentation than FCM algorithm.

5. Conclusion and Discussion

In this paper, the MSFCM algorithm is presented to segment the brain MRI image, which consists of the superpixel method and the FCM algorithm. The image was firstly parsed into several superpixels, and then deep segmentation is to be done for the areas with bigger gray variance than setting threshold. And to get the fuzzy membership of each superpixel, the FCM algorithm is used to cluster the superpixels rather than pixels, and the membership is used to determine the classification for these superpixels. Finally, the segmented brain MRI image is achieved by merging the superpixels with the same classification.

The experiments reveal that the proposed method is more efficient and stable than FCM, and has achieved good results in segmenting MRI images with noise and intensity inhomogeneity. This advantage made it possible to obtain a high accuracy and effectiveness in the human brain MRI image segmentations compared to those outlined by the experts and by the FCM method according to the evidence of similarity metrics. Additionally, the experimental results have also shown that the local exploitation of broadcast method to properly label superpixels classifications and the Butterworth function to measure the similarity between superpixels are highly suitable for medical image applications, including segmenting datasets of sequential medical images within an appropriate computational time.

Highlight

(1) A broadcast method taking advantage of spatial adjacent information is proposed to label superpixels to the proper classification, which improves the accuracy compared with the inflexible label method.

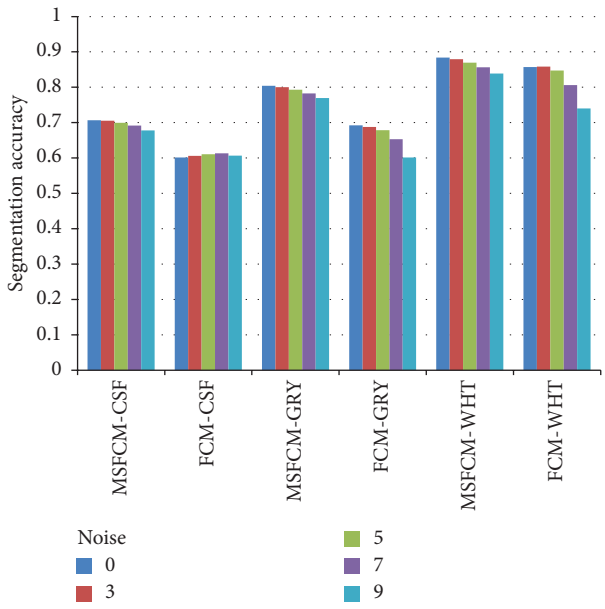


Figure 11: The robustness comparison of MSFCM and FCM.

(2) The Butterworth function is introduced and designed to measure the similarity between superpixels.

(3) The MSFCM is a new multistage segmentation method based on the superpixel method and the FCM algorithm, which combine the advantages of the two methods to solve the influence of noise and the bias filed in the brain MRI medical image segmentation effectively and robustly.

Conflict of Interests

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

Acknowledgments

The work is supported by NSFC Joint Fund with Guangdong under Key Project U1201258, Shandong Natural Science Funds for Distinguished Young Scholar under Grant no. JQ201316, Program for New Century Excellent Talents in University of Ministry of Education of China (NCET-11-0315), and the Natural Science Foundation of Shandong Province (no. ZR2011FQ033).

References

[1] J. S. Duncan and N. Ayache, "Medical image analysis: progress over two decades and the challenges ahead," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 85–106, 2000.

[2] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: a new variational formulation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 430–436, June 2005.

[3] C. Li, C. Xu, A. W. Anderson, and J. C. Gore, "MRI tissue classification and bias field estimation based on coherent local intensity clustering: a unified energy minimization framework," in *Information Processing in Medical Imaging: Proceedings of the 21st International Conference, IPMI 2009, Williamsburg, VA, USA, July 5–10, 2009*, vol. 5636 of *Lecture Notes in Computer Science*, pp. 288–299, Springer, Berlin, Germany, 2009.

[4] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2007–2016, 2011.

[5] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 193–199, 2002.

[6] L. Yi and G. Zhijun, "A review of segmentation method for MR image," in *Proceedings of the International Conference on Image Analysis and Signal Processing (IASP '10)*, pp. 351–357, Zhejiang, China, April 2010.

[7] L. Cheng, J. Yang, X. Fan, and Y. Zhu, "A generalized level set formulation of the mumford-shah functional for brain MR image segmentation," in *Information Processing in Medical Imaging*, vol. 3565 of *Lecture Notes in Computer Science*, pp. 418–430, Springer, Berlin, Germany, 2005.

[8] X. Fan, J. Yang, and L. Cheng, "A Novel segmentation method for MR brainimages based on fuzzy connectedness and FCM," in *Fuzzy Systems and Knowledge Discovery*, vol. 3613 of *Lecture Notes in Computer Science*, pp. 505–513, Springer, Berlin, Germany, 2005.

[9] X. Fan, J. Yang, Y. Zheng, L. Cheng, and Y. Zhu, "A novel unsupervised segmentation method for MR brain images based on fuzzy methods," in *Computer Vision for Biomedical Image Applications*, vol. 3765 of *Lecture Notes in Computer Science*, pp. 160–169, Springer, Berlin, Germany, 2005.

[10] C. Zhu and T. Jiang, "Multi-context fuzzy clustering for separation of brain tissues in magnetic resonance images," *NeuroImage*, vol. 18, no. 3, pp. 685–696, 2003.

[11] D. L. Pham and J. L. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 9, pp. 737–752, 1999.

[12] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY, USA, 1981.

[13] L. Amini, H. Soltanian-Zadeh, C. Lucas, and M. Gity, "Automatic segmentation of thalamus from brain MRI integrating fuzzy clustering and dynamic contours," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 5, pp. 800–811, 2004.

[14] S. Shen, W. Sandham, M. Granat, and A. Sterr, "MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 3, pp. 459–467, 2005.

[15] S. P. Awate, H. Zhang, and J. C. Gee, "A fuzzy, nonparametric segmentation framework for DTI and MRI analysis: with applications to DTI-tract extraction," *IEEE Transactions on Medical Imaging*, vol. 26, no. 11, pp. 1525–1536, 2007.

[16] M. Halt, C. C. le Rest, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET," *IEEE Transactions on Medical Imaging*, vol. 28, no. 6, pp. 881–893, 2009.

[17] X. F. Zhang, C. M. Zhang, W. J. Tang, and Z. W. Wei, "Medical image segmentation using improved FCM," *Science China Information Sciences*, vol. 55, no. 5, pp. 1052–1061, 2012.

[18] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Patten Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2281, 2012.

[19] W. Wu, A. Y. C. Chen, L. Zhao, and J. J. Corso, "Brain tumor detection and segmentation in a CRF, (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features," *International Journal of Computer Assisted Radiology and Surgery*, 2013.

[20] M. Li, H. He, J. Yi, B. Lv, and M. Zhao, "Segmentation and tracking of coronary artery using graph-cut in CT angiographic," in *Proceedings of the 2nd International Conference on Biomedical Engineering and Informatics (BMEI '09)*, pp. 1–4, Tianjin, China, October 2009.

[21] L. Zhang and Q. Ji, "Image segmentation with a unified graphical model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1406–1425, 2010.

[22] L. Gan, Y. Wu, M. Liu, P. Zhang, H. Ji, and F. Wang, "Triplet markov fields with edge location for fast unsupervised multi-class segmentation of synthetic aperture radar images," *IET Image Processing*, vol. 6, no. 7, pp. 831–838, 2012.

[23] X. Zhou, X. Li, T.-J. Chin, and D. Suter, "Superpixel-driven level set tracking," in *Proceedings of the 19th IEEE International Conference on Image Processing (ICIP '12)*, pp. 409–412, Orlando, Fla, USA, October 2012.

[24] X. Gong and J. Liu, "Rock detection via superpixel graph cuts," in *Proceedings of the 19th IEEE International Conference on Image Processing (ICIP '12)*, pp. 2149–2152, Orlando, Fla, USA, October 2012.

[25] H. Jiang, "Solution to scale invariant global figure ground separation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 678–685, Providence, RI, USA, June 2012.

[26] Z. Li, X.-M. Wu, and S.-F. Chang, "Segmentation using superpixels: a bipartite graph partitioning approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 789–796, Providence, RI, USA, June 2012.

[27] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: fast superpixels using geometric flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2290–2297, 2009.

[28] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598, 1991.

[29] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[30] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *Computer Vision—ECCV: Proceedings of the 10th European Conference on Computer Vision, Marseille, France, October 12–18, 2008, Part IV*, vol. 5305 of *Lecture Notes in Computer Science*, pp. 705–718, Springer, Berlin, Germany, 2008.

[31] http://www.bic.mni.mcgill.ca/brainweb/.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 579652, 8 pages
http://dx.doi.org/10.1155/2014/579652

Research Article

Independent Component Analysis of Instantaneous Power-Based fMRI

Yuan Zhong,[1,2] Gang Zheng,[2] Yijun Liu,[3] and Guangming Lu[2]

[1] *School of Psychology, Nanjing Normal University, Nanjing 210097, China*
[2] *Department of Medical Imaging, Jinling Hospital, Nanjing University School of Medicine, Nanjing 210002, China*
[3] *Departments of Psychiatry and Neuroscience, McKnight Brain Institute, University of Florida, Gainesville, FL 32610, USA*

Correspondence should be addressed to Guangming Lu; cjr.luguangming@vip.163.com

Received 20 December 2013; Accepted 30 January 2014; Published 6 March 2014

Academic Editor: Rong Chen

In functional magnetic resonance imaging (fMRI) studies using spatial independent component analysis (sICA) method, a model of "latent variables" is often employed, which is based on the assumption that fMRI data are linear mixtures of statistically independent signals. However, actual fMRI signals are nonlinear and do not automatically meet with the requirement of sICA. To provide a better solution to this problem, we proposed a novel approach termed instantaneous power based fMRI (ip-fMRI) for regularization of fMRI data. Given that the instantaneous power of fMRI signals is a scalar value, it should be a linear mixture that naturally satisfies the "latent variables" model. Based on our simulated data, the curves of accuracy and resulting receiver-operating characteristic curves indicate that the proposed approach is superior to the traditional fMRI in terms of accuracy and specificity by using sICA. Experimental results from human subjects have shown that spatial components of a hand movement task-induced activation reveal a brain network more specific to motor function by ip-fMRI than that by the traditional fMRI. We conclude that ICA decomposition of ip-fMRI may be used to localize energy signal changes in the brain and may have a potential to be applied to detection of brain activity.

1. Introduction

Independent component analysis (ICA) is a data-driven approach that uses higher-order statistical moments to provide solutions to blind-source separation problems [1]. This technique has been shown to be capable of partitioning various physiological and physical signals in functional magnetic resonance imaging (fMRI) studies of brain activation [2–5]. There are primarily two types of ICA methods: (i) temporal ICA (tICA), which is used to detect specific changes in the time series of fMRI signals from brain regions of interest (ROIs) [6], and (ii) spatial ICA (sICA), which is often used to localize brain activity changes and has, so far, been the dominant ICA method used in fMRI applications [7]. Here, we focus only on the sICA method for fMRI data analysis.

In fMRI data, blood oxygenation level-dependent (BOLD) signals represent brain activity changes and can be expressed by fluctuations of T_2^* signals [8]. The BOLD signal is a complex function of neural activity, oxygen metabolism, cerebral blood volume, cerebral blood flow, and other physiological parameters. A basic assumption of sICA for fMRI data analysis is that the observed fMRI signals are a linear sum of various components separated at each voxel [9]. However, The dynamics underlying neural activity and hemodynamic physiology are believed to be nonlinear [10, 11], and they do not automatically satisfy the commonly used latent variables model (see below). Although there are always computational solutions of sICA for fMRI data, they are nonunique [12]. Hence, in order to use sICA for localization of brain activity changes, the problem of converting nonlinear signals to linear ones needs to be solved.

Although unique nonlinear ICA has been proposed in previous studies [13] and there are different approaches for regularizing ICA solutions [12, 14], so far these methods have rarely been applied to fMRI studies. For example, the methods for transforming postnonlinear mixtures in ICA to invertible linear mixtures have been established [15, 16]. In

addition, convolutive postnonlinear mixtures can be treated in the same manner as certain postnonlinear ones, which can then be transformed to linear mixtures [12]. A fundamental difficulty of the nonlinear ICA problem-solving is that its solutions are nonunique if there is no suitable regularization. More generally, a basic method for solving nonlinear ICA problems is to transform nonlinear variables or measurements into linear mixtures such that nonlinear ICA problems can be reduced to traditional linear ICA problems.

In the present study, we propose using instantaneous power as a new regularization approach in ICA for transforming nonlinear fMRI signals to linear forms in order to automatically satisfy the assumption in the "latent variable" ICA model. The instantaneous power-based fMRI (ip-fMRI) approach defines the energy of fMRI signals by their inner products such that the signal energy can be represented by a variance of fMRI signals, which is regarded as integral to their instantaneous power. Based on the regulation, the instantaneous power of fMRI signals is then partitioned into independent components using conventional sICA. There are three steps in establishing our method. First, we briefly describe the theories of ICA and the instantaneous power of fMRI signals. Second, based on the resulting accuracy and receiving operator characteristic (ROC) curves, we describe how simulated fMRI data can be used to evaluate the performance of the ip-fMRI approach and compare it with the traditional fMRI results. Third, we describe how the new approach was further tested by applying it to human data for the analysis of task-induced brain activations.

2. Theory

2.1. The Latent Variable Model of ICA. In classical ICA methods, a statistical "latent variable" model [9, 17] is often used based on the assumption that observed random variables $[x(1,l),\ldots,x(t,l),\ldots,x(T,l)]$ are linear mixtures of latent variables $[s(1,l),\ldots,s(m,l),\ldots,s(M,l)]$ that are non-Gaussian and mutually independent [9]. Consider the following:

$$\begin{aligned} x(1,l) &= a_{11}s(1,l) + a_{12}s(2,l) \\ &\quad + \cdots a_{1m}s(m,l) + \cdots a_{1M}s(M,l) \\ &\vdots \\ x(t,l) &= a_{t1}s(1,l) + a_{t2}s(2,l) \\ &\quad + \cdots a_{tm}s(m,l) + \cdots a_{tM}s(M,l) \\ &\vdots \\ x(T,l) &= a_{T1}s(1,l) + a_{T2}s(2,l) \\ &\quad + \cdots a_{Tm}s(m,l) + \cdots a_{TM}s(M,l). \end{aligned} \tag{1}$$

In the matrix notation, this can be written simply as

$$X = AS, \tag{2}$$

(see [18]), where $X = [x(1,l),x(2,l),\ldots,x(T,l)]^T$ is a vector of observed variables, $S = [s(1,l),s(2,l),\ldots,s(M,l)]^T$ is a vector of latent independent sources, and $A = [a_{t1},a_{t2},\ldots,a_{tM}]$ is the mixing matrix.

In sICA, the row of X corresponds to the voxels in fMRI signals and the column denotes time series. $l = 1,2,\ldots,L$ is the spatial index of voxels in one volume, where L is the total number of voxels, $t = 1,2,\ldots T$ is the temporal index of fMRI time series, and $m = 1,2,\ldots,M$ is the component index. In addition, the total time points of fMRI signals T are no less than the total number of components M according to the problem-solving processes of blind-source separation.

2.2. The Energy and the Instantaneous Power of fMRI Signals. Suppose that $P_x(1,l),P_x(2,l),\ldots,P_x(t,l),\ldots,P_x(T,l)$ represent the energies of observed fMRI signals $x(1,l),\ldots,x(t,l),\ldots,x(T,l)$, respectively. Because $x(t,l)$ represents the signal in one voxel at time t, the energy term $P_x(t,l)$ is an instantaneous value and can be expressed by

$$\begin{aligned} P_x(t,l) &= a_{t1}P_s(1,l) + a_{t2}P_s(2,l) \\ &\quad + \cdots a_{tm}P_s(m,l) + \cdots a_{tM}P_s(M,l), \end{aligned} \tag{3}$$

where $P_s(1,l),P_s(2,l),\ldots,P_s(m,l),\ldots P_s(M,l)$ are the instantaneous powers of signals $s(1,l),\ldots,s(m,l),\ldots,s(M,l)$, respectively. The overall energy of observed data in each voxel can then be represented by

$$E_x(l) = \sum_{t=1}^{T} P_x(t,l). \tag{4}$$

Because the mixing matrix is a normalized weight matrix [9], the energy of observed data is nearly equal to the energy of source signals (supposing that the spatially and temporally white noise are eliminated from observed data) and can be written as

$$E_s(l) = \sum_{m=1}^{M} P_s(m,l) \approx E_x(l). \tag{5}$$

Therefore, even if source signals are nonlinear, their energy signals $P_s(1,l),P_s(2,l),\ldots,P_s(m,l),\ldots,P_s(M,l)$ are additive, which satisfies the requirement of the "latent variable" model of sICA. Then, the remaining question is how to define the energy and instantaneous power of fMRI signals.

To give the definitions, the concept of electric energy can be used as an analogy. In fMRI, if the observed T_2^* signals are taken as the instant voltage or current fluctuation of the resistance, we can define the energy of T_2^* signals by the inner product of the signals, which can be written as

$$E_{T_2^*}(l) = \langle x,x\rangle = \int x^2(t,l)\,dt. \tag{6}$$

Because the energy of fMRI signals is associated with the variations of T_2^* signals [8], the temporal variance of T_2^* signals can be used to define the overall energy of fMRI signals by

$$E_{\mathrm{BOLD}} = \int [x(t,l) - \overline{x}(l)]^2 dt, \tag{7}$$

where $x(t,l)$ is the voxel-wise fMRI signal intensity. $\overline{x}(l) = \int x_r(t,l)dt/T$ is the signal baseline value, which is the mean BOLD value crossing the time course in fMRI during an experimental resting condition and can be obtained through a temporal normalization procedure [19]. According to (5), the instantaneous power of fMRI signals can then be represented by

$$P_{\mathrm{BOLD}}(t,l) = \frac{\partial E_{\mathrm{BOLD}}(l)}{\partial t} = [x(t,l) - \overline{x}(l)]^2. \quad (8)$$

Given that BOLD signals are composed of M independent components of brain activity according to the basic hypothesis underlying the ICA approaches used in fMRI studies, the instantaneous power of BOLD signals can be partitioned into M independent instantaneous powers of brain activity by sICA, as expressed by

$$\begin{aligned} P_{\mathrm{BOLD}}(t,l) = {} & a_{t1}P_{\mathrm{activation\_1}}(1,l) + a_{t2}P_{\mathrm{activation\_2}}(2,l) \\ & + \cdots a_{tm}P_{\mathrm{activation\_}m}(m,l) \\ & + \cdots a_{tM}P_{\mathrm{activation\_}M}(M,l). \end{aligned} \quad (9)$$

Because the instantaneous power of fMRI signals can be considered as a linear mixture of each instantaneous power of brain activity and normally meets with the latent variables model of sICA, it is more suitable to use instantaneous power of fMRI signals partitioned by sICA than to use original fMRI signals.

3. Materials and Methods

3.1. Participants. Eighteen healthy volunteers (11 males; mean age 27.5 years; age range: 22–35 years) participated. The healthy subjects had no history of neurological or psychiatric disorders and were not on any medication for at least a month before the experiment. All the participants were right-handed as assessed using the Edinburgh handedness inventory [20]. The study was approved by the local Ethical Committee of Jinling Hospital, and written informed consent was obtained from all subjects prior to participating.

3.2. Experimental Paradigms. In the experiment, healthy subjects were scanned when performing a hand flexion task using their nondominant (left) hand [21]. The subjects were trained to grip the hand with a frequency of 1 Hz, and they practiced for 100 sec before the scan. A block design was used in the paradigm and the overall task consisted of 10 blocks, 5 task blocks alternating with 5 resting blocks, lasting for 200 sec with each block of 20 sec. During the functional scan, the subjects were instructed to grip the left hand when seeing a stationary cross presenting on the center of the screen throughout each task block and to remain still and fixate on a stationary asterisk throughout each resting block. The paradigm has been expounded in our precious study [21].

3.3. MRI Acquisition. Imaging data were acquired using a 1.5T GE MRI system (Signa) at Jingling Hospital, Nanjing, China. Foam padding was used to minimize head motion and improve participants comfort. fMRI time series of 100 repeated whole brain images were acquired in an orientation parallel to the AC-PC plane using a T_2^*-weighted GRE-EPI sequence. The sequence parameters were TR = 2000 ms, TE = 40 ms, FA = 80°, FOV = 24 × 24 cm^2, 21 continuous slices with a thickness of 4 mm (no gap), and matrix size = 3.75 3.75 mm^2. Anatomical images using a T1-Flair sequence (TR = 2019.3 ms, TE = 25.3 ms, interslice gap = 0.5 mm, and slice thickness = 4 mm) were acquired to facilitate the precise determination of the structures corresponding to the functional regions.

3.4. Preprocessing of Data. The fMRI data of each subject were first preprocessed using SPM8 software package (http://www.fil.ion.ucl.ac.uk/spm/), and spatial realignment was performed to remove head motion artifacts, and the functional scans were spatially normalized to a standard template (Montreal Neurological Institute) and resampled to $2 \times 2 \times 2$ mm^3. If the head motion and rotation parameters of a subject exceeded ±0.5 mm and ±0.5°, respectively, the data was excluded from further analysis. To increase the signal-to-noise ratio (SNR), the data were smoothed spatially using an isotropic Gaussian filter with a full width at half-maximum (FWHM) of 8 mm kernel.

3.5. ICA Analysis. Data from all participants were concatenated into a single dataset and reduced using two stages of principal component analysis (PCA) [7]. The optimal number of ICs was determined by a dimension estimation using the minimum description length (MDL) criterion [22]. ICA was then conducted to decompose the data from all subjects into different spatially independent components (ICs) with the FAST-ICA algorithm. For each IC, the time course corresponded to the waveform of a specific pattern of coherent brain activity, and the intensity of the pattern across the voxels is expressed in the associated spatial map. This analysis was repeated 50 times for assessing the repeatability of ICs [23]. To display the voxels that contributed most strongly to a particular IC, the intensity values in each spatial map were converted to Z-values (standard deviation of image) map [24]. The voxels with absolute ICA amplitudes larger than a specified amplitude threshold (i.e., |amplitude| > 2.5) were selected as the voxels with significant changes in brain activity [25].

3.6. Simulation. A simulation was conducted to evaluate the performance of ICA decomposition of ip-fMRI. A slice of resting-state EPI scans (79 × 95 voxels) was replicated 200 times in order to simulate 200 time points of noise-free fMRI data. Nine 8 × 8 square blobs of voxels were selected for the simulation of localized activity changes (Figure 1(a)). The simulated time courses used in this section are shown in Figure 1(b). Three simulated signals (Signals (A)–(C)) were constructed to represent the brain hemodynamics for event-related activation (Signal (A)), resting-state activities (Signal (B)), and activation in block-designed paradigm (Signal (C)). A slowly varying baseline (Signal (D)) was added to all the voxels. To simulate the noisy environment in

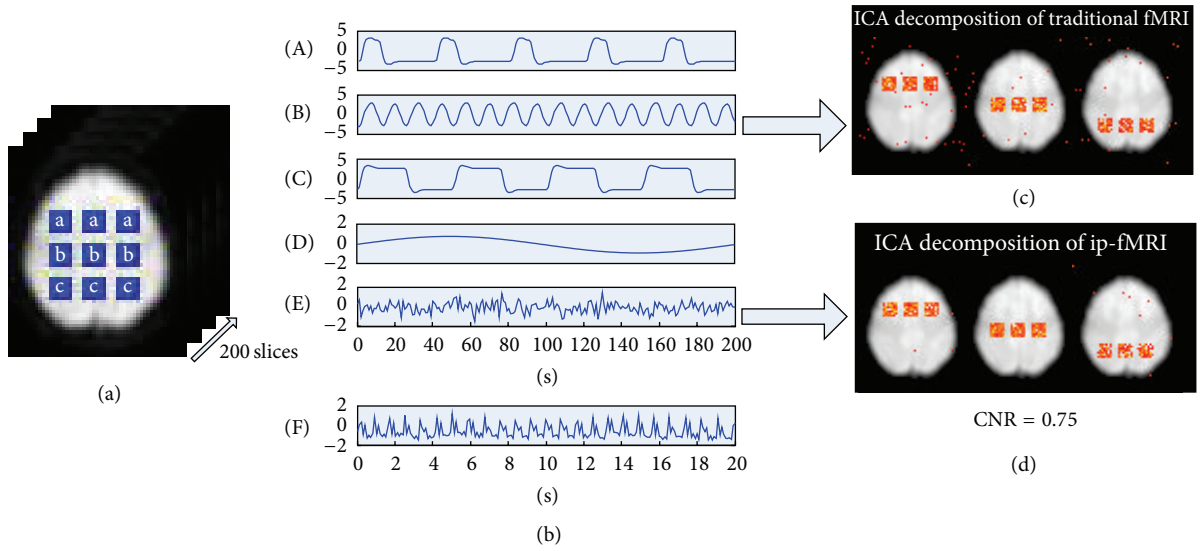


Figure 1: (a) AnEPI scan image with the selected nine regions of interest and simulated fMRI time courses. (b) Signals (A) and (B) (0.08 Hz) are the series of gamma variate functions simulating event-related and resting-state brain hemodynamics, respectively. Signal (C) is the convolution of an HRF and a square wave simulating a block-design fMRI signal. Signal (D) (0.005 Hz) is a sine wave simulating a slowly varying global baseline. Signal (E) is a Gaussian signal simulating the random white noise. Signal (F) (with a mean frequency of 1.2 Hz) is a cardiac signal simulating the structured noise. (c) Spatial components extracted by traditional fMRI. (d) Spatial components extracted by ip-fMRI. Voxels with amplitude values above threshold 2.5 are shown as the points in red to yellow color on the image.

the brain, random noise and structured noise were mixed in the simulated data (all voxels in the brain area). The random noise (Signal (E)) follows a Gaussian distribution with a mean of 0 and a variance of 1. To simulate the structured noise, a cardiac signal (Signal (F)) which has an average frequency of 1.2 Hz was generated. The magnitudes of the signals and noises have been varied to adjust the contrast-to-noise ratio (CNR $\equiv \Delta S/\sigma_{\text{noise}}$). The CNRs ranged from 0.5 to 2, consistent with the CNR values reported in the fMRI literature [26].

The ip-fMRI and fMRI datasets were decomposed separately into spatially independent spatial patterns. Each pattern was associated with a temporal waveform, and only the components with the closest correlation between waveform and simulated true sources were considered. Examples of the spatial components decomposed by the two approaches are shown in Figures 1(c) and 1(d) with the CNR of 0.75. A receiver-operating characteristic (ROC) analysis was then conducted based on the spatial maps to determine the optimal component number, accuracy, and specificity of the two approaches.

4. Results

4.1. Simulation Results. The accuracy was calculated for both the fMRI and ip-fMRI approaches based on our simulated dataset at different ICA amplitude levels and CNRs. Under almost all the conditions tested, ip-fMRI produced more suited results than the traditional fMRI, which was especially evident with low CNRs (Figure 2). In addition, each individual component separated from ip-fMRI was almost consistent when being used to localize the ROIs. Further comparisons of the ROC curves showed that those of ip-fMRI plotted were always over those of traditional fMRI (Figure 3). In terms of specificity, the AUC value of 1 represents a perfect test; while an AUC value of 0.5 or below just gives a chance discrimination. When CNR was set to be 0.5, the traditional fMRI failed to produce reliable results (AUC = 0.4896) while ip-fMRI still performed well (AUC = 0.9751). The simulation results indicate that the ICA decomposition from ip-fMRI outperforms that from the traditional fMRI, even though both of the approaches have an equal AUC value (i.e., = 1) when CNRs are set above 1.

4.2. Experimental Results

4.2.1. Identification of the Functional Network Underlying Hand Movement. The performance of the ICA decomposition of ip-fMRI was further evaluated for the detection of task-induced brain activation. A motor network underlying the left hand flexion task was identified using either approach, which consists of the contralateral primary motor (M1) and sensory (SI) cortices and the supplementary motor area [27]. However, the spatial component extracted from ip-fMRI from each subject became more anatomically focused or more specific to motor function than those from traditional fMRI. For the visualization, the motor component from a randomly selected subject was shown in Figure 4. Furthermore, the correlation coefficient between the temporal waveform associated with each spatial motor component and the designed ON-OFF paradigm was calculated. Two-sample t-test was employed to compare the results between ip-fMRI and traditional fMRI methods. Compared with the traditional fMRI, the result obtained from the two-sample

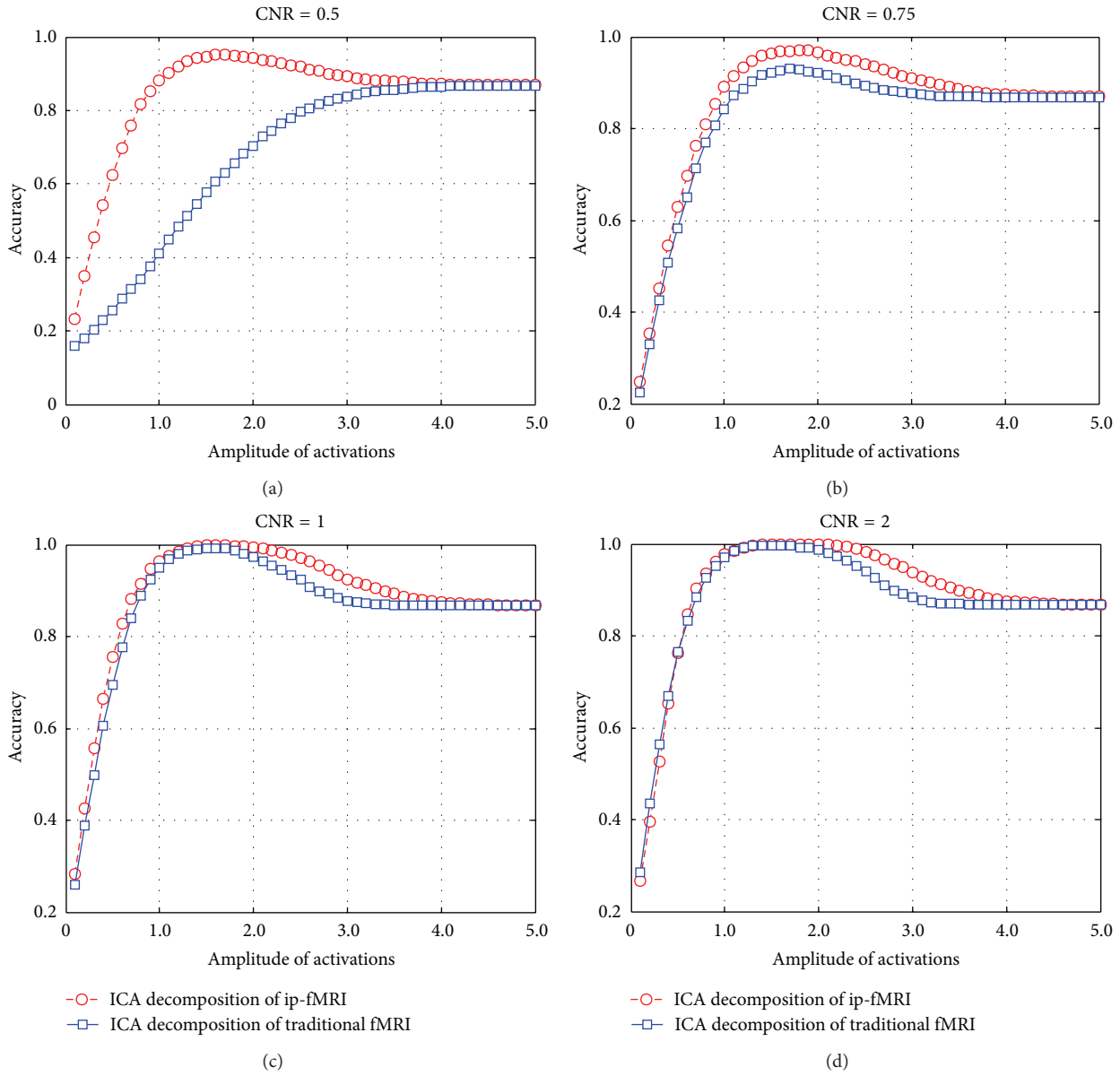


Figure 2: The accuracy analyses of ICA decomposition from ip-fMRI and traditional fMRI at CNRs of 0.5, 0.75, 1, and 2. The plots are the average accuracy curves of the fifty repeated procedures.

t-test clearly showed significant difference ($t = 3.0486$, $P = 0.0055$, uncorrected). This enabled the identification of significant changes in the ip-fMRI method as compared with the traditional fMRI.

5. Discussion

ICA provides a method to separate signals "blindly" into spatially independent components, enabling exploratory analysis of fMRI data [2]. The key assumptions in sICA are that an fMRI dataset consists of a number of spatially independent components that are linearly mixed and spatially fixed. However, fMRI signals are actually nonlinear and are affected by many other artifacts such as those induced by head motion or physiological noises. Thus, fMRI signals may not automatically satisfy the commonly used latent ICA variables model.

To provide a better solution to this problem, we present an instantaneous power approach to resolving the ICA problem in fMRI analysis. We used instantaneous power to regularize fMRI signals such that the distribution of fMRI signal changes follows the temporal pattern in power distribution (defined in (8)). Then, the decompositions separated by ipICA can be simultaneously used to extract a variety of spatially independent components. The spatially independent but temporally coherent components represent the instantaneous power of each fMRI source signal (i.e., changes in brain activity). In other words, because the power of fMRI signals

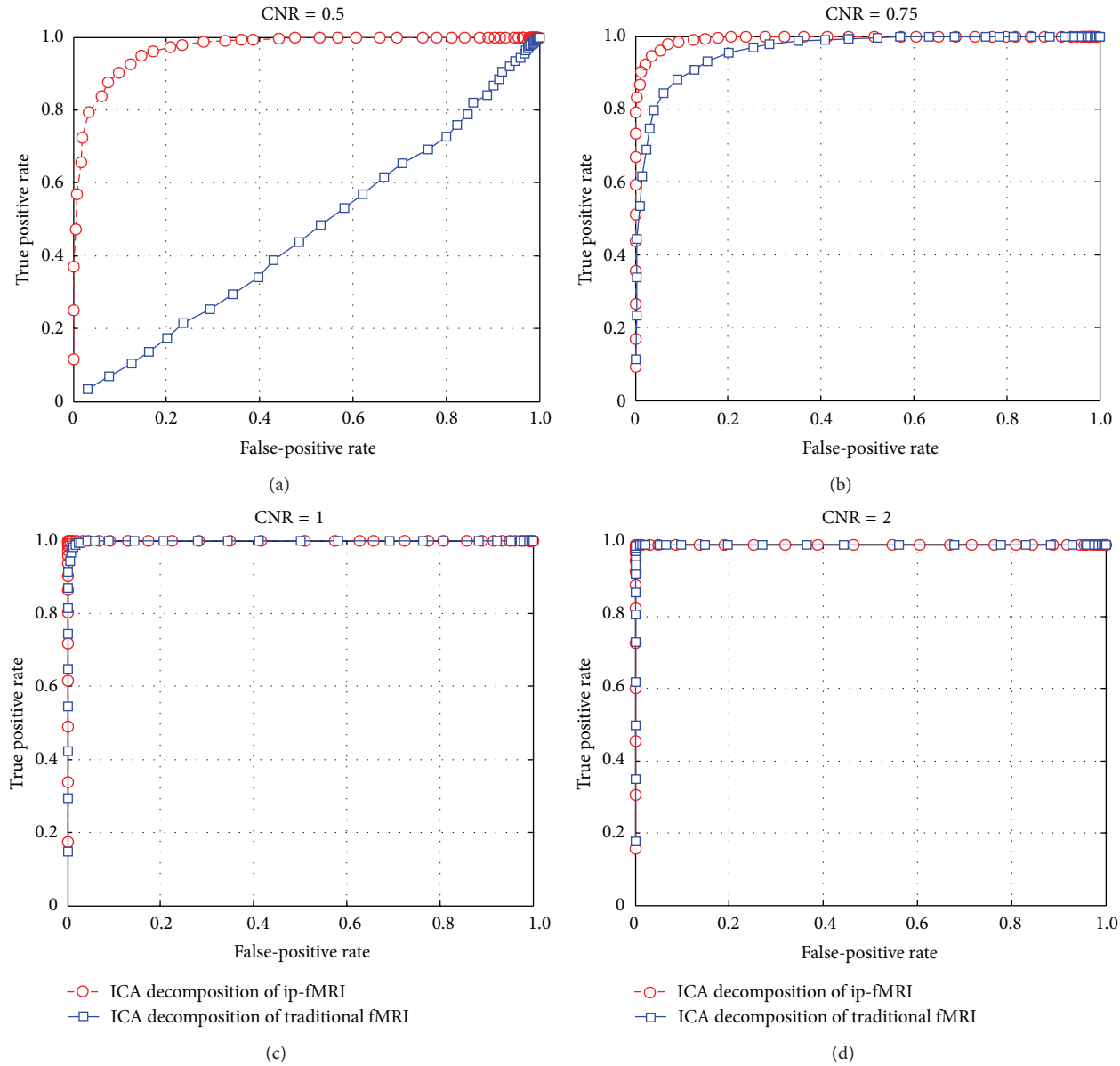


Figure 3: The ROC curves for ICA decomposition of ip-fMRI and traditional fMRI. Plotted are the mean ROC curves of the fifty repeated procedures. For the simulation with a CNR set to be 0.5, the area under the ROC curve (AUC) is 0.9751 and 0.4896 for ip-fMRI and traditional fMRI, respectively. Briefly, CNR = 0.75, AUC = 0.9932 (ip-fMRI) and 0.9723 (traditional fMRI); CNR = 1, AUC = 1 (ip-fMRI) and 0.9993 (traditional fMRI); CNR = 2, AUC = 1 (ip-fMRI) and 1 (traditional fMRI).

can be considered as a linear mixture of each instantaneous power, the components separated by ip-fMRI naturally satisfy the addition theorem to reflect different patterns of brain activity.

We have used a relatively realistic simulation based on a single volume of resting-state EPI data. This simulation has correct noise structures and spatial and temporal correlations with three artificial components added, which are shown in 9 dominant square blobs of the regions simulated. To make a visual comparison between the ICA results decomposed by ip-fMRI and traditional fMRI, Figures 1(c) and 1(d) show the components correlated to the simulated time series, respectively. For a given amplitude threshold, the ICA map from traditional fMRI tends to be noisier than that from ip-fMRI with the CNR of 0.75. Our results from the accuracy curves (Figure 2) and ROC curves (Figure 3) indicate that the performance of ip-fMRI is superior to that of the traditional fMRI under different CNRs or ICA amplitude values.

The proposed new ip-fMRI approach has been further validated using human data. A task-related fMRI experiment was provided for evaluating the new approach. When the actual mixtures are regularized through instantaneous power,

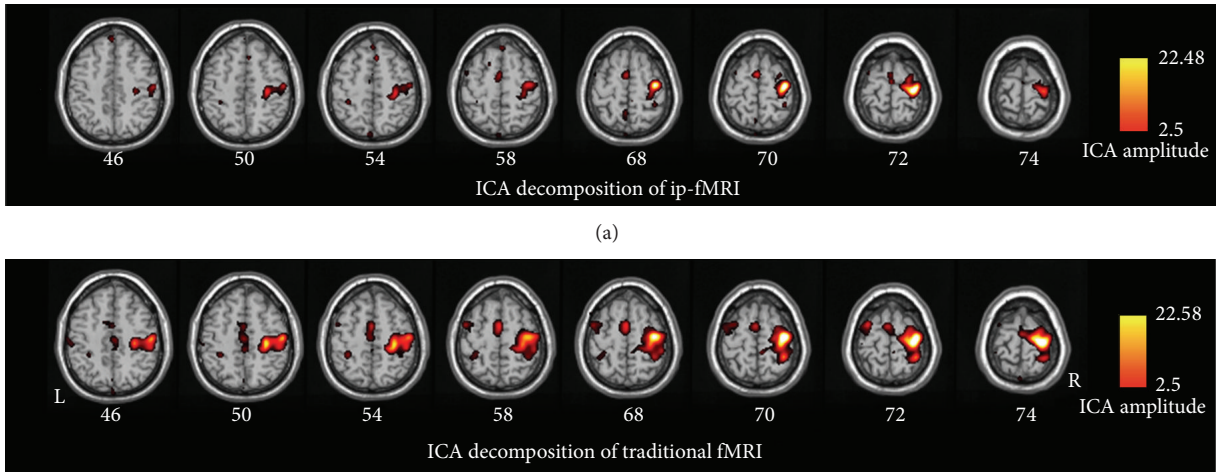


Figure 4: The motor functional networks of a random selected subject detected by ip-fMRI and traditional fMRI methods, respectively. (a) Motor networks detected by ip-fMRI. (b) Motor networks detected by traditional fMRI. The numbers beneath the axial MR image refer to Talairach coordinates.

the extracted spatial sources from the regularized results become more anatomically focused than those without the regulation (Figure 4) and their time courses become more fit into the designed paradigm for hypothesis testing.

In summary, we introduce a new ICA method based on the instantaneous power of fMRI signals to improve the decomposition and interpretation of fMRI data. The decomposed components by ip-fMRI represent the distribution of instantaneous power changes in fMRI signals. Combining the simulated and *in vivo* fMRI data, our results indicate that the spatially independent components decomposed from ip-fMRI are superior to those decomposed from the traditional fMRI in both accuracy and specificity for detecting the brain activity changes. We conclude that the ICA decomposition of ip-fMRI approach may provide a tool for the localization of energy changes in the brain, which may potentially be used to detect altered brain functions.

Conflict of Interests

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

Acknowledgments

This work was supported by grants from the Natural Scientific Foundation of China (Grant no. 81201077 to Yuan Zhong and 81101039 to Gang Zheng), the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Grant no. 11YJC190039 to Yuan Zhong), the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province, China (Grant no. 12kjb310002 to Yuan Zhong), and a Chinese Key Grant for GML (Grant nos. BWS11J063 and 10z026). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the paper.

References

[1] P. Comon, C. Jutten, and J. Herault, "Blind separation of sources—II: problems statement," *Signal Processing*, vol. 24, no. 1, pp. 11–20, 1991.

[2] A. Hyvarinen, "Independent component analysis: recent advances," *Philosophical Transactions of the Royal Society A*, vol. 371, no. 1984, Article ID 0534, 2013.

[3] V. Kiviniemi, J.-H. Kantola, J. Jauhiainen, A. Hyvärinen, and O. Tervonen, "Independent component analysis of nondeterministic fMRI signal sources," *NeuroImage*, vol. 19, no. 2, pp. 253–260, 2003.

[4] M. J. McKeown and T. J. Sejnowski, "Independent component analysis of fMRI data: examining the assumptions," *Human Brain Mapping*, vol. 6, pp. 368–372, 1998.

[5] J. Xu, S. Zhang, V. D. Calhoun et al., "Task-related concurrent but opposite modulations of overlapping functional networks as revealed by spatial ICA," *Neuroimage*, vol. 79, pp. 62–71, 2013.

[6] B. B. Biswal and J. L. Ulmer, "Blind source separation of multiple signal sources of fMRI data sets using independent component analysis," *Journal of Computer Assisted Tomography*, vol. 23, no. 2, pp. 265–271, 1999.

[7] V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Human Brain Mapping*, vol. 13, no. 1, pp. 43–53, 2001.

[8] S. Ogawa, R. S. Menon, D. W. Tank et al., "Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model," *Biophysical Journal*, vol. 64, no. 3, pp. 803–812, 1993.

[9] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[10] R. M. Birn, Z. S. Saad, and P. A. Bandettini, "Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response," *NeuroImage*, vol. 14, no. 4, pp. 817–826, 2001.

[11] S. Faisan, L. Thoraval, J.-P. Armspach, and F. Heitz, "Hidden Markov multiple event sequence models: a paradigm for the spatio-temporal analysis of fMRI data," *Medical Image Analysis*, vol. 11, no. 1, pp. 1–20, 2007.

[12] C. Jutten and J. Karhunen, "Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear mixtures," *International Journal of Neural Systems*, vol. 14, no. 5, pp. 267–292, 2004.

[13] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: existence and uniqueness results," *Neural Networks*, vol. 12, no. 3, pp. 429–439, 1999.

[14] C. Jutten, M. Babaie-Zadeh, and S. Hosseini, "Three easy ways for separating nonlinear mixtures?" *Signal Processing*, vol. 84, no. 2, pp. 217–229, 2004.

[15] S. Achard, D. T. Pham, and C. Jutten, "Blind source separation in post nonlinear mixtures," in *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation (ICA '01)*, pp. 295–300, Citeseer, 2001.

[16] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Blind separating convolutive post-nonlinear mixtures," in *Proceedings of the ICA*, pp. 138–143, 2001.

[17] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[18] K. E. Hild II and S. S. Nagarajan, "Source localization of eeg/meg data by correlating columns of ica and lead field matrices," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2619–2626, 2009.

[19] G. He, J. Mao, A. Perry, J. Gao, P. Fox, and Y. Liu, "The temporal constancy of resting brain activity implicated in fMRI studies," in *Proceedings of International Society of Magnetic Resonance in Medicine*, vol. 9, pp. 17–25, 2001.

[20] R. C. Oldfield, "The assessment and analysis of handedness: the Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, 1971.

[21] Y. Zhong, H. Wang, G. Lu, Z. Zhang, Q. Jiao, and Y. Liu, "Detecting functional connectivity in fmri using pca and regression analysis," *Brain Topography*, vol. 22, no. 2, pp. 134–144, 2009.

[22] Y.-O. Li, T. Adali, and V. D. Calhoun, "Estimating the number of independent components for functional magnetic resonance imaging data," *Human Brain Mapping*, vol. 28, no. 11, pp. 1251–1266, 2007.

[23] J. Himberg, A. Hyvärinen, and F. Esposito, "Validating the independent components of neuroimaging time series via clustering and visualization," *NeuroImage*, vol. 22, no. 3, pp. 1214–1222, 2004.

[24] V. D. Calhoun, T. Adali, V. B. McGinty, J. J. Pekar, T. D. Watson, and G. D. Pearlson, "fMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis," *NeuroImage*, vol. 14, no. 5, pp. 1080–1088, 2001.

[25] M. J. McKeown, S. Makeig, G. G. Brown et al., "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapping*, vol. 6, pp. 160–188, 1998.

[26] F. Esposito, E. Formisano, E. Seifritz et al., "Spatial independent component analysis of functional MRI time-series: to what extent do results depend on the algorithm used?" *Human Brain Mapping*, vol. 16, no. 3, pp. 146–157, 2002.

[27] M. Bosevski, V. Borozanov, M. Vavlukis, G. Pemovska, and L. Georgievska-Ismail, "Carotid ultrasound, blood lipids and waist determination can predict a future coronary revascularisation in the type 2 diabetic cohort," *Prilozi*, vol. 28, no. 2, pp. 127–136, 2007.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 938350, 8 pages
http://dx.doi.org/10.1155/2014/938350

Research Article

Deep Learning Based Syndrome Diagnosis of Chronic Gastritis

Guo-Ping Liu,[1] Jian-Jun Yan,[2] Yi-Qin Wang,[1] Wu Zheng,[1] Tao Zhong,[2] Xiong Lu,[3] and Peng Qian[1]

[1] *Basic Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[2] *Center for Mechatronics Engineering, East China University of Science and Technology, Shanghai 200237, China*
[3] *Technologies and Experiment Center, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*

Correspondence should be addressed to Guo-Ping Liu; guoping2013penn@gmail.com and Jian-Jun Yan; jjyan@ecust.edu.cn

Received 22 November 2013; Accepted 10 January 2014; Published 5 March 2014

Academic Editor: Yuanjie Zheng

In Traditional Chinese Medicine (TCM), most of the algorithms used to solve problems of syndrome diagnosis are superficial structure algorithms and not considering the cognitive perspective from the brain. However, in clinical practice, there is complex and nonlinear relationship between symptoms (signs) and syndrome. So we employed deep leaning and multilabel learning to construct the syndrome diagnostic model for chronic gastritis (CG) in TCM. The results showed that deep learning could improve the accuracy of syndrome recognition. Moreover, the studies will provide a reference for constructing syndrome diagnostic models and guide clinical practice.

1. Introduction

In recent years, the standardization and objectification of TCM diagnosis have gradually became a research hotspot with the development of mathematical statistics, data mining, and pattern recognition technology. Many researches are emerged in large numbers. An entropy-based partition method for complex systems is applied to establish endothelial dysfunction diagnostic criteria for Yin deficiency syndrome. Moreover, the experimental results are highly consistent with the findings of clinical diagnosis [1]. Su et al. [2] employed the correlation coefficient, similarity D, the angle cosine, and spectral similarity to study the correlation between the symptoms (signs) and the five syndromes of liver cirrhosis. The research can provide a basis for differentiating patients with nonspecific clinical manifestations. Multilabel learning [3] combined with the feature selection had been used to improve the syndrome recognition rate of chronic gastritis.

Although a large number of machine learning methods have been used in the standardization and objectification of TCM diagnosis, researchers can provide a reference for clinical syndrome differentiation. However, in clinical practice, diagnosis of TCM is from the brain and has some hierarchical nature, complexity, and nonlinearity. There is a complex and nonlinear relationship between symptoms (signs) and syndrome. Most of the algorithms are not considering the hierarchical nature of diagnosis from the brain's cognitive perspective. This is likely to cause misunderstanding and bias.

Inspired by the hierarchical structure of the brain, neural network researchers have been working on multilayer neural network. Back propagation algorithm (BP) is a classical multilayer network algorithm, but the theoretical and experimental results showed that BP was not suitable for training the data with multiple hidden layer units [4]. Traditional machine learning and signal processing techniques were only to explore the shallow structure containing a single layer and nonlinear transformation. Typical shallow layer learning included traditional hidden Markov model (HMM), conditional random fields (CRF), maximum entropy model (MaxEnt), and support vector machine (SVM). The function ability of representing shallow layer structure has its limitations. However, deep learning [5] can succinctly represent complex functions.

Hinton Research Group proposed the deep network and deep learning concept in 2006. Hinton et al. [6, 7] proposed

unsupervised training drill greedy algorithm for solving optimization problems and then proposed the automatic multiencoder deep belief networks based on the deep structure (DBN). LeCun et al. [8] proposed convolutional neural networks (CNNs), the first true multilayer structure learning algorithms, which use relative spatial relationships, reducing the number of parameters to improve the performance of BP training. In addition, the study of deep learning also appeared in many deformed structures such as automatic denoising encoder [9, 10], DCN [11] and sum-product [12]. Deep learning method has been applied to machine vision [13–15], speech recognition [16, 17], and other areas to improve data classification and identification of effects and set off a new craze machine field.

Deep learning is distinctly more in line with the human brain thinking; it can use high-dimensional abstract features to express some of the original low-dimensional features. It is a good method to find the relationship between the symptoms each other and between the symptoms and syndromes. This idea is consistent with the diagnosis ideas of TCM.

At the same time, patients may simultaneously have more than one syndrome in clinical practice. Therefore, in this paper, we proposed to apply the deep learning method to establish the multilabel learning model of CG. Through the deep learning algorithm, we try to find a complex and nonlinear relationship between symptoms and syndromes of CG and to improve the syndrome cognition rate of CG.

2. Material and Methods

2.1. Research Subjects. Chronic gastritis (CG) samples were collected from a clinic, inpatient department, and gastroscopy room of the digestive system department of the Longhua Hospital, the Shuguang Hospital of Shanghai University of Traditional Chinese Medicine, the Xinhua Hospital, the Putuo District Central Hospital, and the Shanghai Hospital of Traditional Chinese Medicine. The Shanghai Society of Medical Ethics approved this work. All patients signed an informed consent form. A total of 919 valid subjects were enrolled after excluding cases with TCM inquiry diagnosis scales that lacked information or cannot be diagnosed with CG. Among the 919 patients, 354 were male (38.5%, with an average age of 44.61 yr ± 14.54 yr) and 565 were female (61.5%, with an average age of 48.70 yr ± 12.74 yr).

2.2. Inclusion Criteria. Patients who met the diagnostic standards for CG and TCM syndromes and patients who were informed and have agreed to join this investigation were included.

2.3. Diagnostic Criteria. Western Diagnostic Standards include diagnosing whether a patient has CG based on gastroscopy results, pathologic results, and clinical performance, according to the Consensus of National Seminar on CG held by the Chinese Medical Association Digestive Diseases Branch in 2007 [18].

Chinese Diagnostic Standards include the following eight syndromes (patterns):

(1) damp heat accumulating in the spleen-stomach;

(2) dampness obstructing the spleen-stomach;

(3) spleen-stomach qi deficiency;

(4) spleen-stomach cold deficiency;

(5) liver stagnation;

(6) stagnated heat in the liver-stomach;

(7) stomach Yin deficiency;

(8) blood stasis in the stomach collateral.

We referred to the diagnoses in "Guideline for Clinical Research of New Traditional Chinese Medicine" [19] issued by the Ministry of Health and "National Standard of People's Republic of China: Syndrome Part of TCM Clinical diagnosis and Treatment Terminology" [20] issued by the China State Bureau of Technical Supervision.

2.4. Exclusion Criteria

(1) mentally ill patients and patients with other severe systemic diseases;

(2) patients who have difficulty in describing their conditions;

(3) patients who are not informed or refuse to cooperate.

2.5. Method for Establishing TCM Inquiry Diagnosis Scales. The research group was composed of Shanghai senior clinical experts on the digestive system, clinical doctors, and researchers. The final TCM inquiry diagnosis scales were drafted based on past experience in the production of scales [21], a wide range of literature about TCM spleen and stomach diseases, related documents in core magazines and journals for over 15 years, and reports about the frequency of symptoms associated with syndromes in CG diseases in TCM. The scales were also amended and fixed by two rounds of expert consultation and statistical tests. The scales include eight dimensions such as cold or heat, sweat, head, chest and abdomen, urine and stool, diet and taste, sleep, mood, woman aspects, and contents of disease history, inspection, and palpation. More than 113 variables were ultimately included in these scales.

2.6. Investigation Methods. The clear definitions of symptoms, the specific methods, and the order of inquiry diagnosis were given in the scales. All samplers must have undergone unified training. The group members assemble regularly and discuss the information of typical patients to ensure the consistency of the collected data.

2.7. Diagnosis Methods. Three senior chief doctors with plenty of experience in clinical practices were invited for inquiry diagnosis of the cases in terms of the CG diagnostic standards made by our research group. If two of them have

the same diagnosis results, the case was included. Otherwise, the case was not adopted until at least two of them came to the same conclusion.

2.8. Data Input and Process Methods

(1) Build a database with Epidata software.

(2) Input data two times independently.

(3) The Epidata software compares the two data sets and checks out mistakes.

(4) Check the investigation form logically in case of filling errors.

2.9. Analysis Methods

2.9.1. Multilabel Learning Based on Deep Learning. Deep belief network (DBN) is a deep architecture, which is suitable to deliver nonlinear and complicated machine learning information. At the same time, the process of syndrome differentiation is considered to be nonlinear and complicated. Applying the DBN based multilabel on syndrome differentiation modeling is more appropriate. A DBN model is actually a multilayer perception neural network with one input layer, one output layer, and several middle hidden layers unit. The higher-level layer connects to its lower layer by a Restricted Boltzmann Machine (RBM) which uses the result of the lower layer to activate the next higher-level layer.

Our study applies the common deep learning method to deal with multilabel learning problem. The multilabel classification algorithms can be generally divided into two different categories [22]: problem transformation methods and algorithm adaptation methods. Some of them consider the correlations among the labels and some of them do not. For the convenient reason, we chose a simple method that ignores the correlations among labels to build the model, that is, binary relevance (BR) method. The deep learning model deep belief network (DBN) will be combining with binary relevance method, respectively, to deal with multilabel learning task. Binary relevance (BR) approach [23] directly transforms multilabel problem N binary classifiers Hn: X $\rightarrow$ {l, –l}, and each independent classifier deals with only one label. In this paper, DBN will take the place of the N binary classifiers. For example, multilabel learning model of CG syndrome diagnosis will be established for accomplishing six labels with six deep learning processes in this paper. The details of the process of multilabel learning methods based on deep learning are shown in Figure 1.

2.9.2. Multilabel Learning Framework Based on Deep Belief Nets. The following text will describe the learning process of deep belief network in detail. In this model, the original features were used directly in the multilabel learning of deep belief network. Figure 2 shows the approximate learning process of multilabel learning based on deep belief network. We put sample features and relevant parameters into the unsupervised RBM training model for training and then shift up the hidden layer to higher layer. This process is repeatedly executed until current hidden layer becomes the highest hidden layer, and so on; several unsupervised RBM models can be trained from visible layer to highest hidden layer and then obtain an initial set of the weighting parameters. Later on, the samples' original features are taken as the input layer of neural network, a label is chosen as output layer of neural network, and the middle hidden layer is taken as the hidden layer of neural network; a neural network model is trained from visible input layer to output layer. The weighting parameters in every layer can further be updated through the forward propagation and afterward propagation. After training, the category labels of training have been finished. Then, another label is chosen to be trained, until all labels are finished. The predicting process is the same as its training process, which means the labels are also predicted one by one. When each label is predicted, the neural network is used, which takes the samples' features as the input layer of the number of hidden layers and the number of hidden layer units that stayed the same as in training process and

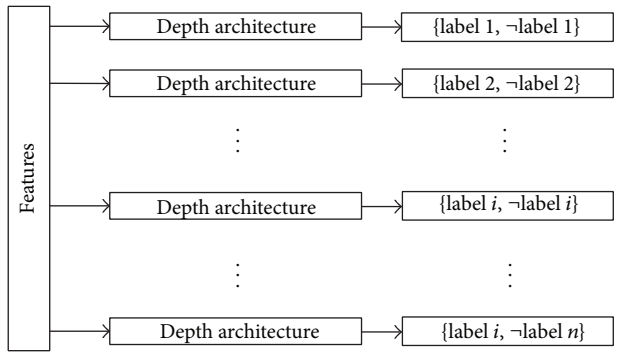


Figure 1: The process of deep learning multilabel learning.

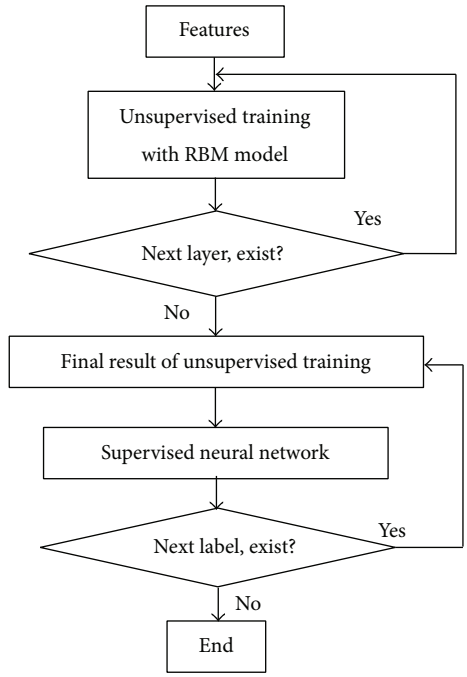


Figure 2: The process of deep belief nets multilabel learning.

executes the prediction through the forward propagation with the weighting parameters in every layer. We can map the corresponding higher expression of original features through trains corresponding model in the lower level to higher until the highest level expression results is presentation. The details of the process of multilabel learning methods based on deep belief nets are shown in Figure 2.

3. Experimental Design and Evaluation

The evaluation index of single label learning is usually accuracy, recalling rate and F1 measure value, but evaluation is different from single-label learning. The following five evaluation metrics specifically designed for multilabel learning are expressed as follows [24].

Average precision evaluates the average fraction of labels ranked above a particular label $y \in Y$, which actually are in Y. The performance is perfect when $\text{avgprecS}(f) = 1$; the bigger the value of $\text{avgprecS}(f)$ is, the better the performance is:

$$\begin{aligned}&\text{avgprecS}(f)\\&= \frac{1}{p}\sum_{i=1}^{p}\frac{1}{|Y_i|}\\&\quad\times \sum_{y\in Y_i}\frac{\left|\left\{y' \mid \text{rank}_f\left(x_i, y'\right) \le \text{rank}_f\left(x_i, y\right), y' \in Y_i\right\}\right|}{\text{rank}_f\left(x_i, y\right)}.\end{aligned} \tag{1}$$

Coverage evaluates how far on average we need to go down the list of labels to cover all the proper labels of the instance. It is loosely related to precision at the level of perfect recall. The smaller the value of $\text{coverageS}(f)$ is, the better the performance is:

$$\begin{aligned}\text{coverageS}(f) &= \frac{1}{p}\sum_{i=1}^{p}\max_{y\in Y_i}\text{rank}_f(x_i, y) - 1\\ \text{rank}_f(x_i, y) &= 1 - f(x_i, y).\end{aligned} \tag{2}$$

Ranking loss evaluates the average fraction of label pairs that are reversely ordered for the instance. The performance is perfect when $\text{rlossS}(f) = 0$; the smaller the value of $\text{rlossS}(f)$ is, the better the performance is:

$$\begin{aligned}\text{rlossS}(f) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{|Y_i|\left|\overline{Y_i}\right|}&\left|\left\{(y_1, y_2) \mid f(x_i, y_1)\right.\right.\\&\left.\left.\le f(x_i, y_2), (y_1, y_2) \in Y_i \times \overline{Y_i}\right\}\right|,\end{aligned} \tag{3}$$

where $\overline{Y}$ denotes the complementary set of Y in $y \cdot y = \{1, 2, \ldots, Q\}$ being the finite set of labels.

Hamming loss evaluates how many times instance-label pairs are misclassified, that is, a label not belonging to the instance is predicted or a label belonging to the instance is not predicted:

$$\text{hloss}_\Gamma(f) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\left|f(x_i)\,\Delta Y_i\right|, \tag{4}$$

where Δ denotes the symmetric difference between two sets.

One-error evaluates how many times the top-ranked label is not in the set of proper labels of the instance. The performance is perfect when $\text{one-error}_\Gamma(f) = 0$:

$$\text{one-error}_\Gamma(f) = \frac{1}{m}\sum_{i=1}^{m}\arg\max_{y\in Y} f(x_i, y)\, f(x_i, y) \notin Y_i. \tag{5}$$

For any predicted π, π equals 1 if π holds and 0 otherwise. Note that, for single-label classification problems, a one-error is identical to an ordinary classification error.

4. Results

We compared the model performance with different nodes' numbers of hidden layer and different multilabel learning algorithms. At the same time, we compared accuracy rates of 6 syndromes using DBN with different hidden layer. The results are shown in the following sections, respectively.

4.1. Comparison of Model with Different Nodes' Numbers. In order to illustrate the performance of deep learning framework on chronic gastritis inquiry data, a series of experiments have been carried out. Firstly, to confirm appropriate value of the deep architecture parameter, we set an experiment to confirm the scale of node in each hidden layer. Secondly, deep learning multilabel framework will be compared with other multilabel learning algorithm with either feature select or not. Finally, we compared the accuracy rates in 6 syndromes using different multilabel methods. In the experiments, five evaluation measures are employed: average precision, coverage, hamming loss, one-error, and ranking loss. Average precision expresses "the bigger the better" and the others express "the smaller the better." The symbol "↓" indicates "the smaller the better" while "↑" indicates "the bigger the better." Tenfold cross validation is employed on both data sets in order to predict reliably. A symbol "±" connects the means of classification result calculated ten times and their standard deviations. The best results are represented in bold.

Firstly, we experiment on an only one hidden layer DBN to find an appropriate node number value hid in the hidden layer; hid is chosen from [5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300]. For the process speed, the samples will be handled in batches, each batch containing 100 samples. The other parameters: the learning rate is set to 0.1, the biggest iterations are set to 100, the smooth is set to 0.5, and the damping factor is set to 2e-4. Table 1 shows the results of five evaluation measures of DBN with one layer. The best results are represented in bold.

As shown in Table 1, when hid = 80, the experimental results in five evaluation standards, as a whole, are the best, where average precision is 0.824, coverage is 0.158, one-error is 0.278, and ranking loss is 0.116 which achieves the best and hamming loss is 0.139 which is worse than the best result (0.135). But when the hid exceeds 30, the results of all the values of hid show little difference, which indicate that as long as there are enough hidden nodes and full learning, the experimental results cannot show too much difference.

Table 1: Results of model with different nodes' number (mean ± std.).

Hid	Average precision ↑	Coverage ↓	Hamming loss ↓	One-error ↓	Ranking loss ↓
5	0.791 ± 0.019	0.191 ± 0.013	0.154 ± 0.011	0.304 ± 0.029	0.151 ± 0.016
10	0.802 ± 0.025	0.175 ± 0.019	0.144 ± 0.014	0.301 ± 0.043	0.133 ± 0.021
20	0.802 ± 0.029	0.175 ± 0.022	0.145 ± 0.012	0.303 ± 0.041	0.133 ± 0.026
30	0.817 ± 0.027	0.164 ± 0.021	0.139 ± 0.012	0.283 ± 0.039	0.120 ± 0.023
40	0.812 ± 0.017	0.168 ± 0.016	0.138 ± 0.015	0.292 ± 0.029	0.125 ± 0.017
50	0.815 ± 0.019	0.166 ± 0.018	0.137 ± 0.010	0.289 ± 0.027	0.123 ± 0.018
60	0.816 ± 0.023	0.164 ± 0.020	**0.135 ± 0.009**	0.287 ± 0.039	0.121 ± 0.020
70	0.818 ± 0.018	0.164 ± 0.015	0.139 ± 0.009	0.283 ± 0.028	0.120 ± 0.016
80	**0.823 ± 0.018**	**0.158 ± 0.015**	0.139 ± 0.014	**0.278 ± 0.028**	**0.116 ± 0.016**
90	0.821 ± 0.020	0.160 ± 0.016	0.137 ± 0.012	0.278 ± 0.034	0.116 ± 0.017
100	0.817 ± 0.023	0.166 ± 0.017	0.136 ± 0.006	0.281 ± 0.045	0.123 ± 0.018
200	0.818 ± 0.021	0.164 ± 0.016	0.139 ± 0.009	0.282 ± 0.035	0.120 ± 0.018
300	0.815 ± 0.023	0.164 ± 0.019	0.141 ± 0.011	0.287 ± 0.039	0.122 ± 0.022

Table 2: Results of model with different multilabel learning (mean ± std.).

	Average precision ↑	Coverage ↓	Hamming loss ↓	One-error ↓	Ranking loss ↓
ML-KNN	0.754 ± 0.031	0.206 ± 0.017	0.166 ± 0.017	0.380 ± 0.059	0.173 ± 0.020
BSVM	0.794 ± 0.037	0.180 ± 0.023	0.166 ± 0.022	0.320 ± 0.065	0.138 ± 0.029
Rank-SVM	0.682 ± 0.018	0.255 ± 0.029	0.232 ± 0.014	0.497 ± 0.025	0.227 ± 0.019
BP-MLL	0.514 ± 0.028	0.395 ± 0.036	0.313 ± 0.010	0.750 ± 0.048	0.390 ± 0.044
CLR	0.784 ± 0.024	0.185 ± 0.023	0.172 ± 0.016	0.343 ± 0.045	0.143 ± 0.021
ECC	0.793 ± 0.021	0.193 ± 0.018	0.150 ± 0.013	0.277 ± 0.038	0.193 ± 0.023
REKAL	0.781 ± 0.026	0.209 ± 0.024	0.152 ± 0.012	0.331 ± 0.036	0.167 ± 0.026
LEAD	0.803 ± 0.019	0.174 ± 0.016	0.151 ± 0.014	0.304 ± 0.034	0.133 ± 0.015
DBN	**0.823 ± 0.018**	**0.158 ± 0.015**	**0.139 ± 0.014**	**0.278 ± 0.028**	**0.116 ± 0.016**

4.2. Comparison of Model with Different Multilabel Learning. We selected the best result for one hidden layer and its optimal nodes' number DBN model and compared the five evaluation parameters obtained using ML-KNN, Ensembles of Classifier Chains (ECC), BSVM, BP-MLL, Rank-SVM, CLR, REkAL, and LEAD algorithms. For BSVM, we chose the kernel function as linear; for ML-KNN, we set the neighbor number to 10 and chose the Euler distance to measure the sample distance; for Rank-SVM, we set the maximum iterations as 50 and chose the linear kernel function; for BP-MLL, we set the number of hidden neurons layer as 20% of the number of features and set the number of neural node as 100; for CLR and ECC, we set the size of integration as 10 and set the sample proportion as 67%; for REKAL, we set the size of subset as 3 and chose LP as the multiclass algorithm. The results are shown in Table 2.

As shown in Table 2, the result of DBN was significantly better than that of other algorithms. Although DBN is actually a neural network model as well as BP-MLL, the result shows that DBN is obviously superior to BP-MLL with 31.5% higher in average precision measure. It indicates that DBN model is better to deal with TCM CG inquiry data than BP-MLL.

4.3. The Comparison of Accuracy Rates of 6 Syndromes. In order to have a further discussion on the effect of the depth of deep architecture, the DBN method was compared with different numbers of layers of accuracy rates for various syndromes. The recognition accuracies of the six common syndromes of CG are shown in Table 3.

As shown in Table 3, there are four syndromes with the DBN algorithm that achieved the highest accuracy rate, that is, the pattern of damp heat accumulation in the spleen-stomach, dampness obstructing the spleen-stomach, spleen-stomach qi deficiency, and liver stagnation achieved at 90.1%, 81.2%, 75.3%, and 83.9%, respectively, followed by BSVM, Rank-SVM, and ML-kNN, whose performances are almost the same. BP-MLL performed second best on the pattern liver stagnation with 83.1% but performed the worst with the other three syndromes. For the pattern of spleen-stomach cold deficiency, the accuracy of DBN has very close performance with ML-kNN and BP-MLL at 96.6%, followed by BSVM at 94.3% and Rank-SVM only at about 80%. For the pattern of stagnated heat in the liver-stomach, BP-MLL algorithm achieved the highest accuracy rate at 91.0%, which is only 0.2% and 0.5% higher than ML-kNN and DBN, and Rank-SVM has the lowest accuracy of 79.9%.

Table 3: Results of recognition accuracy (%) for six common syndromes (mean ± std.).

Syndromes (patterns)	ML-kNN	BSVM	BP-MLL	Rank-SVM	DBN
Damp-heat accumulating in the spleen-stomach	86.9 ± 3.6	88.4 ± 2.5	24.7 ± 3.5	88.0 ± 2.8	**90.1 ± 0.024**
Dampness obstructing the spleen-stomach	73.7 ± 4.4	80.0 ± 3.5	68.3 ± 5.2	76.2 ± 4.4	**81.2 ± 0.019**
Spleen-stomach qi deficiency	68.9 ± 6.5	71.2 ± 2.3	53.8 ± 3.9	67.9 ± 6.8	**75.3 ± 0.044**
Spleen-stomach deficiency cold	96.6 ± 1.7	94.3 ± 2.7	96.6 ± 1.7	79.3 ± 3.6	**96.6 ± 0.021**
Liver stagnation	82.7 ± 5.6	82.6 ± 4.9	83.1 ± 5.4	81.0 ± 4.7	**83.9 ± 0.028**
Stagnated heat in liver-stomach	90.8 ± 2.3	90.1 ± 3.0	**91.0 ± 2.2**	79.9 ± 4.8	90.5 ± 0.030

From the comparison of experimental results, DBN method obtains the satisfied comprehensive performance in the multilabel learning for syndrome classification on CG data.

5. Discussion

A syndrome is a unique TCM concept. It is an abstractive conception of a variety of symptoms and signs. It is a pathological summarization of a certain stage of a disease and it covers disease location, etiology, and the struggle between the body's resistance and pathogenic factors. Different syndromes have different clinical manifestations. Symptoms, which are the external manifestations of a disease and a syndrome, refer to subjective abnormalities and the abnormal signs of patients elicited by doctors using the four diagnostic methods. The etiology, location, nature, the struggle between the body's resistance and pathogenic factors, and the condition at a certain stage of the disease process are highly summarized using syndrome differentiation. Syndrome differentiation involves three steps: (a) determining symptoms and signs through inspection, auscultation, inquiry, and palpation; (b) making an overall analysis of the information; and (c) making a diagnostic conclusion. All these steps are based on TCM theory.

Figure 3 shows the TCM syndrome diagnosis of the hierarchical structure diagram. $X1$, $X2, \ldots$ and $X5$ are directly observed and we call them symptoms (signs) variables. In this study, it denotes the symptoms and signs of CG. $H1$ and $H2$ are the syndrome factors, which are the preliminary summarize of syndrome and the foundation of the syndrome diagnosis. $S1$ and $S2$ are the results of the syndrome diagnosis. H and S are indirectly measured through their manifestations and we call them latent variables, which represent the different hierarchical syndromes of CG.

In clinical syndrome diagnosis, there is certain complex and nonlinear relationship between symptoms and each other and between symptoms and syndromes. The occurrence of a symptom may be accompanied by other symptoms together.

Multiple symptoms concurrent phenomena can be understood: some abstract syndromes factors can represent the collection of several concurrent symptoms. This syndrome diagnosis hierarchy is consistent with the human brain cognitive.

In the traditional information methods of TCM, most research is not considered from a cognitive point of view and from the TCM nonlinear, complex, and multilayered aspects. Simple feature selection is likely to cause the incomplete expression of feature subset and feature conversions easily lead to uncertainty.

Deep learning can use high-dimensional abstract features to express some of the original low-dimensional features without the need for the person to participate in the selection of features. Therefore, deep learning is more consistent with human brain's cognitive thinking. This idea is consistent with the diagnosis ideas of TCM. It is a good method to find the relationship between the symptoms and each other and between the symptoms and syndromes. This idea is consistent with the diagnosis ideas of TCM.

This paper introduces the basic concept of the deep learning method, using the DBN to establish a multilabel learning algorithm and apply this established algorithm on the TCM syndrome differentiation of CG. Firstly, a simple RBM model of different numbers of hidden layer node was tried to find out the appropriate layer on CG. The best result is when the scale of nodes is 80. The average precision, coverage, one-error, and ranking loss were the best; they were 0.823, 0.158, 0.278, and 0.116. Only the hamming loss gets an old stuff value of 0.139 and then the multilabel learning based on DBN was compared with other popular multilabel learning algorithms on CG data in both multilabel learning task and single label learning task. In the multilabel learning task, the multilabel learning based on DBN achieves the best in all the five evaluation measures, especially the average precision (82.3%) being 2% higher than LEAD which is the second best performance with 80.3%. In the single label-learning task, each syndrome was treated as single label classification by various algorithms: ML-kNN, BSVM, BP-MLL, Rank-SVM, and DBN. DBN achieves better than other algorithms with five syndromes, that is, the pattern of damp heat accumulation in the spleen-stomach with the accuracy 90.1%, dampness obstructing the spleen-stomach with the accuracy 81.2%, spleen-stomach qi deficiency with the accuracy 75.3%, spleen-stomach deficiency cold with the accuracy 96.6, and liver stagnation with the accuracy 83.9%. Only the pattern of Stagnated heat in liver-stomach performed third best with the accuracy 90.5% less than BP-MLL with the accuracy 91% and NL-kNN with the accuracy 90.8%. The perfect result demonstrates that the multilabel learning based on DBN method is superior to other multilabel learning methods.

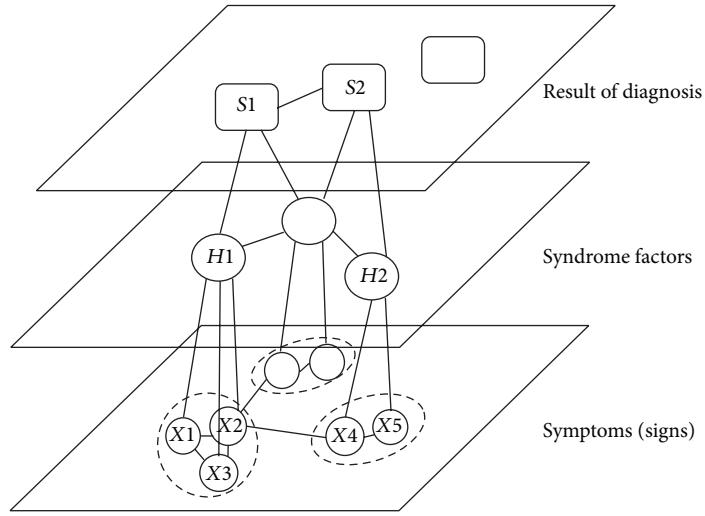


Figure 3: The TCM syndrome diagnosis of the hierarchical structure diagram.

6. Conclusions

To fully understand the characteristics of multilabel data of TCM in syndrome diagnosis, a deep learning model DBN is used to establish a multilabel learning framework and apply to TCM syndrome differentiation modeling for CG dates which are regarded as nonlinear and complicated. DBN based multilabel learning can perform outstanding for its capacity of high level information expression.

An experiment is set to find appropriate scale of nodes in one hidden layer DBN architecture with CG data. The result indicates that with only enough scale of nodes, but not too much, the DBN architecture can improve the performance of deep learning.

Moreover, DBN based multilabel learning was compared with other multilabel algorithms. Compared results indicated that DBN dealing with multilabel task performs better than other algorithms. The results are measured by five evaluation indexes; that is, average precision, coverage, hamming loss, one-error, and ranking loss. And all the indexes of DBN based multilabel learning achieve the best.

The study has shown that DBN based on multilabel learning is effective to deal with the task of modeling of TCM dates. In addition, the study will serve as a reference for establishing diagnostic criteria and a diagnostic model for CG and a better guide for clinical practice.

Conflict of Interests

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 81270050, 30901897, and 81173199).

References

[1] H. L. Gao, Y. L. Wu, J. Chen et al., "The complex system based on entropy method to partition the syndrome diagnosis standard research of application," *China Journal of Traditional Chinese Medicine and Pharmacy*, vol. 23, pp. 307–309, 2008.

[2] Y. Su, L. Wang, and H. Zhang, "Posthepatitic cirrhosis of TCM syndrome and disease information similarity analysis," *Chinese Journal of Integrated Traditional and Western Medicine*, vol. 29, pp. 398–402, 2009.

[3] G.-P. Liu, J. Yan, and Y.-Q. Wang, "Application of multilabel learning using the relevant feature for each label in chronic gastritis syndrome diagnosis," *Evidence-Based Complementary and Alternative Medicine*, vol. 2012, Article ID 135387, 9 pages, 2012.

[4] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning," *Journal of Machine Learning Research*, vol. 11, no. 3, pp. 625–660, 2010.

[5] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," in *Large-Scale Kernel Machines*, L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, Eds., pp. 321–335, MIT Press, Cambridge, Mass, USA, 2007.

[6] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[7] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds., pp. 153–160, MIT Press, 2007.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[9] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096–1103, ACM Press, Helsinki, Finland, July 2008.

[10] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: learning useful

representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. 12, pp. 3371–3408, 2010.

[11] Y. Dong and D. Li, "Deep convex net: a scalable architecture for speech pattern classification," in *Proceedings of the 12th Annual Conference of International Speech Communication Association*, pp. 2285–2288, 2011.

[12] H. Poon and P. Domingos, "Sum-product networks: a new deep architecture," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 689–690, November 2011.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012.

[14] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1915–1929, 2013.

[15] K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, and Y. LeCun, "Learning convolutional feature hierarchies for visual recognition," in *Advances in Neural Information Processing Systems 23*, 2010.

[16] G. Dahl, D. Yu, L. Deng, and A. Acero, "Large vocabulary continuous speech recognition with context-dependent DBN-HMMS," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4688–4691, 2011.

[17] A. Mohamed, T. N. Sainath, G. Dahl, B. Ramabhadran, G. Hinton, and M. Picheny, "Deep belief networks using discriminative features for phone recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 5060–5063, 2011.

[18] Chinese Medical Association Digestive Diseases Branch, "Consensus of national seminar on chronic gastritis," *Chinese Journal of Digestive Endoscopy*, vol. 24, pp. 58–63, 2007.

[19] X. Y. Zheng, *Chinese Herbal Medicine Clinical Research Guiding Principles (for Trial Implementation)*, China Medical Science and Technology Press, Beijing, China, 2004.

[20] State Bureau of Technical Supervision, *National Standards of Peoples Republic of China. Syndrome Part of TCM Clinical Diagnosis and Treatment Terminology*, Standards Press of China, Beijing, China, 1997.

[21] G. P. Liu, Y. Q. Wang, Y. Dong et al., "Development and evaluation of scale for heart system inquiry of TCM," *Journal of Chinese Integrative Medicine*, vol. 7, no. 1, pp. 1222–1225, 2009.

[22] G. Tsoumakas and I. Katakis, "Multi-label classification: an overview," *International Journal of Data Warehousing and Mining*, vol. 3, pp. 1–13, 2007.

[23] G. Nasierding, G. Tsoumakas, and A. Z. Kouzani, "Clustering based multi-label classification for image annotation and retrieval," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC '09)*, pp. 4627–4632, San Antonio, Tex, USA, October 2009.

[24] R. E. Schapire and Y. Singer, "Boostexter: a boosting-based system for text categorization," *Machine Learning*, vol. 39, no. 2-3, pp. 135–168, 2000.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 536308, 8 pages
http://dx.doi.org/10.1155/2014/536308

Research Article

Cirrhosis Classification Based on Texture Classification of Random Features

Hui Liu,[1] Ying Shao,[1] Dongmei Guo,[2] Yuanjie Zheng,[3] Zuowei Zhao,[4] and Tianshuang Qiu[1]

[1] *Department of Biomedical Engineer, Dalian University of Technology, Dalian 116024, China*
[2] *Department of Radiology, Second Affiliated Hospital, Dalian Medical University, Dalian 116027, China*
[3] *Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA*
[4] *Second Affiliated Hospital, Dalian Medical University, Dalian 116027, China*

Correspondence should be addressed to Hui Liu; liuhui@dlut.edu.cn

Received 22 November 2013; Accepted 14 January 2014; Published 24 February 2014

Academic Editor: Rong Chen

Accurate staging of hepatic cirrhosis is important in investigating the cause and slowing down the effects of cirrhosis. Computer-aided diagnosis (CAD) can provide doctors with an alternative second opinion and assist them to make a specific treatment with accurate cirrhosis stage. MRI has many advantages, including high resolution for soft tissue, no radiation, and multiparameters imaging modalities. So in this paper, multisequences MRIs, including T1-weighted, T2-weighted, arterial, portal venous, and equilibrium phase, are applied. However, CAD does not meet the clinical needs of cirrhosis and few researchers are concerned with it at present. Cirrhosis is characterized by the presence of widespread fibrosis and regenerative nodules in the hepatic, leading to different texture patterns of different stages. So, extracting texture feature is the primary task. Compared with typical gray level cooccurrence matrix (GLCM) features, texture classification from random features provides an effective way, and we adopt it and propose CCTCRF for triple classification (normal, early, and middle and advanced stage). CCTCRF does not need strong assumptions except the sparse character of image, contains sufficient texture information, includes concise and effective process, and makes case decision with high accuracy. Experimental results also illustrate the satisfying performance and they are also compared with typical NN with GLCM.

1. Introduction

Liver cirrhosis is one of the leading causes of death by disease [1]. Making a definite diagnosis and staging of the cirrhosis is crucial which will help doctor to offer the timely and appropriate therapeutic method. In recent decades, CAD has drawn an increasing attention for its convenient and noninvasive diagnosis procedure, meaningful diagnosis result. As it is well known, the more cirrhosis categories, the more valuable the classification result is, for it can provide a constructive assistance to doctors and facilitate the process for doctors to produce a specific treatment for patient. Therefore, the study of cirrhosis classification is developing form two categories classification of cirrhosis, such as methods provided by Chen et al. [2], Lee et al. [1, 3], Hui et al. [4], and Li et al. [5], to precisely class three stages (normal stage, early stage, and middle and advanced stage). Thus, a CAD system which can generate more precise stages is an inevitable tendency in the future.

The early CAD systems of cirrhosis mostly apply CT images, such as the methods provided by Li et al. [5] and Chen et al. [2, 6]. Compared with CT, MRI provides many advantages including high resolution of soft tissue, no radiation, and multiparameters imaging modalities and it is becoming an effective modality in assessing cirrhosis. However, few researchers are concerned with the cirrhosis CAD system with MRI by far. So in this paper, multisequences MRI, including T1-weighted, T2-weighted, arterial phase, portal venous phase, and equilibrium phase, are applied for cirrhosis classification which divide cases into three stages (normal stage, early stage, and middle and advanced stage).

Cirrhosis CAD usually consists of two steps: the first step involves feature extraction and effective feature selection, and the second step is to train classifier based on the selected

features and identify cirrhosis [4]. Feature extraction and selection occupy an important position that directly affects the performance of classifier. Cirrhosis is characterized by the presence of widespread fibrosis and regenerative nodules in the hepatic. The fibrosis and nodules formation causes distortion of the normal hepatic architecture, resulting in characteristic texture patterns [1]. Therefore, extracting texture feature is the primary task of CAD. Jiang et al.'s cirrhosis classification method [7] achieves 56% accuracy with texture features directly extracted from region of interests (ROI), such as the ROI mean gray value and ROI standard deviation. Because simple texture features are unsatisfactory, Jiang et al. add morphological features to improve classification performance. Actually, effective texture feature that can fully describe cirrhosis feature is a key problem of CAD. GLCM based feature is a classical method to extract texture features for cirrhosis classification [4, 8–11]. However, GLCM has many weaknesses. For example, GLCM is a statistical texture description for two specific grayscales with specific distance and direction which will cause strong assumptions about the texture being studied in practice. And it is difficult to contain all texture information according to the GLCM texture features with such strong assumptions. It should also be noted that complex process of computing GLCM and its 14 typical texture parameters will take long time especially for large image. And with GLCM based features it will need a necessary procedure to select effective features with a heavily and complicated test to improve the classifier's performance. Virmani et al. [9] uses all the 14 GLCM texture features for two stages classification (normal and cirrhosis). Its classification accuracy is 95.86%, achieved by using an NN classifier with stratified 10-fold cross validation method. For two stages classification, GLCM texture features are fine even though with the price of time and complexity. However, the classification with the GLCM texture features is far from satisfactory for three stages (normal, early, and middle and advanced); the accuracy is about 89.52% for normal, 70.98% for early, and 74.27% for middle and advanced stages that is verified in this paper, which causes most of traditional classifiers far from application.

Recently, texture classification from random features [12] proposed by Liu et al. provides an effective way to extract texture features. Firstly, it does not need strong assumptions about the texture images, except the sparse character of the texture image. Secondly, it almost does not lose information when extracting texture feature. Because no information is lost in the three stages which are extracting local image from image, extracting patch from local image, and stretching patch into patch vector, the process of compressing patch vector retains salient information when image is sparse character. Thus, the texture feature, which is compressed patch vector, almost does not lose information. Thirdly, Liu et al.'s method does not need feature selection, reducing method complexity. Finally, it costs little time because of the simple computation procedure which includes extracting local image from image, extracting patch from local image, stretching patch into patch vector, and compressing patch vector that only needs matrix multiplication. And omitting feature selection can also save time.

Considering the above reasons, we propose CCTCRF method and apply Liu et al.'s method to extract texture features. Liver organization of hepatic MRIs studied in this paper is a kind of image with texture character, and it also meets the sparse character requirement as an assumption in Liu et al.'s method. As it is well known, the key factor to obtain excellent performance in cirrhosis classification is preserving salient cirrhosis texture information, while Liu et al.'s method exactly meets the requirement according to the process of extracting texture feature. Furthermore, due to the weak assumption of Liu et al.'s method, we can apply it into five sequences MRIs all have sparse character without considering specific parameters for specific sequence, and the experiment result also demonstrates the effectiveness of the five sequences MRIs texture features. In addition, as the number of multisequences MRI used in our research is large, the little time costing and complexity reducing in feature extraction and selection is also a reason to choose Liu et al.'s method. Thus, we apply Liu et al.'s method as texture extraction method in our research. Except for the benefits of texture feature extraction based on Liu et al.'s method, CCTCRF also has a remarkable performance on case classification, compared with typical GLCM texture based CAD system.

In this paper, we propose the CCTCRF method with five sequences MRI, involving T1-weighted, T2-weighted, hepatic arterial phase, portal venous phase, and equilibrium phase images to class the patient samples into normal stage, early stage, and middle and advanced stage. In order to illustrate the effectiveness of CCTCRF method, we also do a contrast experiment with GLCM texture features and typical neural network (NN).

This paper is organized as follows. Section 2 presents the materials used in our study and describes the NN classification method with GLCM texture features to be compared and the CCTCRF theory applied in this paper. Section 3 describes the experiments performed to prove the effectiveness of CCTCRF and the NN classifier with GLCM texture features experiment as comparison study. The conclusion is in Section 4.

2. Materials and Methods

2.1. Materials. The MR images are collected from the Second Affiliated Hospital of Dalian Medical University between September 2012 and September 2013. The database containing 55 patient cases is used in this study. Of the 55 patient cases, 26 cases are at normal stage, 13 are early cirrhosis stage, and 16 are the middle and advanced cirrhosis, shown in Table 1.

Most of the 55 patient cases have five sequences MRI, T1-weighted, T2-weighted, arterial phase, portal venous phase, and equilibrium phase, except a few of them who are short of portal venous phase or equilibrium phase images. The cases are scanned by a 3.0-T superconductivity MR scanner (Signa, Siemens, Germany) or 1.5-T MR scanner (Signa, GE, USA); the scan order is T1-wighted, T2-wighted, and three kinds of dynamic enhancement scan which are arterial phase, portal venous phase, and equilibrium phase, respectively,

Table 1: The number of collected cases.

	Normal	Early	Middle and advanced
T1-weighed	26	13	16
T2-weighed	26	13	16
Arterial	26	13	16
Portal venous	26	13	15
Equilibrium	20	13	14

after injecting Gd-DTPA 25 s, 65 s, and 120 s. The parameters of T1-weighted, arterial phase, portal venous phase, and equilibrium phase from 3.0-T superconductivity MR scanner are TR = 3.9 ms and TE = 1.4 ms; T2-weighted parameters are TE = 105 ms. The parameters of T1-weighted, arterial phase, portal venous phase, and equilibrium phase from 1.5-T MR scanner are TR = 175.0 ms and TE = 4.2 ms; T2-weighted parameters are TR = 7058.8 ms and TE = 89.5 ms.

2.2. Methods

2.2.1. Texture Classification from Random Features. The method of texture classification from random features [12] is inspired by theories of sparse representation and compressed sensing. And it presents a simple and powerful approach for texture image classification based on random projection. It includes four parts, patch extraction, compressed texton dictionary, histogram of textons learning, and the classification. At the feature extraction step, texture feature, compressed patch vector, is extracted with patch extraction, patch vector generation, and patch vector compressing. The texture features are embedded into a bag-of-words model to perform texture classification. The extraction of compressed patch vector with random projection is simple and takes full advantages of the sparse nature of texture images. Liu et al.'s method outperforms traditional feature extraction methods which involve careful design and complex steps. Compared with the four state-of-the-art texture classification involving Patch, Patch-MRF, MR8, and LBP on the databases of CUReT, Brodatz, and MSRC, Liu et al.'s method leads to an important improvement in classification accuracy and reductions in feature dimensionality [12].

2.2.2. NN Classification with GLCM Texture Features. NN classifier [13] is a classic method in pattern classification and recognition, widely used in speech recognition [14, 15], image recognition [16–18], handwritten character recognition [19, 20], radar target recognition [21], and sonar target recognition [22, 23]. NN simulates the progress of neure that obtains knowledge from outside world and learns by itself to achieve the class function. GLCM is a classic method to extract texture feature, provided by Haralick [24]. It describes the probability that a couple of pixels, at θ direction and at a distance of d pixels, appear i gray level and j gray level. Then, it deduces 14 texture features including angular second moment, contrast, correlation, variance, inverse differences moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measures of correlation, Maximum correlation coefficient 1, and Maximum correlation coefficient 2 [25] according to GLCM [26]. Thus, NN classifier with GLCM texture features is widely used in classification.

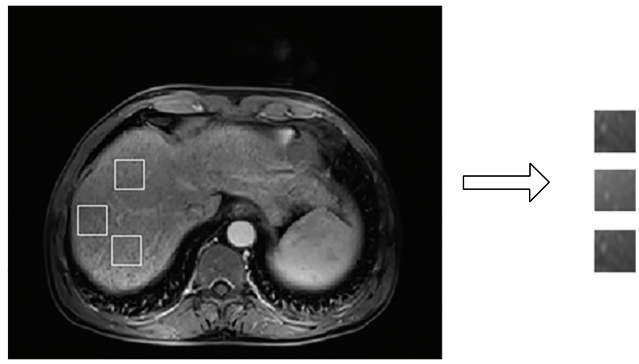


Figure 1: Extracting ROI.

2.2.3. Cirrhosis Classification Based on Texture Classification from Random Features (CCTCRF). We propose a new method CCTCRF which distinguishes hepatic cirrhosis patients into normal stage, early stage, and middle and advanced stage with five sequences MRI, which are T1-weighted, T2-weighted, arterial phase, portal venous phase, and equilibrium phase. Liu et al.'s method is applied to extract texture features and class intermediate samples. In addition, a final case decision making is produced with the intermediate samples classification results. CCTCRF includes four parts: ROI segment, texture feature extraction, ROI classification system, and case decision.

ROI segment step is shown in Figure 1. The images used for segment are manually selected according to the requirements that the images need to contain a clear and relatively whole liver. ROIs with size of $n \times n$ are extracted from five sequences MRIs, and considering the extraction principle of the diffuse distribution of liver, large blood vessels within the liver are excluded [8]. In this paper, the size of ROI is 30×30 or 60×60, and it depends on the range in liver that can be extracted. And we extract many ROIs from the same image when it has enough region accord with the extraction principle.

Figures 2(a)–2(c) are the ROI examples of normal stage, early stage, and middle and advanced stage, respectively, and the image order is T1-weighted, T2-weighted, arterial phase, portal venous phase, and equilibrium phase for each stage. According to clinical information, hepatic MRI conveys the different texture information according to different cirrhosis stage: normal hepatic tissue appears as delicate texture and uniform medium gray level, while early cirrhosis and middle and advanced cirrhosis appear as coarse particles or diffuse small nodular with saltatory gray with different extent [4].

ROIs are manually segmented by an experienced radiologist; the number distribution of selected ROI is illustrated in Table 2.

At the feature extraction step, we extract texture feature from ROI. Processing a ROI of n^2 pixels by extracting square

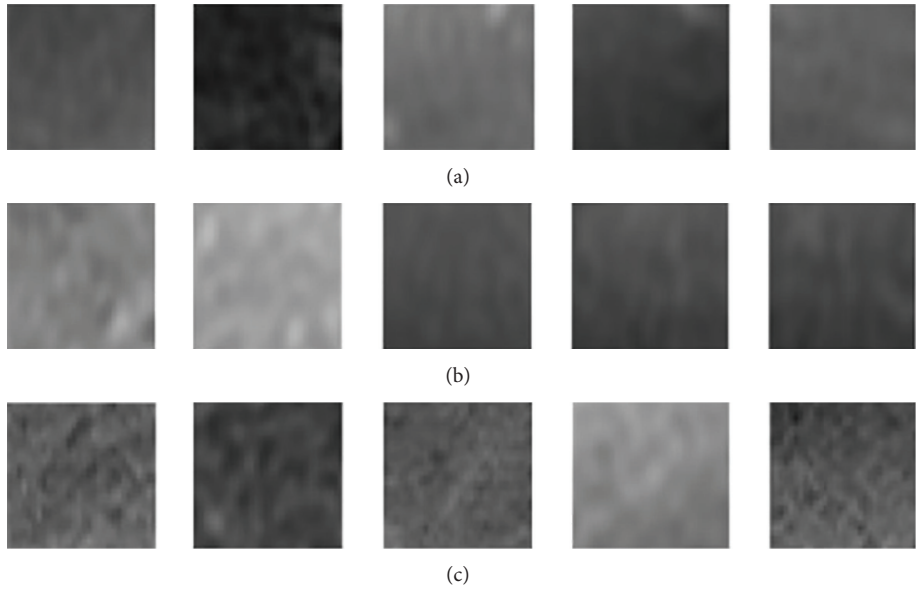


Figure 2: Normal, early, and middle and advanced stage ROIs form T1-weighted, T2-weighted, arterial phase, portal venous phase, and equilibrium phase.

Table 2: ROIs distribution.

	Normal	Early	Middle and advanced
T1-weighed	142	93	235
T2-weighed	75	64	91
Arterial	127	87	189
Portal venous	119	80	175
Equilibrium	94	84	153

patch $\{P_{i,j}\}$ of m pixels and $\sqrt{m}$ is much less than n, from every pixel located in $P_{i,j}$. And in order to get the square, those pixels on the ROI boundary are removed. So the range of i, j is $(\sqrt{m}/2, n-\sqrt{m}/2)$. Then, stretch each patch $\{P_{i,j}\}$ into a vector of size m and labelled as p_l. Next, compress patch vector p_l with Φ, which is sampled from independent zero-mean, unit-variance normal distribution, to obtain compressed patch vector $x_l = \Phi p_l$ as the texture feature. Consider

$$\mathbf{X} = \{x_l = \Phi p_l \mid p_l \in \mathbf{P}\}, \tag{1}$$

where $\mathbf{P}$ is the set of the patch generated from a ROI.

Before processing x_l as the input vector of classification, it needs to be normalized and there are two kind of normalization methods [12], formulated as (2) and (3) or no normalization as follows.

(1) Weber's law

$$x \longleftarrow x \times \left[\frac{\log\left(1 + \|x\|_2/0.03\right)}{\|x\|_2}\right]. \tag{2}$$

(2) Unit norm

$$x \longleftarrow \frac{x}{\|x\|_2}. \tag{3}$$

Any kind of normalization can be selected based on the classification performance after the next process.

Patch vector, which comes from ROI, is the gray level assemble in practice. Thus, it reflects the different texture information of tissue and organ with T1 value and T2 value or proton density as well as gray level for MRI. It also reflects local information of a ROI and we can control the local range by adjusting m to achieve the ideal scope of cirrhosis texture. Actually, to compress patch vector is a process of dimensionality reduction and is important in handling high dimensional data since it mitigates the curse of dimensionality and other undesired properties of high dimensional spaces [12], and the compressed patch vector does not need to feature selection which has heavy and complicated algorithm or analysis. In this paper, random projection which refers to the technique of projecting a set of points from a high dimensional space to a randomly chosen low dimensional subspace is the way of compressing patch. Random projection's effectiveness in information-preserving and dimensionality-reduction power is evidenced in the emerging theory of compressed sensing (CS) [27–29], which states that, for sparse and compressible signals, a small number of nonadaptive linear measurements in the form of random projections can capture most of the salient information in the signal and allow for perfect reconstruction of the signal. And it has been widely used in information retrieval, face recognition [30], and machine learning [31, 32]. Thus, using the compressed patch vector x as the following input is reasonable and effective.

Our ROI classification system is illustrated in Figures 3-4, consisting of four steps. Suppose we have C cirrhosis stages and each stage has S samples.

(1) *Compressed Texton Dictionary Learning Step.* Compressed texton dictionary W is learned directly in the compressed

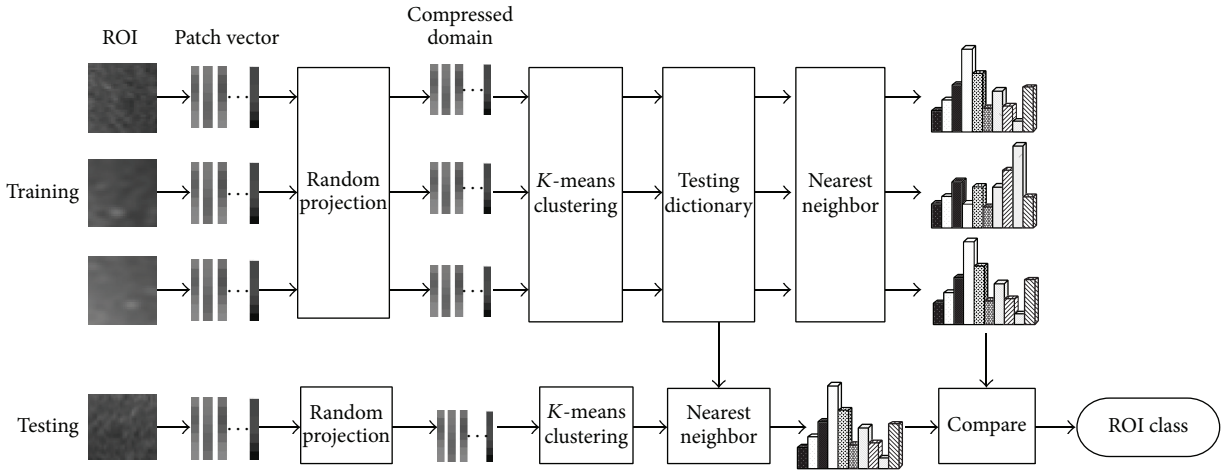


Figure 3: ROI training and testing.

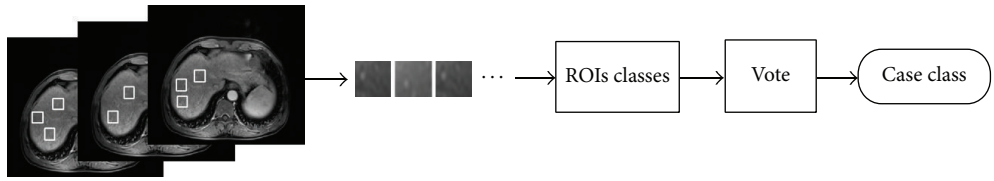


Figure 4: Case decision making step.

domain **X**. It learns K textons with K-means for each stage compressed patch vectors. Next, the compressed texton dictionary W is formulated by concatenating the K textons of each texture stage and the size of dictionary is CK (i.e. $W = CK$).

(2) *Histogram of Textons Learning, Shown in Figure 3 (Top).* By labeling each of extracted patches x with the closest texton in W, $h_{c,s}$ which is a histogram of compressed textons belonging to one of the samples from stage c is learned for each training sample. Each stage is represented by a set of models $H_c = \{h_{c,s}\}_s$ [12].

(3) *The Classification Step Is Shown in Figure 3 (Bottom).* Computing h_{new} which is a histogram from one of the samples belongs to test data. Using nearest neighbor classifier, h_{new} is classified and the distance between two histograms is measured using the χ^2 statistic:

$$\chi^2(h_1, h_2) = \frac{1}{2}\sum_{k=1}^{CK}\frac{[h_1(k) - h_2(k)]^2}{h_1(k) + h_2(k)}. \quad (4)$$

Then, h_{new} belongs to stage i when the distance between h_{new} and h_i is the nearest compared with other histograms. Thus, the stage of corresponding ROI which h_{new} derives from is determined; it is equal to h_{new} stage.

Case decision making stage is as follows. As we have shown in Figure 4, a case of one sequence images can extract many ROIs that the number of ROI is determined by the image region scope that meets extraction principle. In other words, M sequence images of case A have many classification results that belong to corresponding ROI. When all ROIs belong to the same case and same sequence is classed to the same stage c, surely the case of M sequence stage is c. However, when the ROIs classification results have more than one stage, case A of M sequence is determined by the vote principle that the minority stage subordinate to the majority stage.

3. Experiment Results

MATLAB R2010a is used to implement the CCTCRF experiment and the NN classifier with GLCM texture features experiment.

3.1. The CCTCRF Experiment Result. We use five sequences MRI in the experiment and the number of patients is shown in Table 1. Because images come from five different MR sequences, we need to do five experiments to class cases with the same sequence images in order to obtain the individual performance of every kind of sequence.

Take T1-weighted images for classification, for instance. There are 26 normal cases, 13 early cirrhosis cases, and 16 middle and advanced cirrhosis cases of T1-weighted. The train cases of normal stage, early stage, and middle and advanced stage are 13, 7, and 8; test cases are 13, 6, and 8. Extracting, respectively, 142, 93, and 235 ROIs from three kinds of cases according to the principle of extracting ROI and train cases of normal, early, and middle and advanced stage which are 54, 42, and 180 ROIs, respectively, the remaining ROIs belong to test cases. Each ROI is processed by extracting patches $\{P_{i,j}\}$

Table 3: Distribution of cases numbers.

	Normal		Early		Middle and advanced	
	Test	Train	Test	Train	Test	Train
T1-weighted	1–13	14–26	1–7	8–13	1–8	9–16
T2-weighted	1–13	14–26	1–7	8–13	1–8	9–16
Arterial	1–13	14–26	1–7	8–13	1–8	9–16
Portal venous	1–13	14–26	1–7	8–13	1–6, 8, 9	10–16
Equilibrium	1–13	14–17, 19, 20, 25	1–7	8–13	1–5, 8, 9	10–16

Table 4: CCTCRF experiment result (%).

	Normal		Early		Middle and advanced	
	ROI	Case	ROI	Case	ROI	Case
T1-weighted	100	100	100	100	94.45	100
T2-weighted	100	100	97.30	100	91.89	100
Arterial phase	100	100	100	100	100	100
Venous	100	100	100	100	86.96	100
Equilibrium	100	100	100	100	98.48	100

of size $\sqrt{25} \times \sqrt{25}$ around each pixel position (i, j) except those pixels on the ROI boundary. Stretching $\{P_{i,j}\}$ into p_l, that is a vector of 25 pixels, we compressed p_l into a 15-dimension compressed patch vector using independent zero-mean, unit-variance normal distribution. Choose unit norm as normalization method on account of unit norm which has the best evaluation among the three normalization ways for T1-weighted. Then, the texture feature, compressed path vector with normalization, is achieved. Next, input the texture features into the ROI classification system, the distribution of cases number is shown in Table 3. After obtaining the ROIs stages, we carry out the case decision making step to achieve the test cases stages of T1-weighted.

What is said above is for T1-weighted images. Except the number of each stage case and parameters, T1-weighted, T2-weighted, arterial phase, portal venous phase, and equilibrium phase have the same process. The parameters in texture feature extraction stage include the size of ROI and patch and the dimension of compressed patch vector. In compressed texton dictionary learning step of ROI classification system, parameters include textons with K-means for each stage and which way to normalize compressed patch vector. We adjust these parameters according to the special sequence in order to achieve the desired performance of the sequence.

To evaluate the effectiveness of our study, we do five experiments using different sequences, respectively. The portal venous phase and equilibrium phase distribution are different from others because of the absence of portal venous phase in middle and advanced stage number 7, equilibrium phase in normal stage numbers 18 and 21–24, and middle and advanced stage numbers 6 and 7. Yet, we still keep the fact that the test cases are completely new compared with train cases.

The accuracy of every kind of sequence is demonstrated in Table 4. Firstly, it has a perfect result of case accuracy, that is, 100%, which means that all the cases can be classed correctly from case point. Secondly, ROI accuracy is not as good as case accuracy. Yet, it has completely no influence on case accuracy when taking all ROIs results of one case into account and it still has a remarkable performance, especially for normal stage. Normal stage can absolutely be separated from early stage and middle and advanced stage according to no matter ROI classification result or case classification result. To early stage, CCTCRF can almost be separated from others according to ROI classification result, except that ROI accuracy of T2-weighted images is 97.30%, because one of two ROIs of number 3 case is erroneous classified. Yet, case number 3 still classifies correctly after considering all the number 3 case ROIs.

Thirdly, the middle and advanced ROI accuracy is not as good as normal or early stage because the cirrhosis texture in middle and advanced liver is very irregular and is accompanied by morphological changes of liver which causes the difficulty to extract suitable ROI. However, it has no influence on case accuracy when considering all ROIs results, just like early stage number 3. In a word, all the cases have an encouraging case accuracy, that is, 100%, and maintain a remarkable ROI accuracy especially for normal and early stage which is more meaningful for doctors and patients.

3.2. NN Classifier with GLCM Texture Features in Comparison Study. We use five sequences MRI for NN classifier with GLCM texture features experiment and the number of patients is shown in Table 1. The texture features based on GLCM is classical statistics features. GLCM is a matrix that describes the probability of a couple of pixels whose gray levels are i and j, and the distance and direction between the couple pixels is d and θ. We did three stages (normal stage, early stage, and middle and advanced stage) classification experiment with 14 kinds of GLCM texture features of ROIs. And ROIs were extracted from MRI with the same extraction principle which is excluding the diffuse distribution of liver, large blood vessels within the liver [8]. We perform tenfold cross validation method to execute the classification and the result is shown in Figure 5. Obviously, the NN classifier

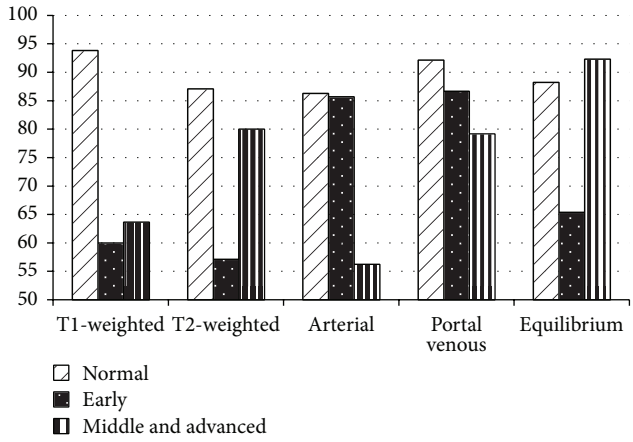


Figure 5: NN classification result.

with GLCM texture features does not have a harmonious or remarkable performance in cirrhosis classification with any sequence of MRI. However, it might be improved by adding feature selection and adjusting parameters, including adopting different angle and step to compute GLCM according to different MR sequence.

4. Conclusions

In this paper, we describe CCTCRF method with five sequences MRI, T1-weighted, T2-weighted, hepatic arterial phase, portal venous phase, and equilibrium phase, to class cirrhosis into three stages (normal, early, and middle and advanced stage). The experiment results show that CCTCRF method surpasses the typical classifier NN with GLCM texture features in cirrhosis classification, but with significant reductions in complexity and time. There are many advantages of CCTCRF than previous studies such as NN classifier with GLCM in cirrhosis texture classification.

Instead of GLCM texture features, we choose compressed patch vector as the texture features. Firstly, compared with GLCM which needs strong assumption including angle and step, compressed patch vector does not need strong assumptions about the texture images, except the sparse character of the image. And we apply the texture extraction method into five MRIs that all have sparse character without considering specific parameters adjusting for specific sequence and the experiment result also demonstrates the effectiveness of the five sequences MRI texture features. Secondly, it almost does not lose information when extracting texture feature, because no information is lost in the three steps which are extracting local image from image, extracting patch from local image, and stretching patch into patch vector. Meanwhile, the process of compressing patch vector process retains salient information when image is sparse. Thus, the texture feature, which is compressed patch vector, almost does not lose information. However, GLCM texture features must lose texture information under such a strong assumption condition. Thirdly, the texture features do not need feature selection, because it is already simplified with compressing and has a remarkable performance in experiment without feature selection. So, the omitting of feature selection reduces method complexity without reducing accuracy. Finally, CCTCRF texture feature extraction costs little time compared with GLCM texture feature extraction, which heavy computation burden and time-consuming are its big problem especially without feature selection. CCTCRF texture feature extraction has the simple texture feature extraction procedure which includes extracting local image from image, extracting patch from local image, stretching patch into patch vector, and compressing patch vector that only needs matrix multiplication. Meanwhile, omitting feature selection can also save time. Thus, the feature extraction of CCTCRF method is more suitable for the further popularization of cirrhosis CAD system that needs to provide diagnosis fast and accurately.

Further, CCTCRF experiment result has confirmed that the five sequences hepatic MRIs of T1-weighted, T2-weighted, arterial phase, portal venous phase, and equilibrium phase are all prominent for every stage cirrhosis classification. In addition, the case decision making step achieves a remarkable performance on the basis of ROI classification result which is already satisfying, and it eliminates few special ROIs inaccurate classification with an overall decision making of one case.

The promising results of this paper motivate a further research of cirrhosis diagnosis and classification. In the future, we will use more cases to verify CCTCRF's effectiveness and add morphological features into middle and advanced classification to improve its ROI accuracy. Furthermore, we will class the cirrhosis with more subtle stages with CCTCRF and apply CCTCRF in hepatic fibrosis classification which is more meaningful for doctors and patients and our medical development is our research direction in the future.

Conflict of Interests

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 61003175 and 81071127), the Fundamental Research Funds for the Central Universities, and the National Key Technology Support Program no. 2012BAJ18B06.

References

[1] G. Lee, M. Kanematsu, H. Kato et al., "Unsupervised classification of cirrhotic livers using MRI data," in *Medical Imaging: Computer-Aided Diagnosis*, vol. 6915 of *Proceedings of SPIE*, San Diego, Calif, USA, 2008.

[2] Y. Chen, J. Luo, T. Tateyama et al., "Statistical shape model of the liver and effective mode selection for classification of liver cirrhosis," in *Proceedings of the 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM '12)*, pp. 446–449, Taipei, Taiwan, 2012.

[3] G. Lee, Y. Uchiyama, X. Zhang et al., "Classification of cirrhotic liver in gadolinium-enhanced MR images," in *Medical Imaging: Computer-Aided Diagnosis*, vol. 6514 of *Proceedings of SPIE*, San Diego, Calif, USA, February 2007.

[4] L. Hui, G. D. Mei, and L. Xiang, "Cirrhosis classification based on MRI with duplicative-feature support vector machine (DFSVM)," *Biomedical Signal Processing and Control*, vol. 8, no. 4, pp. 346–353, 2013.

[5] W. Li, Y. Zeng, X. Zhang, Y. Huang, L. Long, and H. Fujita, "Staging the hepatic fibrosis on CT images: optimizing the slice thickness and texture features," in *Proceedings of the 2nd International Symposium on Bioelectronics and Bioinformatics (ISBB '11)*, pp. 267–270, Suzhou, China, November 2011.

[6] Y. Chen, J. Luo, C. Dong et al., "Computer-aided diagnosis and quantification of cirrhotic livers based on morphological analysis and machine learning," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 264809, 8 pages, 2013.

[7] H.-Y. Jiang, Y. Zhao, and X.-F. Yang, "Computer aided diagnosis on hepatocirrhosis degree based on MRI," *Journal of Northeastern University*, vol. 28, no. 1, pp. 31–34, 2007.

[8] Y. F. Zeng, X. J. Zhang, W. Yan et al., "Computer aided interpretation of fibrous texture in hepatic magnetic resonance images," *Advanced Materials Research*, vol. 647, pp. 325–330, 2013.

[9] J. Virmani, V. Kumar, N. Kalra, and N. Khandelwal, "Prediction of cirrhosis based on singular value decomposition of gray level co-occurence marix and a neural network classifier," in *Proceedings of the 4th International Conference on Developments in E-Systems Engineering (DeSE '11)*, pp. 146–151, Dubai, UAE, December 2011.

[10] J. Virmani, V. Kumar, N. Kalra, and N. Khandelwal, "SVM-based characterisation of liver cirrhosis by singular value decomposition of GLCM matrix," *International Journal of Artificial Intelligence and Soft Computing*, vol. 3, no. 3, pp. 276–296, 2013.

[11] S. Gao, Y. Peng, H. Guo et al., "Texture analysis and classification of ultrasound liver images," *Bio-Medical Materials and Engineering*, vol. 24, no. 1, pp. 1209–1216, 2014.

[12] L. Liu and P. Fieguth, "Texture classification from random features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 574–586, 2012.

[13] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '89)*, vol. 1, pp. 593–605, Washington, DC, USA, June 1989.

[14] D. Albesano, R. Gemello, and F. Mana, "Hybrid HMM-NN modeling of stationary-transitional units for continuous speech recognition," *Information Sciences*, vol. 123, no. 1, pp. 3–11, 2000.

[15] F. Jelinek, "Continuous speech recognition by statistical methods," *Proceedings of the IEEE*, vol. 64, no. 4, pp. 532–556, 1976.

[16] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, 1997.

[17] H. A. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 38–44, Santa Barbara, Calif, USA, June 1998.

[18] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998.

[19] I. Guyon, J. S. Denker, and Y. LeCun, "Time delay neural network for printed and cursive handwritten character recognition," Google Patents, 1992.

[20] N. Arica and F. T. Yarman-Vural, "An overview of character recognition focused on off-line handwriting," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 31, no. 2, pp. 216–233, 2001.

[21] G. Zhang, H. Rong, W. Jin, and L. Hu, "Radar emitter signal recognition based on resemblance coefficient features," in *Rough Sets and Current Trends in Computing*, vol. 3066 of *Lecture Notes in Computer Science*, pp. 665–670, 2004.

[22] A. Kundu, G. C. Chen, and C. E. Persons, "Transient sonar signal classification using hidden Markov models and neural nets," *IEEE Journal of Oceanic Engineering*, vol. 19, no. 1, pp. 87–99, 1994.

[23] B. Barshan, B. Ayrulu, and S. W. Utete, "Neural network-based target differentiation using sonar for robotics applications," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 4, pp. 435–442, 2000.

[24] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[25] K. Li, *Study on breast lesions computer-aided diagnosis between malignant and benign based on DCE-MRI [M.S. thesis]*, Dalian University of Technology, Dalian, China, 2011.

[26] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[27] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[28] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[29] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[30] N. Goel, G. Bebis, and A. Nefian, "Face recognition experiments with random projection," in *Biometric Technology for Human Identification II*, vol. 5779 of *Proceedings of SPIE*, pp. 426–437, Orlando, Fla, USA, March 2005.

[31] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," in *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '01)*, pp. 245–250, San Francisco, Calif, USA, August 2001.

[32] X. Z. Fern and C. E. Brodley, "Random projection for high dimensional data clustering: a cluster ensemble approach," in *Proceedings of the 20th International Conference on Machine Learning (ICML '03)*, pp. 186–193, Washington, DC, USA, August 2003.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 726712, 6 pages
http://dx.doi.org/10.1155/2014/726712

Research Article

A Novel Blind Separation Method in Magnetic Resonance Images

Jianbin Gao,[1,2,3] Qi Xia,[4] Lixue Yin,[3] Ji Zhou,[1] Li Du,[5] and Yunfeng Fan[1]

[1] *School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2] *Key Laboratory of Integrated Electronic System, Ministry of Education, Chengdu 611731, China*
[3] *Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu 610072, China*
[4] *School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China*
[5] *Southwest China Research Institute of Electronic Equipment, Chengdu 610036, China*

Correspondence should be addressed to Jianbin Gao; gaojianbin0769@gmail.com

Received 22 November 2013; Accepted 17 January 2014; Published 23 February 2014

Academic Editor: Yuanjie Zheng

A novel global search algorithm based method is proposed to separate MR images blindly in this paper. The key point of the method is the formulation of the new matrix which forms a generalized permutation of the original mixing matrix. Since the lowest entropy is closely associated with the smooth degree of source images, blind image separation can be formulated to an entropy minimization problem by using the property that most of neighbor pixels are smooth. A new dataset can be obtained by multiplying the mixed matrix by the inverse of the new matrix. Thus, the search technique is used to searching for the lowest entropy values of the new data. Accordingly, the separation weight vector associated with the lowest entropy values can be obtained. Compared with the conventional independent component analysis (ICA), the original signals in the proposed algorithm are not required to be independent. Simulation results on MR images are employed to further show the advantages of the proposed method.

1. Introduction

Blind source separation (BSS) aims at recovering unknown source signals only from the observed data. It has received considerable attention for its potential applications in a lot of fields, such as biomedical signal processing, image processing, and digital communications. The basic instantaneous linear mixture model used in BSS is as follows:

$$x(t) = As(t), \tag{1}$$

where $s(t) = [s_1(t), s_2(t), \ldots s_m(t)]^T$ is a $m \times 1$ vector of source signals which represents the samples of unobserved source signals, $x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$ is a $n \times 1$ vector of mixed signals observed by n sensors, and $A = [a_1, a_2, \ldots, a_m]$ is an unknown $n \times m$ mixing matrix of full rank. Assume the weight matrix $W \in R^{n \times m}$ and the output $\widehat{s}(t) = Wx(t)$ at time t. The goal of a BSS algorithm is to find a weight matrix W such that $\widehat{s}(t)$ is a permutation of source signals $s(t)$ up to a scaling factor. It is known that when $n > m$ the principal component analysis (PCA) technique can be used to reduce the dimensionality of observations. For this reason, we only consider the case that $n = m$ in this paper.

Since the pioneering work by Hyvärinen et al. [1], various separation algorithms have been proposed for different BSS subjects [2–6], for example, Oja et al. considered the nonnegative assumption and proposed some algorithms to separate these nonnegative sources [7–9]. A particular application of these algorithms is the blind separation of mixed images. More recently, much attention has been paid to BSS methods that make use of a priori information, such as sparse component analysis which works under the assumption that the sources can be represented by sparse signals. For the application of image separation, we here consider the a priori information that comes from the observation that most of neighbor pixels in a small patch are smooth. The local smoothness means little randomness, that is, the lower entropy values in the small patch compared to that of

the image space. As stated in [10] that the smooth degree of any linear mixture of the source images is between the greatest smooth degree and the smallest smooth degree of the source images, we can formulate a proper entropy like function so that the source images would have the lowest entropy and their mixtures would have higher entropy values. By taking the entropy like function as the objective function, the global search technique is used to obtain the lowest entropy values of image signal, that is, the source images. The two-dimensional matrix formats will be treated for utilizing the full information carried by images. The result of experiment demonstrated that our method provides a good separation performance even for rich texture images. In [11], by using SVD technique, the mixed images are decomposed into three parts and the global stochastic optimization technique is used to recover the source images by searching for the lowest entropy values of images. Although Guo and Garland's algorithm has achieved a better performance compared to the conventional ICA method, however, it will cause large memory requirements. In other words, their algorithm is infeasible on most of computers. In [10], the separation performance is low in separating texture images.

2. Materials and Methods

2.1. Problem Statement. For natural image signal, the neighbor pixels in a small patch always present strongly smooth property, which means that images are locally smooth (see [10] for more details). However, the smooth property is inapplicable to those images with rich texture. For example, the textural image is coarser according to the six features of image proposed by Tamura et al. To evaluate the coarseness of natural image and textural image, the following coarseness measure is used, which is defined in [12],

$$F_{crs} = \frac{1}{m \times n} \sum_{i}^{m} \sum_{j}^{n} S_{\text{best}}(i, j), \tag{2}$$

where m and n are the effective width and height of the picture, respectively. The coarseness of natural image (Figure 1(a)) and textural image (Figure 1(b)) is, respectively, 0.9754 and 30.9368. Two segments of the natural image and the textual image, which are transformed to a vector in row-wise order, respectively, are drawn in Figure 1(c). Both images are of unit variance. From this key observation, the textural image is more random than natural image.

Motivated by the special phenomenon depicted in Figure 1(c), we can formulate (3) as follows:

$$y(k) = s(t) - s(t-k), \tag{3}$$

where k represents a positive integer. The key point here is that $y(k)$ has a large amount of columns in which only one element is not equal to zero. This phenomenon is shown in Figure 2.

By substituting (3) into (1), we can obtain the following equation:

$$z(k) = x(t) - x(t-k) = A(s(t) - s(t-k)). \tag{4}$$

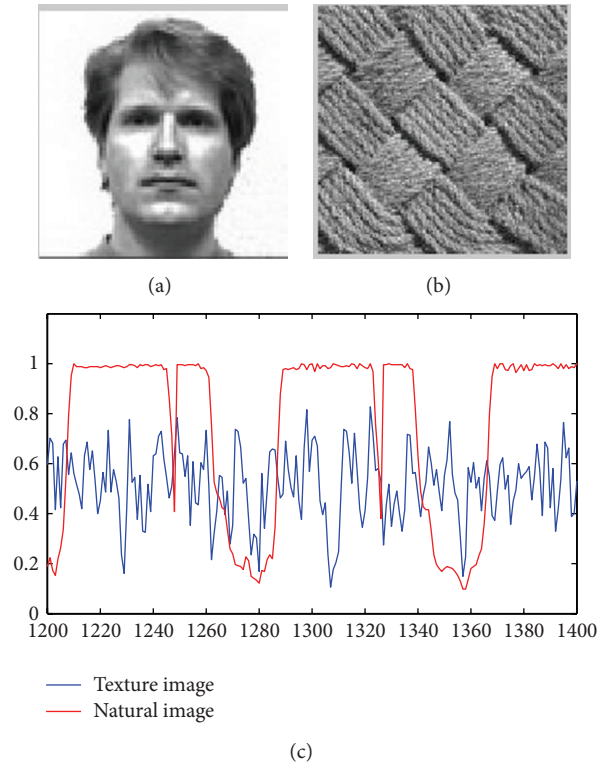


Figure 1: (a) Natural image; (b) textural image; (c) the segments of textural image and natural image.

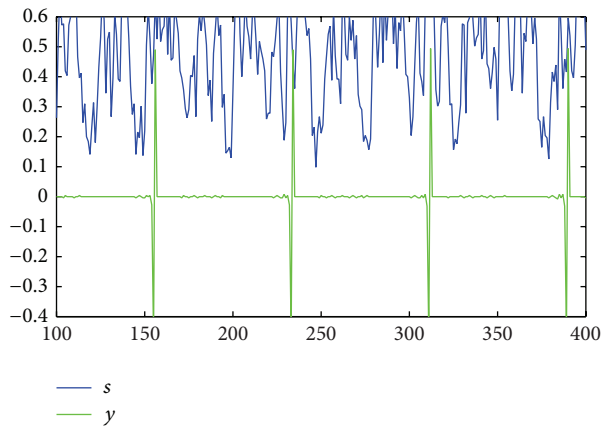


Figure 2: Comparison of s and y.

For image signals, some hypotheses are made naturally.

(A) Pixels are positive.

(B) There at least exist m vectors within the matrix $z(k)$ which can construct the $n \times m$ matrix $\overline{A}$ which is a generalized permutation of the mixing matrix A up to a scaling factor.

For image signals, the hypothesis A is very normal, and hypothesis B can be met by modulating parameter k. Since

$y(k)$ has a large amount of columns that only one element is not equal to zero, the matrix $z(k)$ would contain some columns which can form the matrix $\overline{A} \in R^{n\times m}$, which is a permutation of A up to a scaling factor. Namely, the matrix $\overline{A}$ is equivalent to the mixing matrix A by using normalization. Thus, the new matrix can be achieved as follows:

$$\bar{s} = \overline{A}^{-1} x, \tag{5}$$

where (−1) represents inverse operator. As stated above, the entropies of pure images would be lower than those of the mixed images. We take the form of (6) to minimize an entropy like function to approximate the entropy of $\bar{s}$ [11, 13]:

$$E = -\sum_i \sum_j p_{ij} \ln p_{ij}, \tag{6}$$

where $p_{ij} = (|x(i+1, j+1) + x(i, j) - x(i, j+1) - x(i+1, j)|)/(\sum |x(i+1, j+1) + x(i, j) - x(i, j+1) - x(i+1, j)|)$. The probability distribution p_{ij} is the second derivative of the data [11]. In other words, the entropy like function E estimates the smoothness of the images. The pure source images can be obtained by solving the following minimization problem:

$$\min F_{\text{obj}} = E + \max\left\{p \times \left(e^{q/(1-\text{corr})} - 1\right)\right\}, \tag{7}$$

where corr denotes the 2D correlation coefficient between any two matrices and the parameter $q = 0.0002$ and $p = 100000$, empirically. Thus, we can take (7) as the objective function of global search algorithm.

2.2. The Standard Global Search Algorithm. In this paper, we use global search algorithm, such as Genetic Algorithm (GA) which is a useful solution to optimize and search problems, to achieve pure images. Generally, GA, one of the popular global stochastic optimization techniques, has been used to separate blind sources. This algorithm belongs to the larger class of evolutionary algorithms which are stemmed from the natural genetics and biological evolutionary process. The GA evaluates a population and generates a new one iteratively, with each successive population referred to as a generation. Given the current generation at iteration $t, G(t)$, the GA generates a new generation $G(t+1)$, based on the previous generation, applying a set of genetic operations. The GA uses three basic operators to manipulate the genetic composition of a population: selection, crossover, and mutation [14]. Selection process determines the individuals for reproduction and the number of offspring that an individual can produce. Generally, we select ninety percent of individuals to produce new individuals and keep ten percent of individuals which have minimum values. During the selection process, each individual of current population is assigned a fitness value derived from the corresponding objective function value. Then, the selection algorithm selects individuals for reproducing on the basis of their relative fitness values. In our method, the fitness values are calculated using linear ranking method with pressure two which can prevent premature convergence [12]. The fitness of ith individual in the population is defined as follows:

$$F(x_i) = 2 - \text{Max} + 2(\text{Max} - 1)\frac{x_i - 1}{p - 1}, \tag{8}$$

where Max is always chosen in [1.1, 2], which is used to determine the selective pressure such that no individuals generate an excessive number of offspring. And x_i is the position of the ith individuals in the reordered population based on their corresponding objective function values. The crossover operator mixes the genes of two chromosomes selected in the phase of reproduction, in order to combine the features, especially the positive ones of them. In the proposed algorithm, the simplest form of crossover is used, that is, one-point crossover.

2.3. The Proposed Algorithm. Based on the state above, we first utilize (3) to construct the new dataset t. Then we discard those columns, whose all elements are equal to zero and change those columns, whose all elements are negative values, to positive value by multiplying by −1. After that, reindex the vectors in $z(k)$. In order to obtain a large amount of information of mixed image, the value of k should not be large. Those columns, whose elements are all equivalent to zero, should be discarded from $z(k)$. At the same time, those columns, whose elements are negative, should be changed to nonnegative value by multiplying by −1. Then we will call GA twice. In the first time, the first pure image would be obtained by minimizing the objective function using GA. That is to say, we can find m vectors, which can form the matrix $\overline{A}$, from $z(k)$. It means that the first pure image can be obtained when the correlation coefficient is the minimum. The first pure image would be saved and the GA is called again. Then we compare all separated images with the first pure image. In the case of only one separated image highly related with the pure image and the others which are mostly uncorrelated with the pure image, we can select a matrix of $\overline{A}$ to separate mixed images in each iteration. At the end of iteration, we can achieve a best matrix from $\overline{A}$.

The blind separation algorithm based on GA can be summarized as follows.

(1) Form dataset $z(k)$ by computing $z(k) = x(t) - x(x - k)$.

(2) Change $z(k)$ to be nonnegative and discard those columns whose all elements are equal to zero.

(3) Reindex the vectors in $z(k)$.

(4) Use GA to get the first pure image and save this image as reference image.

(5) Call GA and compare all separated images with the first pure image by computing the correlation iteratively. Thus, a best separated matrix can be achieved.

(6) Estimate $\overline{S}(t)$ by computing (5).

3. Results

3.1. Separation of Texture Image. In order to evaluate the proposed method, we tested MR images [15]. The digital imaging

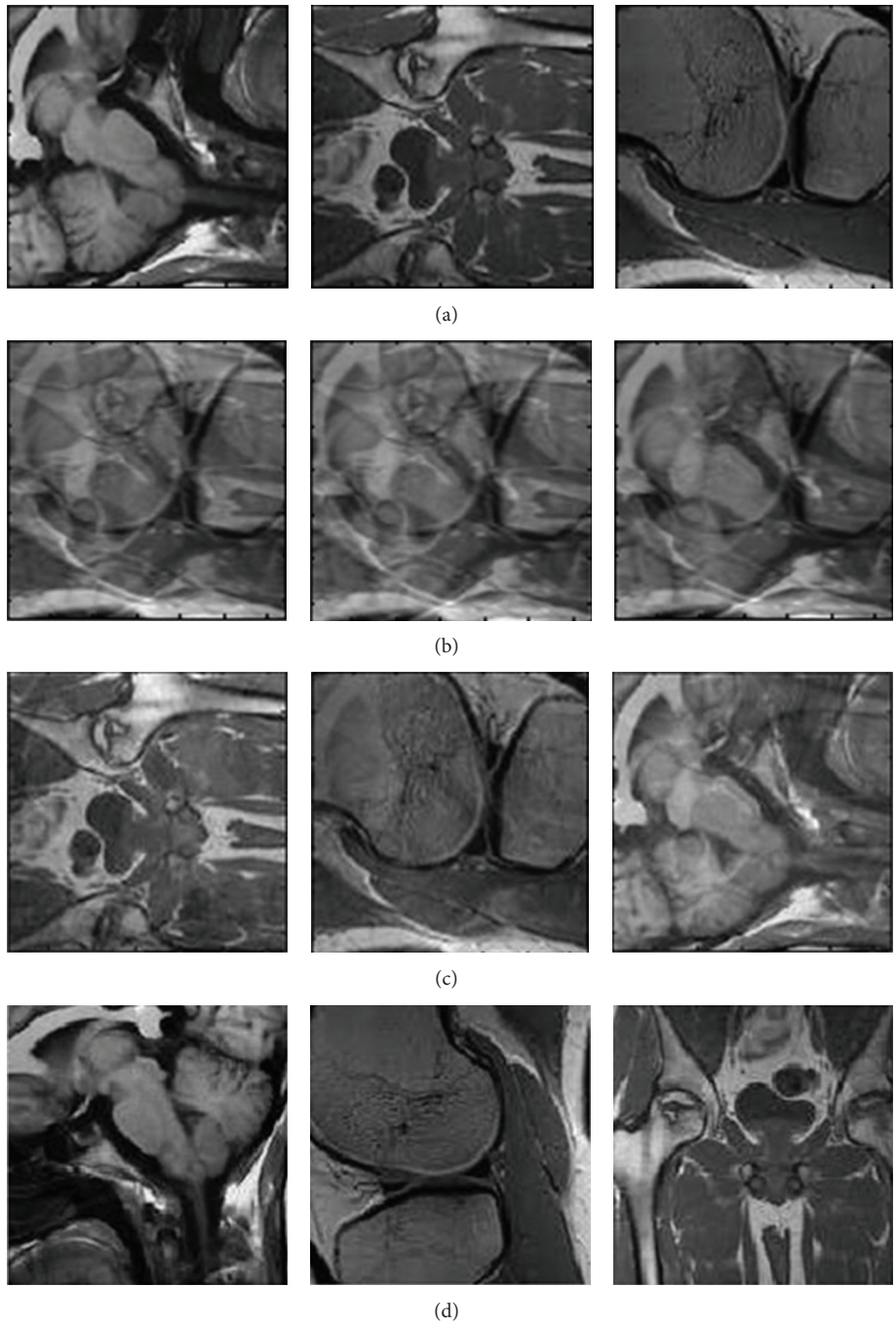
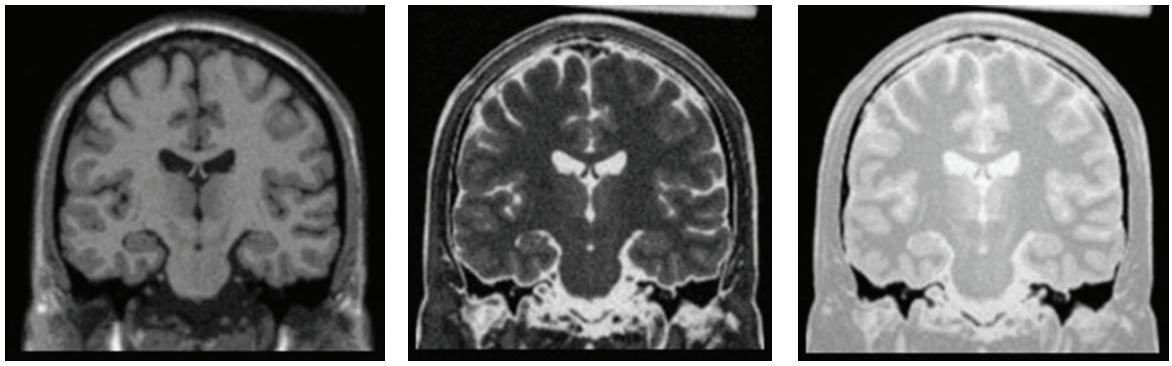


Figure 3: Results of standard ICA and the proposed method on 3 correlated MRIs with noise variance 0.01. (a) Source images; (b) mixed sources; (c) separation results using ICA method; (d) separation result using the proposed method.

and communications in medicine (DICOM) standard was created by the National Electrical Manufacturers Association (NEMA) to aid the distribution and viewing of medical images, such as MR scans, and ultrasound. For this experiment, we have collected 3 MR scans whose correlation coefficients are between 0.6 and 0.8. The separated results are presented as follows. Figures 3(a), 3(b), and 3(c) illustrate the results of proposed algorithm. For simplicity, we here only plot the separated images with the variance $\sigma^2 = 0.01$.

To evaluate the separation performance, the following performance index (PI) in [9] is used, which is defined by

$$\mathrm{PI} = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|C_{ij}|}{\max_k |C_{ik}|} - 1\right) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|C_{ij}|}{\max_k |C_{kj}|} - 1\right), \tag{9}$$



(a) (b) (c)

Figure 4: Simulated MR scans. (a) Spin-lattice; (b) spin-spin; (c) proton density.

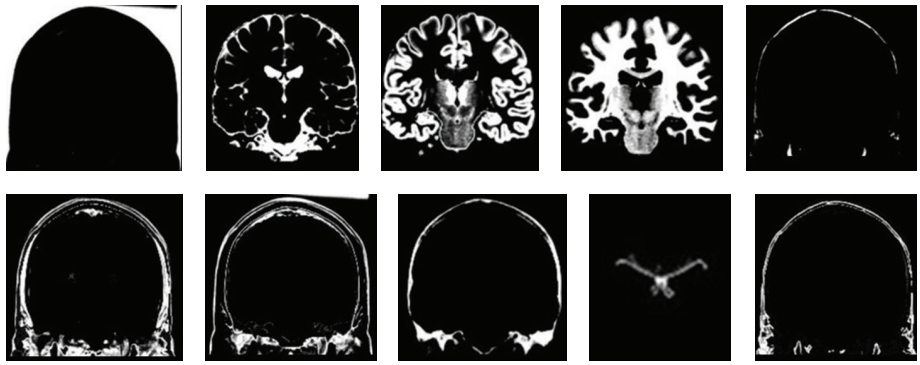


Figure 5: Ground truth images of different brain tissue substances.

Table 1: Separation performance with different variances of Gaussian white noise (average over 100 trials).

Algorithm	PI ($\sigma^2 = 0$)	PI ($\sigma^2 = 0.01$)	PI ($\sigma^2 = 0.02$)
Proposed method	$2.3154e-012$	0.01324	1.5321
ICA method	4.2405	4.2305	4.7959

where $C_{ij} = \overline{A}^{-1}A$ is the combination of the separating and mixing matrix. The PIindex is equal to zero if and only if the matrix $\overline{A}$ is a permutation of A. The comparisons of the two methods with different variances of zero mean Gaussian white noise averaging over 100 trials are listed in Table 1.

From Table 1, we can see that the proposed algorithm in a noise scenario is superior accuracy compared to the ICA method.

4. Discussion

The brain has a number of constituents in the context of a MRI scan of the brain, such as gray matter, white matter, cerebrospinal fluid (CSF) fat, muscle/skin, and glial matter. Now since each is unique, they would show unique characteristics under a magnetic field. However, while taking a scan, we get on MRI image of the entire brain. These scans can be considered as an equivalent to the mixtures of the blind source. The blind source separation technique can be used for this to separate out the various constituents such as gray matter, white matter, and CSF. These images of independent sources can be used for better diagnosis. The MR scans are from the McGill Simulated Brain Database as shown in Figure 4 [16].

Actually, the images [16] for these scans would be as shown in Figure 5.

Magnetic Resonance Imaging can give much better soft tissue contrast than that of CT for brain imaging, so MRI is superior to CT. It means that even small changes in the proton density and composition in the tissue are well represented by MRI. Some new methods and techniques can be used to improve scans obtained by MRI to improve diagnosis. Only in the past decade, various algorithms have been proposed to separate physiologically different components from EEG or MEG data [17, 18], financial data [19], and even in fMRI [20, 21]. However, for MRI, BSS-based methods have not gained much attention. Nakai et al. utilized ICA for the purpose of separating physiologically independent components from MRI scans [22]. They took MR images of 10 normal subjects, 3 subjects with brain tumor, and 1 subject with multiple sclerosis and performed ICA on the data. They reported success in improving contrast for gray and white matter, which was conducive to the diagnosis of brain tumor. The demyelination in multiple sclerosis cases was also enhanced

in the images. The ICA method could potentially separate out all the tissues which had different relaxation characteristics according to their research result which shows much promise in biomedical domain. Take a set of MR frames as a single multispectral image, where each band is taken during a particular pulse sequence. Then ICA can be used on the data to separate out the physiologically independent components. Generally, a classifier such as the SVM would be used to improve the contrast of the separated components.

5. Conclusions

In this paper, a novel GA-based algorithm is proposed to separate MR images blindly by using smooth information in both noisy-free and noise scenarios. In order to take advantage of MR scans structure, we use an entropy like function to represent the local smooth property of near pixels. Let the entropy like function be the objective function, the GA is used for searching for the lowest entropy values. The performance of the proposed method is tested on NEMA MR image database. Simulations confirm the efficiency and effectiveness of the proposed algorithm. Because the standard GA method is sensitive to strong noise (see Table 1), further work is on the way to extend our method to the high noise scenario.

Conflict of Interests

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

Acknowledgments

This work is supported in part by the Postdoctoral Science Foundation of China (2013M542267), Open Research Foundation of Integrated Electronic System of the Ministry of Education of China (20120105), the Fundamental Research Funds for the Central Universities (ZYGX2013J118), the Fundamental Research Funds for the Central Universities (ZYGX2011J067), the NSFC of China for International Young Scientists under Grant 61250110543, and National Natural Science Foundation of China (41371341).

References

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[2] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1004–1034, 1995.

[3] E. Oja and M. D. Plumbley, "Blind separation of positive sources by globally convergent gradient search," *Neural Computation*, vol. 16, no. 9, pp. 1811–1825, 2004.

[4] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[5] W.-Q. Guo and T.-S. Qiu, "Adaptive blind estimation of evoked potentials in EEG based on minimum dispersion coefficient and revolving transform," *Chinese Journal of Biomedical Engineering*, vol. 26, no. 5, pp. 647–651, 2007.

[6] M. D. Plumbley and E. Oja, "A "nonnegative PCA" algorithm for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 66–76, 2004.

[7] M. D. Plumbley, "Conditions for nonnegative independent component analysis," *IEEE Signal Processing Letters*, vol. 9, no. 6, pp. 177–180, 2002.

[8] A. Cichocki and S.-I. Amar, *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, 2002.

[9] M. Ye, Y. Liu, M. Liu, F. Li, and Q. Liu, "Blind image extraction by using local smooth information," in *Proceedings of the 5th International Conference on Natural Computation*, vol. 3, pp. 415–420, Tianjin, China, 2009.

[10] J. Gao and M. Ye, "Blind separation of locally smooth images based on genetic algorithm," *Journal of Computational Information Systems*, vol. 6, no. 8, pp. 2465–2472, 2010.

[11] L. Guo and M. Garland, "The use of entropy minimization for the solution of blind source separation problems in image analysis," *Pattern Recognition*, vol. 39, no. 6, pp. 1066–1073, 2006.

[12] H. Tamura, T. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[13] Y. Zeng and M. Garland, "An improved algorithm for estimating pure component spectra in exploratory chemometric studies based on entropy minimization," *Analytica Chimica Acta*, vol. 359, no. 3, pp. 303–310, 1998.

[14] J. E. Baker, "Adaptive selection methods for genetic algorithms," in *Proceedings of the 1st International Conference on Genetic Algorithms (ICGA '85)*, pp. 101–111, 1985.

[15] ftp://medical.nema.org/medical/Dicom/Multiframe/.

[16] http://brainweb.bic.mni.mcgill.ca/brainweb/anatomic_normal.html.

[17] S. Makeig, A. J. Bell, T. P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," *Advances in Neural Information Processing Systems*, pp. 145–151, 1996.

[18] M. Kawakatsu, "Application of ICA to MEG noise reduction," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, 2003.

[19] Y. S. Abu-Mostafa, "Independent component analysis in financial data," in *Computational Finance*, The MIT Press, 2000.

[20] V. D. Calhoun T, Adali, L. K. Hansen, J. Larsen, and J. J. Pekar, "ICA of functional MRI data: an overview," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation*, Nara, Japan, 2003.

[21] M. J. McKeown and T. J. Sejnowski, "Independent component analysis of fMRI data: examining the assumptions," *Human Brain Mapping*, vol. 6, no. 5-6, pp. 368–372, 1998.

[22] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 302458, 12 pages
http://dx.doi.org/10.1155/2014/302458

Research Article

Left Ventricular Endocardium Tracking by Fusion of Biomechanical and Deformable Models

Hussin Ketout and Jason Gu

Electrical and Computer Engineering, Dalhousie University, 1360 Barrington Street, Halifax, NS, Canada B3J 2X4

Correspondence should be addressed to Hussin Ketout; hussin.ketout@dal.ca

Received 27 July 2013; Accepted 19 November 2013; Published 21 January 2014

Academic Editor: Rong Chen

This paper presents a framework for tracking left ventricular (LV) endocardium through 2D echocardiography image sequence. The framework is based on fusion of biomechanical (BM) model of the heart with the parametric deformable model. The BM model constitutive equation consists of passive and active strain energy functions. The deformations of the LV are obtained by solving the constitutive equations using ABAQUS FEM in each frame in the cardiac cycle. The strain energy functions are defined in two user subroutines for active and passive phases. Average fusion technique is used to fuse the BM and deformable model contours. Experimental results are conducted to verify the detected contours and the results are evaluated by comparing themto a created gold standard. The results and the evaluation proved that the framework has the tremendous potential to track and segment the LV through the whole cardiac cycle.

1. Introduction

Echocardiography is an important imaging modality that enables the cardiologist to evaluate the structure and functions of the heart. Because of noninvasive characteristics, low cost, and being nonionizing radiation, echocardiography has been largely applied in the evaluation of cardiac function. One of the most important applications of echocardiography is in determining systolic and diastolic ventricular volumes of the patient, both of which are used to calculate the left ventricular ejection fraction, muscle contraction ratio of cardiac cavities, local ejection fraction, myocardial thickness, and the ventricle mass [1]. To calculate the above-mentioned parameters, the cardiac muscle contour on the echocardiography image needs to be identified. The border detection process simplifies image analysis and greatly reduces the amount of data which needs to be processed, while preserving the structural information about the contours of the object under study [2].

However, in clinical practice, this task still relies on manual outlining. Manual outlining of these borders is slow, time consuming, and tedious task. Moreover, the resulting outlines vary between different observers and suffer from a subjective bias [3].

Automatic LV border detection and tracking over the cardiac cycle in echocardiographic image sequences remain open and a challenging problem due to many difficulties related to the heart and its dynamics and other difficulties related to the echocardiography ultrasound machine.

Echocardiography has a poor image quality and resolution with various image artifacts like speckle, shadowing, and side lobes [3]. The images of echocardiography suffer from signal dropout. This dropout in the echocardiography signal makes part of the LV invisible which yields an open contour [3]. Moreover, the relation between the physical property of the monitored tissue and the intensity of the pixels cannot be described in a simple way [4]. Due to the highly anisotropic information of the 2D echocardiography images and the mentioned artifacts, the locations of the real contour do not always correspond to the locations of the strong image features like strongest edges. Relaying on the strongest image features will not lead to the detection of the desired contour as traced by cardiologists [4]. Another challenge is the Gray level intensity variability and low signal to noise ratio [3, 5].

The difficulties and challenges that relate to the heart which impede the LV detection of LV boundaries are.

(i) The heart is a highly deformable object with a wide range motion.

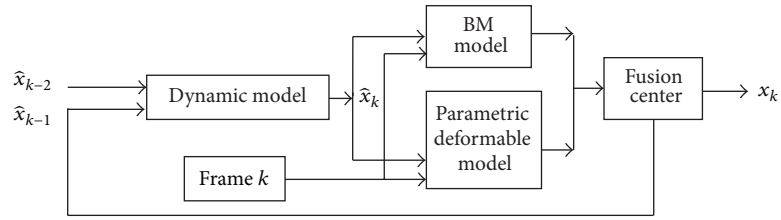


Figure 1: LV endocardium contour tracking framework.

(ii) The papillary muscle is impeding the detection of the real LV boundary.

(iii) The mitral valve also is one of the artifacts and causes a problem for the detecting the LV boundaries.

To tackle the challenges and difficulties due to echocardiographic images and due to LV complex motion, an averaging framework is employed which is based on fusing the parametric deformable and BM models computed contours. It has been found that the different contour detectors used for a special detection task usually complement each other with respect to the information extracted from the patterns to be classified [4]. As a result, combining different contour detectors in an efficient way, is expected to achieve better contour detection results than any single contour detector, and hence produce incremental gains in overall performance.

The biomechanical model is used to mimic the mechanics of the heart muscles specifically myocardium. This particular model is centered on using finite deformation elasticity technique. Once the modeling is done using this technique, the subsequent functional is solved using FEM. The accurate model of myocardial mechanics can simulate the heart wall motion during the diastolic and systolic stages of the cardiac cycle [6, 7].

Instead of depending only on the echocardiographic images to detect the LV boundaries, a FEM model for mimicking the LV movement during a cardiac cycle is developed to obtain the LV boundaries from the functions of transversely isotropic material biomechanical model introduced by Lin and Yin [8].

The paper is organized as follows. Section 2 presents the tracking framework. Section 3 introduces the BM model along with the constitutive equations that are used in our work. The boundary conditions and the ABAQUS solver are illustrated also in detail. Section 4 is about the parametric deformable model. Section 5 gives the details of the averaging fusion technique. Experimental results are given in Section 6. Evaluation of the results is provided in Section 7. The discussion and the analysis of the results are given in Section 8. The future work is mentioned in Section 9. The paper's conclusion is given in Section 10.

2. Tracking Framework

The endocardium contour tracking framework for 2D echocardiographic image sequence is shown in Figure 1. The user needs to click inside the ROI and the level set [10] will be used to provide the initial contour x_0. The predicted state of the LV contour will be provided by the dynamic model. To enable modeling of motion in addition to position, second-order dynamical model is used and given as follows [11, 12]:

$$\hat{x}_k = A_1\hat{x}_{k-1} + A_2\hat{x}_{k-2} + (I - A_2 - A_1)\,\overline{x}, \tag{1}$$

where $\overline{x}$ is the mean shape of the LV, $\hat{x}_k$ is the LV contour at time step k, and $\hat{x}_{k-1}$ is the LV contour at time step $k-1$.

3. BM Model Endocardial Contour Detection

Unlike the imaging based techniques, the BM model endocardial contour detection is based on finding the LV deformations by using the heart model and solving the model equations using ABAQUS FEM. By using the BM model and getting LV deformations through the FEM, this technique enables us to avoid and tackle all the challenges and difficulties of segmenting and tracking the LV through echocardiography image sequence that was mentioned in the introduction. The following are the parts of BM model solution.

3.1. Constitutive Equations. A constitutive equation is a mathematical model that characterizes the relationship between stress and deformation. To mimic the LV movement, our work is based on the Lin-Yin model [8] for the constitutive equations based on hyperelastic material theory. In this model, the strain energy functional W is divided into two components: one is the passive (W_{pass}) and the other is the active (W_{active}). Lin and Yin used an exponential function form for the passive strain energy function given as follows:

$$\begin{aligned} W_{\text{pass}} &= c_1\left(e^{Q} - 1\right), \\ Q &= c_2(I_1 - 3)^2 + c_3\,(I_1 - 3)\,(I_4 - 1) + c_4(I_4 - 1)^2, \end{aligned} \tag{2}$$

where c_1, c_2, c_3, and c_4 are the material properties parameters that determined experimentally.

The active strain energy function is given in the polynomial form as follows [8]:

$$\begin{aligned} W_{\text{active}} = {} & c_5 + c_6\,(I_1 - 3)\,(I_4 - 1) + c_7(I_1 - 3)^2 \\ & + c_8(I_4 - 1)^2 + c_9\,(I_1 - 3) + c_{10}\,(I_4 - 1). \end{aligned} \tag{3}$$

Also, c_5, c_6, c_7, c_8, c_9, and c_{10} are the material properties parameters that are determined experimentally [8]. I_1 and

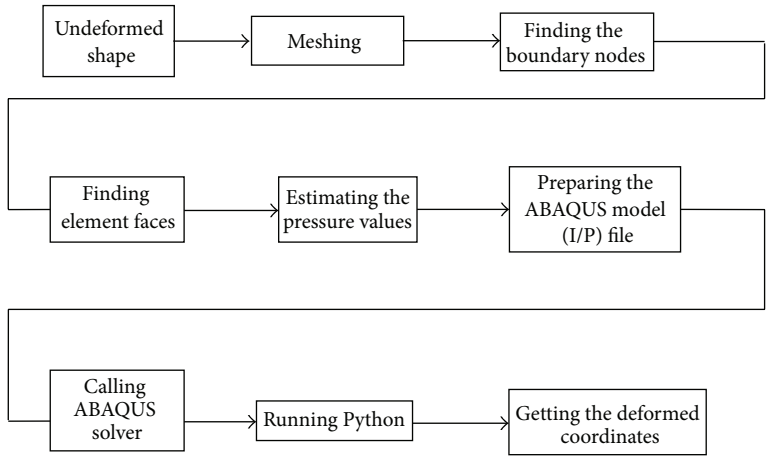


Figure 2: FEM solution block diagram.

I_4 are the invariants of right Cauchy deformation tensor and they are given by the following equations:

$$I_1 = \operatorname{tr} C_{ij} = C_{11} + C_{22} + C_{33},$$
$$I_4{}^2 = N_i C_{ij} N_j. \tag{4}$$

3.2. BM Model Endocardial Contour Detector Framework. The BM model endocardial contour detector for 2D echocardiographic image sequence is shown in Figure 2. The undeformed shape is obtained by using previous frame in the cardiac cycle. The meshing of LV domain is obtained by dividing the domain into triangles as shown in Figure 3. After meshing the LV domain, the next step is finding and identifying all the nodes at the boundary of the LV. This step is necessary to find the nodes where the boundary conditions should be applied. After finding the nodes that form the boundaries of the LV, the elements which relate to the boundary nodes will be identified. After knowing the boundary elements, the faces of each boundary element must be recognized. Each triangle element has three faces. These faces are ordered in anticlockwise as shown in Figure 4. Then, next the face of the triangle element at the boundary should be finding and identified as Face1, Face2, or Face3. After identifying all the faces at the boundary of the LV, the same faces will be grouped together in one group. After that, all the groups will be combined together to form one SURFACE. This SURFACE is where the pressure should be applied during the systole and diastole stages of the cardiac cycle.

3.3. Boundary Conditions. From the heart anatomy, LV is bridled by the atria, RV (right ventricle), and the aorta. The quantitative information of the boundary conditions between these parts of the heart is unknown. To prevent rigid body motion of the left ventricle during the deformation calculation, the basal plane motion should be suppressed.

The load applied to the endocardial SURFACE is the blood pressure and the blood pressure in human heart depends on time and location. The fluid dynamics of the blood pressure in the left ventricle should be taken into account to estimate the spatial distribution of the blood pressure as a function in time. During LV contraction, the pressure gradients are very small compared to the absolute pressure. From this, we can assume a uniform parabolic distribution of the pressure along the endocardial SURFACE of the LV.

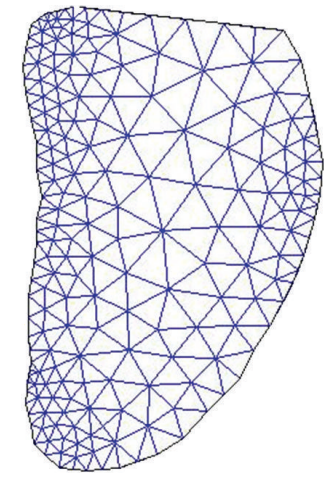


Figure 3: Meshing the LV into M domain cells.

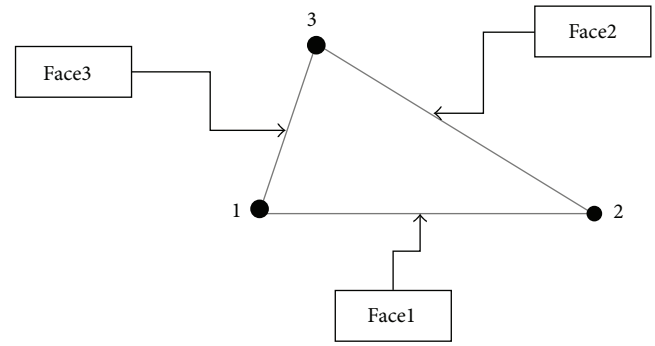


Figure 4: Identifying triangle element's faces.

From tracking of the cardiac cycle, the duration of systole phase lasts for 16 frames starting at the QRS ECG signal. The pressure starts rising up to 80 mmHg and reaches the peak of 120 mmHg and down to 80 mmHg at the end of systole stage.

In the literature [13], the blood pressure of the left ventricle varying with time is simulated as follows:

$$t = 0.055 + k * 0.0090625,$$
$$P = -944.38t^2 + 245.54t, \quad 0 \le t \le 0.2. \tag{5}$$

The total duration of time is divided over the 16 frames which covers the systole stage (k represents the frame number). According to the frame number k, the value of t will be calculated using first equation of (5). The pressure value will be calculated from the second equation of (5). The minimum pressure will be 10.666 Kpa which corresponds to 80 mmHg while the maximum pressure value is 15.96 Kpa which corresponds to 120 mmHg. These values are the normal systolic blood pressure for an intact heart.

In diastole stage, the pressure is approximately fixed and a value of 5 Kpa is used as indicated in the pressure curve in the cardiac cycle diagram.

3.4. ABAQUS Solver. For each frame of the echocardiographic image sequence, ABAQUS model (input file) is prepared by MATLAB script that identifies the entire model and history data which are saved in the input file and run by ABAQUS to estimate the deformations of the LV at that moment in the cardiac cycle. Beside, the input file, two user subroutines are prepared and used in the framework. To identify the biomechanical model to ABAQUS, both the passive and active strain energy functions should be declared in a specific user subroutine in ABAQUS called UANISO-HYPER_STRAIN.F [14, 15]. ABAQUS will be called from MATLAB script and it will run in background mode.

After the completion of the ABAQUS model, a Python program will be called from MATLAB script to read the results from ODB file (Output Data Base) [16, 17].

The Python program will read the deformations (displacements) at each boundary node and add it to the original coordinates to find the deformed contour. After the completion of these calculations, the MATLAB running will continue to find the LV deformed contour using parametric deformable model and then fuse it with BM estimated contour as illustrated in the following sections.

4. Parametric Deformable Model

Kass et al. [18] introduced the concept of active contour models (ACM), or Snake in his paper "Snakes: active contour models." Snakes are used in the area of image processing to detect the object boundaries. Snake is modeled as parametric curve that evolves into a position where its energy functional is minimized. The position of the Snake is given by the parametric curve $C(s) = [x(s), y(s)]$ with $s = [0, 1]$. Kass et al. introduced the following energy functional for the Snake:

$$E_{\text{Total}} = \int_0^1 \left(E_1(c(s)) + E_2(c(s))\right) ds, \tag{6}$$

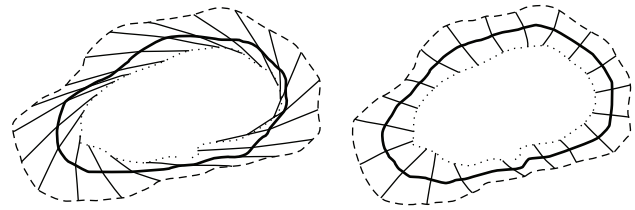


Figure 5: Averaging steps, firstly, finding the correspondence between each control point in both contours and lastly, the averaged contour with the bold line [9].

where

$$E_1 = \int_0^1 \left(\alpha \left\|c'(s)\right\|^2 + \beta \left\|c''(s)\right\|^2\right) ds,$$
$$E_2 = \int_0^1 P(c(s))\, ds. \tag{7}$$

E_1 is the internal energy term and E_2 represents the external energy term. The first term in the internal energy represents the elasticity and the second term represents the curvature. The influence of the two terms is controlled by the parameters α and β, respectively. The external energy (image energy) attracts the Snake to the boundaries of the object in the images. The image energy here will be defined as follows:

$$E_2 = -\left\|\nabla I(x, y)\right\|^2, \tag{8}$$

where I is the image function. Following this, the Snake function will be minimized in the position with high gradient values.

parametric deformable models are used at this stage due to the way that they represent their curves with a set of control points in the same manner that we used to represent the contour curve in the BM model. BM model represents the curve as a set of nodes. This similarity enables us to use point to point mapping to fuse both contours of Snake and BM model as illustrated in the next section.

5. Fusion Using Averaging Technique

Averaging fusion technique is based on establishing one-to-one correspondence between the control points of the Snake and BM model contours [9]. The first step in the averaging fusion technique is to compute the average contour C_{avg} using the following formula:

$$y_i = \frac{1}{M} \sum_{j=1}^{M} x_{ji}. \tag{9}$$

After getting the average curve, at each point on the C_{avg} curve, a normal to the curve is calculated. An efficient method is used to compute the normal given by [19] based on a 2×2 scatter matrix given as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \tag{10}$$

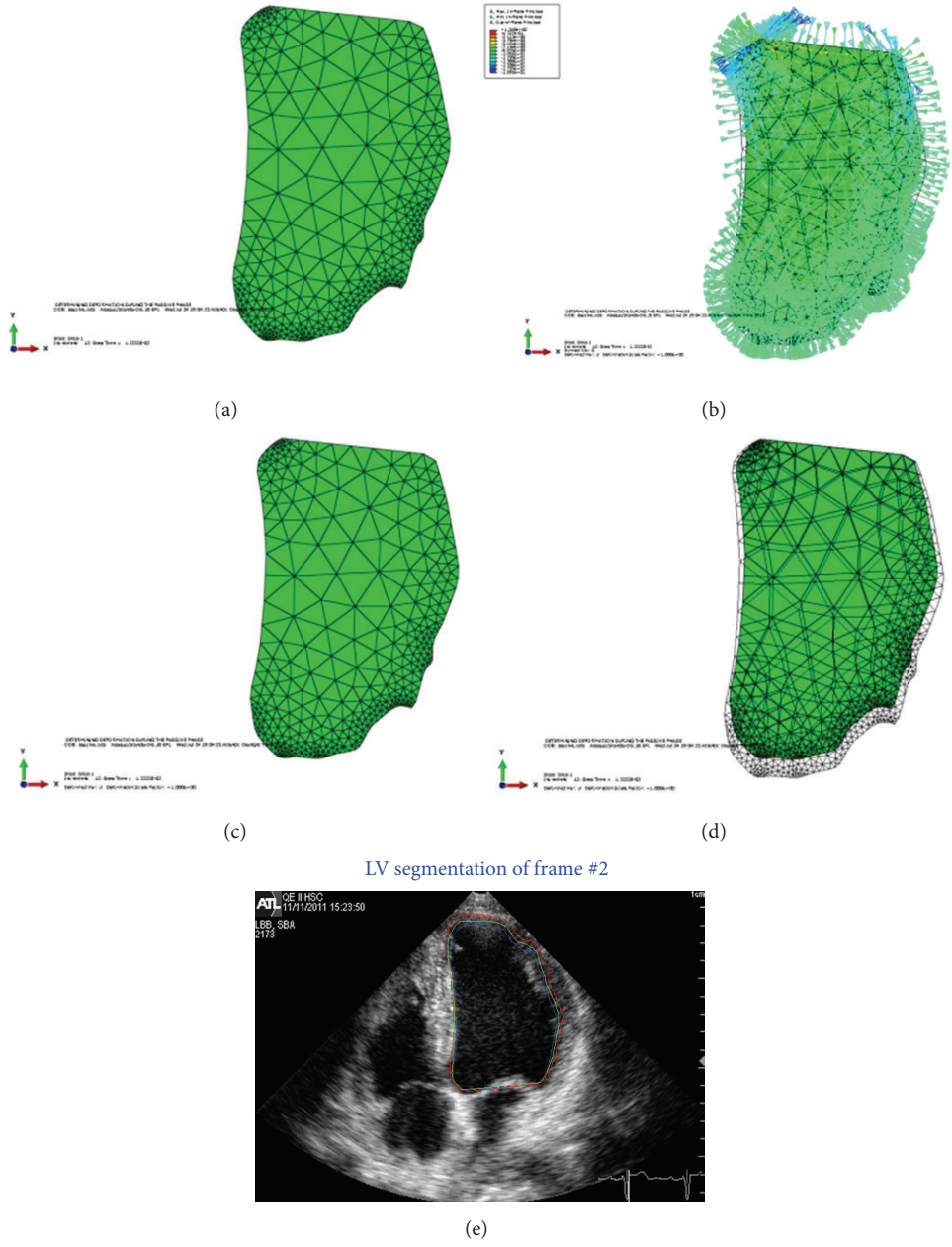


FIGURE 6: (a) Undeformed shape. (b) Load applied to the undeformed shape of LV at contraction phase. (c) Deformed shape. (d) Superposition of the undeformed and deformed shapes of LV. (e) Plotting all contours in the ultrasound image, blue for BM model, red for Snake, and green for the fused contour.

The matrix A is given by following formula:

$$A = \frac{1}{\sum_i \left[(x_i - \overline{x})^2 + (y_i - \overline{y})^2\right]} \times \begin{bmatrix} \sum_i (x_i - \overline{x})^2 & \sum_i (x_i - \overline{x})(y_i - \overline{y}) \\ \sum_i (x_i - \overline{x})(y_i - \overline{y}) & \sum_i (y_i - \overline{y})^2 \end{bmatrix}. \quad (11)$$

First, the eigenvalues of the matrix are calculated and represented as α_M and α_m. After that the orthonormal eigenvectors (a_M, a_m) are calculated using the following formula:

$$a_M = \frac{(a_{12}, \alpha_M - a_{11})}{\sqrt{a_{12}^2 + (\alpha_M - a_{11})^2}},$$
$$a_m = \frac{(a_{11} - \alpha_m, a_{12})}{\sqrt{a_{12}^2 + (\alpha_M - a_{11})^2}}. \quad (12)$$

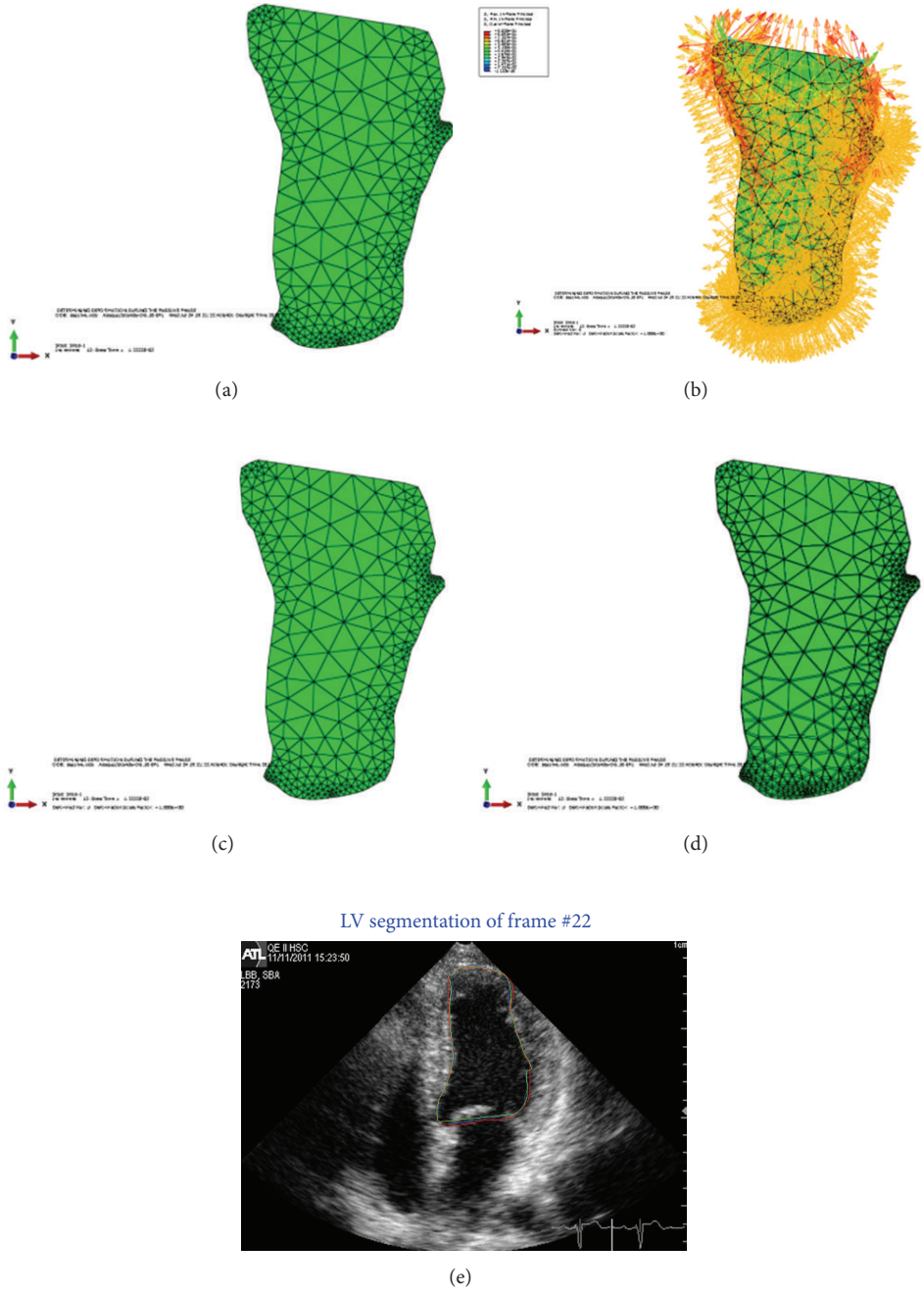


Figure 7: (a) Undeformed shape. (b) Load applied to the undeformed shape of LV at the relaxation phase. (c) Deformed shape. (d) Superposition of the undeformed and deformed shapes of LV. (e) Plotting all contours in the ultrasound image, blue for BM model, red for Snake, and green for the fused contour.

After computing the normal, the intersection between the normal vector with BM and Snake contours will be computed. These intersection points give us a new correspondence between both contours which will be averaged again using (9). This procedure will be iterated until we find that there is no change in the computed averaged points of both contours. Usually the iteration process takes 5 iterations to compute the final averaged contour [9]. Figure 5 illustrates the steps of the averaging technique.

6. Experimental Results

Our implemented framework is used to estimate the deformations of the left ventricle at each frame of the cardiac cycle for the 2D echocardiographic image sequence. The tracking starts at the QRS of the ECG signal which marks the end of diastole and starting of the systole phase. The BM model estimates the LV deformation at each control point (nodes at the boundary), adding these calculated deformations to

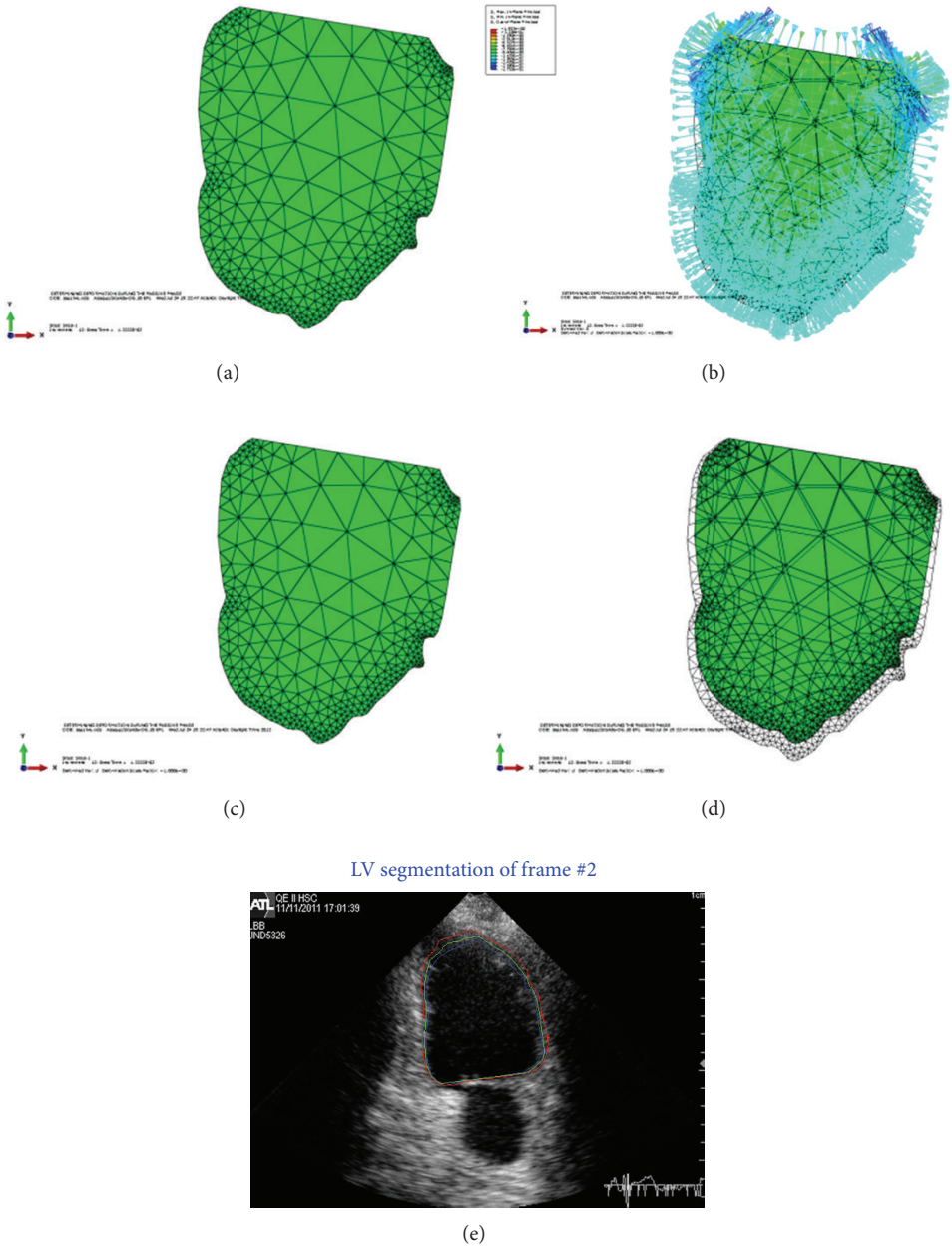


Figure 8: (a) Undeformed shape. (b) Load applied to the undeformed shape of LV at the contraction phase. (c) Deformed shape. (d) Superposition of the undeformed and deformed shapes of LV. (e) Plotting all contours in the ultrasound image, blue for BM model, red for Snake, and green for the fused contour.

the previous contour resulting in the final contour. The FEM will run for a certain time with the computed pressure (load) at each frame of the cardiac cycle. The BM model estimated contour is fused with the Snake contour to get the endocardial contour of the LV. Some experimental results are conducted to verify the framework. In Figure 6, a sample of four-chamber view is used to test the framework. The first image, Figure 6(a), shows the undeformed shape of the LV. The second image, Figure 6(b), shows the pressure applied to this frame. The load is positive in the contraction phase and applied at all the faces of the elements at the boundary of the LV. This positive pressure will let the LV boundaries to contract and the volume of the LV will be less in the deformed shape. No pressure is applied at the base and this surface should be kept fixed to avoid rigid body motion.

Starting from the undeformed shape and applying the pressure to the endocardial SURFACE, the LV will contract and the deformed shape will be as shown in the third image Figure 6(c). The forth image, Figure 6(d) shows superimposing the undeformed and deformed shapes of the LV before and after contraction. As in the real LV, more contraction will occur at the apex and lateral wall while less contraction will occur at the septal. Figure 6(e) shows plotting of the all estimated contours, the BM model, Snake, and the fused

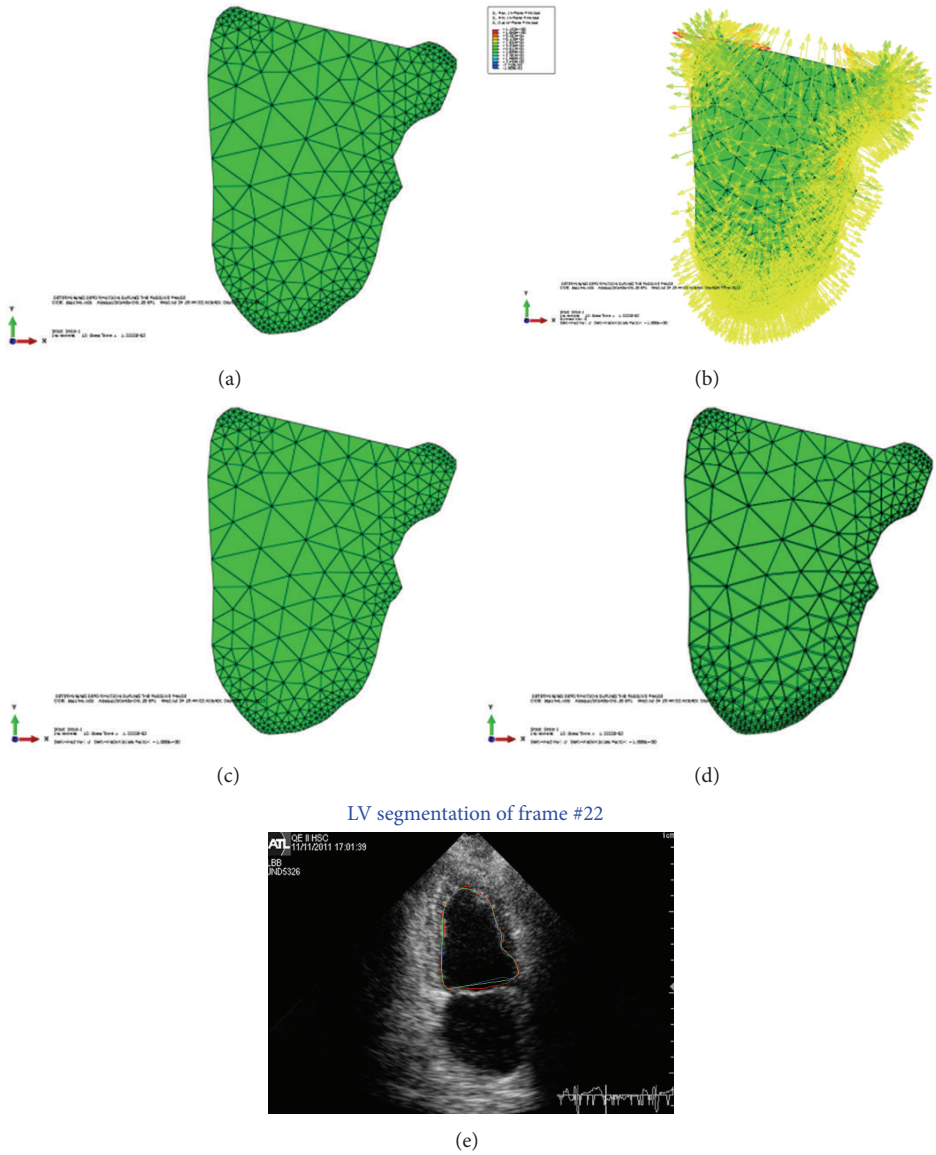


Figure 9: (a) Undeformed shape. (b) Load applied to the undeformed shape of LV at the relaxation phase. (c) Deformed shape. (d) Superposition of the undeformed and deformed shapes of LV. (e) Plotting all contours in the ultrasound image, blue for BM model, red for Snake, and green for the fused contour.

contour in the current frame of 2D echocardiographic image sequence.

Next figure shows the results of applying the framework to a certain image in the passive phase of the cardiac cycle. Figure 7(a) shows the undeformed shape before applying the load.

Figure 7(b) shows the load applied to the undeformed shape of the LV. As shown in the figure, the pressure in this case is negative (arrows are pointing outward) to let the LV boundaries expand to simulate the relaxation of the LV muscles in the passive phase. As mentioned in the contraction results, no pressure was applied at the base; only pressure was applied to the faces of the elements at the LV boundaries. The third image, Figure 7(c), shows the LV shape after applying the load. The undeformed and the deformed shapes are superimposed together to show the amount of deformation that happened after applying the load as shown in Figure 7(d). In the last image of Figure 7(e), the LV computed contours are plotted in the 2D echocardiographic image. Two-chamber view results are given in the next two figures to show the performance of the framework. Figure 8 shows the estimation of the LV contour in the active phase while Figure 9 shows the results in passive phase.

Figure 10 shows the segmentation of LV area from the sequence of two-chamber view. The sequence starts from the QRS signal that marks the end of diastole stage and beginning of the systole stage. 20 frames are used to show the deformations of the LV and its area at each frame and the robustness of the framework to extract the exact area of the LV.

Figure 10: Segmentation of 20 frames of two-chamber view starting from the end of diastole to the end of systole.

7. Evaluation of the Results

The results of the framework are evaluated by comparing them to a gold standard created from three manually plotted contours traced by three cardiologists. The procedure of creating the gold standard is given in [9] is done by taking the average of the three curves after establishing the correspondence between the points in each curve. Ten samples are used to create the gold standard. Average perpendicular distance (APD) is used as an error metric to measure the closeness of the estimated contours to the gold standard. Figure 11 shows samples of comparing the BM model, Snake, and the fused computed contours with the created gold standard contour. Figure 11(a) shows the plotting of the BM estimated contour versus the gold standard of the same frame which is at the end of diastole stage. The computed contour matches the gold standard without leaking outside the boundaries at parts of the contour that has signal drop out. Figure 11(b) shows the plotting of the Snake on the same frame versus the gold standard. The computed contour has leakage outside the real boundaries of the LV due to the signal drop out. Due to the open contour, Snake does not find the edges at that part of the contour. Unlike the Snake, BM does not depend on the ultrasound image to estimate the contour. As shown in Figure 11(a), this shortcoming of the ultrasound can be avoided by using BM model. The fusing of the BM and Snake contours is shown in Figure 11(c). The fused contour shows more closeness to the gold standard than both BM and Snake contours. Also, more smoothness with rejecting the overshoots has been done by fusing both contours.

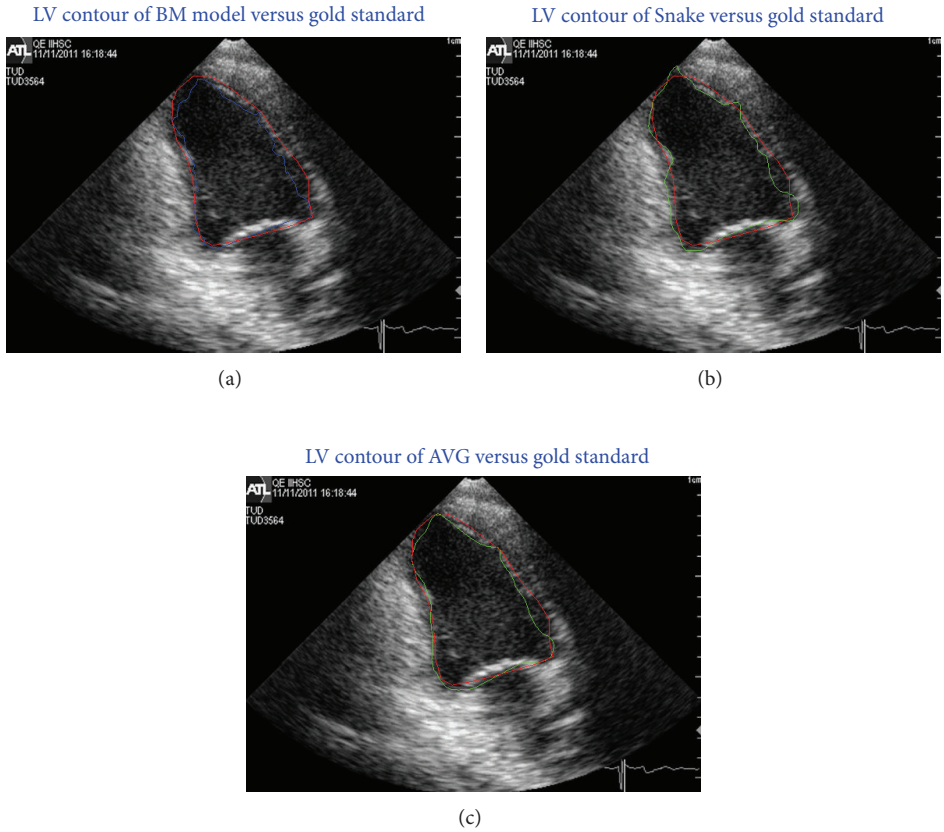


Figure 11: (a) Plotting the BM contour (blue line) versus gold standard contour (red line), (b) Snake contour (green line) versus gold standard contour (red line), and (c) AVG (green line) contour versus their gold standard (red line) at the end of the diastole stage.

In the same manner, Figure 12 shows the comparison of the three contours with the gold standard at the end of the systole stage.

In Figure 13, the computed area and ejection fraction (EF) values are compared to the gold standard.

8. Discussion and Analysis of the Results

By incorporating BM model in the averaging fusion technique framework and from the experimental results and the evaluation, the framework achieves high robustness and stability overall the samples during the cardiac cycle. The computed contours show a high closeness to the gold standard. BM model plays a dominant rule in the framework by providing the concrete base that the framework stands on during the cardiac cycle. The BM model works independently from the ultrasound images and can provide accurate detection to the LV boundaries where the deformable models fail to do so. BM model keeps the deformable models inside the ROI by overcoming the difficulties of twisting and rotation of the LV and preventing the deformable models from leaking outside the ROI when there are missing parts of LV boundaries or signal drop out.

By employing averaging fusion techniques, we ensure that the fused contour is close to the boundaries of the LV by removing the outliers in the deformable models and modifying the BM model contour. The fused contour and the dynamic model provide robustness and accurate starting point by initializing the current frame with closeness contour to the desired one. This accurate initialization ensures the quality of the contour detection and reduces the required running time for the deformable models by reducing the number of iterations that the deformable models required.

The constitutive equations of the BM model resemble and simulate the intact heart. From the experimental results and the evaluation, the BM model and the framework overall have accurate detection of the contour, area and ejection fraction for the cases in which the heart is normal or suffers from fewer complications and abnormality. In the cases where the patient has severe heart abnormality, the heart of patient does not contract as the normal (dysfunction case) and the stroke volume is less than normal value too. In this situation, the BM contour will not match exactly the real contour and it will affect the accuracy of the fused contour.

From the APD data and the statistical analysis, the fused contour scores the highest accurate results to be the closet contour to the gold standard with 1.313 ± 0.0206 mm as shown in Table 1. Also, the fused contour has the outstanding results in computing the area and ejection fraction values. In calculating the ejection fraction, all the computed values are located in acceptance range of the Bland and Altman plot as shown in Figure 13.

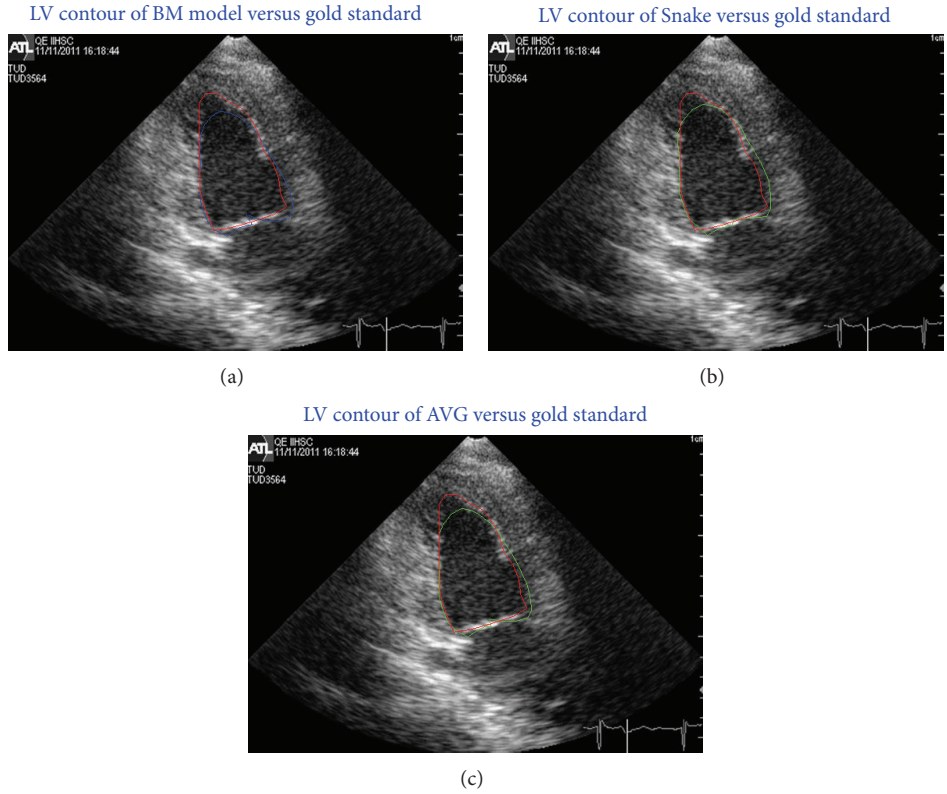


FIGURE 12: (a) Plotting the BM contour (blue line) versus gold standard contour (red line), (b) Snake contour (green line) versus gold standard contour (red line), and (c) AVG (green line) contour versus gold standard (red line) at the end of the systole stage.

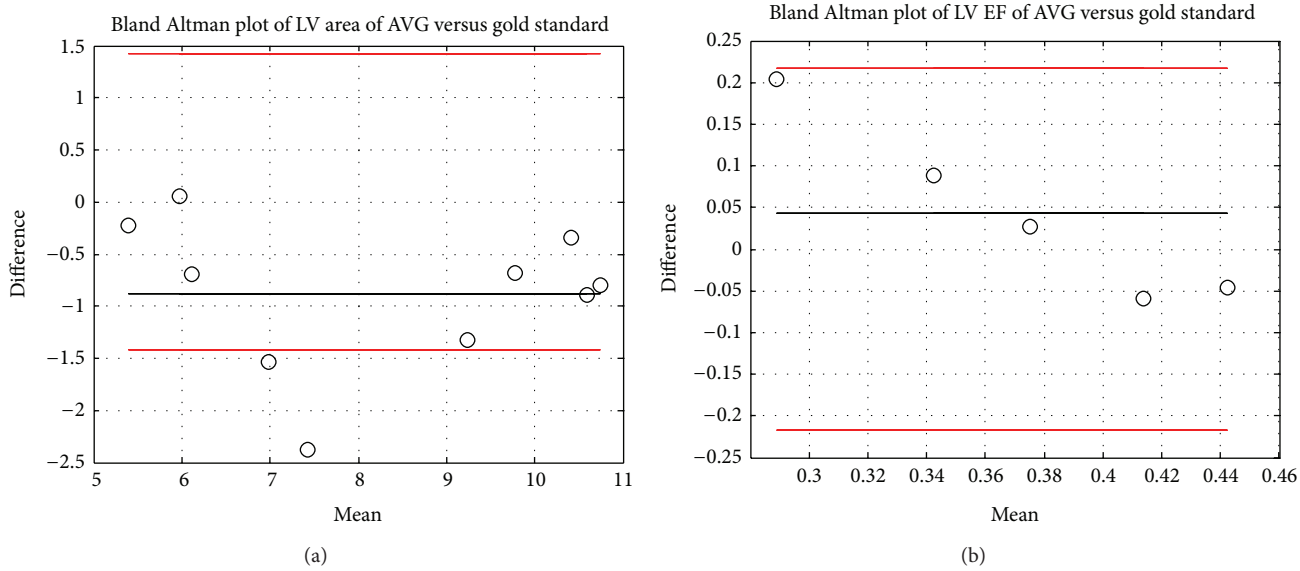


FIGURE 13: Bland Altman plot for LV computed (a) area and (b) EF.

9. Future Work

Based on the LV tracking performance in 2D echocardiography image sequence, the framework will be extended for 3D echocardiography using 3D BM and deformable models.

10. Conclusion

A novel approach to integrate a left ventricular BM model within the framework of FEM has been presented and fused the estimated contour with a parametric deformable model.

Table 1: Computing the mean of the measured APDs.

Method	Mean (mm)
BM model	1.566
Snake	1.515
AVG	1.313

LV deformations through 2D echocardiographic image sequence are tracked for the whole cardiac cycle by this framework to tackle the challenges and difficulties of the ultrasound images and the heart. The BM model uses the constitutive equations of both the passive and active strain energy functions to simulate the myocardial tissue movement. Nonlinear deformations are estimated by solving the constitutive equations using ABAQUS FEA. Averaging fusion technique is used to fuse the BM and the deformable model contours. The experimental results are conducted and evaluated with a created gold standard. According to the experimental results and the evaluation, this approach shows a tremendous potential to track the LV endocardium during the cardiac cycle and tackle the difficulties and challenges due to 2D echocardiography and the heart.

Conflict of Interests

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

References

[1] S. Chan and G. Sainarayanan, "Fuzzy-based boundary enhancement for echocardiogram using local image characteristics," *Malaysian Journal of Computer Science*, vol. 19, no. 2, pp. 151–159, 2006.

[2] H. Ketout, J. Gu, and G. Horne, "Improved dempster and shafer theory to fuse fuzzy inference system, neural networks and CNN endocardial edge detection," *The Mediterranean Journal of Electronics and Communications*, vol. 7, no. 3, pp. 267–275, 2011.

[3] G. Sainarayanan, N. Mal Murugan, and S. Chan, "A novel method for echocardiogram boundary detection using adaptive neuro-fuzzy systems," in *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA '07)*, vol. 3, pp. 415–419, December 2007.

[4] J. G. Bosch, S. C. Mitchell, B. P. F. Lelieveldt et al., "Automatic segmentation of echocardiographic sequences by active appearance motion models," *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1374–1383, 2002.

[5] L. F. Zhang and E. A. Geiser, "An effective algorithm for extracting serial endocardial borders from 2-dimensional echocardiograms," *IEEE Transactions on Biomedical Engineering*, vol. 31, no. 6, pp. 441–447, 1984.

[6] P. M. F. Nielsen, I. J. Le Grice, B. H. Smaill, and P. J. Hunter, "Mathematical model of geometry and fibrous structure of the heart," *American Journal of Physiology: Heart and Circulatory Physiology*, vol. 260, no. 4, pp. H1365–H1378, 1991.

[7] M. P. Nash and P. J. Hunter, "Computational mechanics of the heart," *Journal of Elasticity*, vol. 61, no. 1–3, pp. 113–141, 2000.

[8] D. H. S. Lin and F. C. P. Yin, "A multiaxial constitutive law for mammalian left ventricular myocardium in steady-state barium contracture or tetanus," *Journal of Biomechanical Engineering*, vol. 120, no. 4, pp. 504–517, 1998.

[9] V. Chalana and Y. Kim, "A methodology for evaluation of boundary detection algorithms on medical images," *IEEE Transactions on Medical Imaging*, vol. 16, no. 5, pp. 642–652, 1997.

[10] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: a new variational formulation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 430–436, June 2005.

[11] A. Blake and M. Isard, *Active Contours*, 1st edition, 2000.

[12] A. Blake, M. Isard, and D. Reynard, "Learning to track the visual motion of contours," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 179–212, 1995.

[13] F. Dorri, P. F. Niederer, and P. P. Lunkenheimer, "A finite element model of the human left ventricular systole," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 9, no. 5, pp. 319–341, 2006.

[14] Dassault Systèmes Simulia Corp., Providence, RI, USA, Abaqus User Subroutines Reference Manual, 2010.

[15] Dassault Systèmes Simulia Corp., Providence, RI, USA, ABAQUS THEORY MANUAL, 2010.

[16] Dassault Systèmes Simulia Corp., Providence, RI, USA, ABAQUS Scripting User's Manual.

[17] Dassault Systèmes Simulia Corp., Providence, RI, USA, ABAQUS Analysis User's Manual, Vol. 1 : Introduction, Spatial Modeling, Execution and Output.

[18] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[19] I. M. Anderson and J. C. Bezdek, "Curvature and tangential deflection of discrete arcs: a theory based on the commutator of scatter matrix Pairs and its application to vertex detection in planar shape data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 27–40, 1984.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 276589, 8 pages
http://dx.doi.org/10.1155/2014/276589

Research Article

A Machine Learning Approach for Specification of Spinal Cord Injuries Using Fractional Anisotropy Values Obtained from Diffusion Tensor Images

Bunheang Tay, Jung Keun Hyun, and Sejong Oh

Department of Nanobiomedical Science, Dankook University, Cheonan 330-714, Republic of Korea

Correspondence should be addressed to Sejong Oh; sejongoh@dankook.ac.kr

Received 30 July 2013; Revised 5 November 2013; Accepted 27 November 2013; Published 21 January 2014

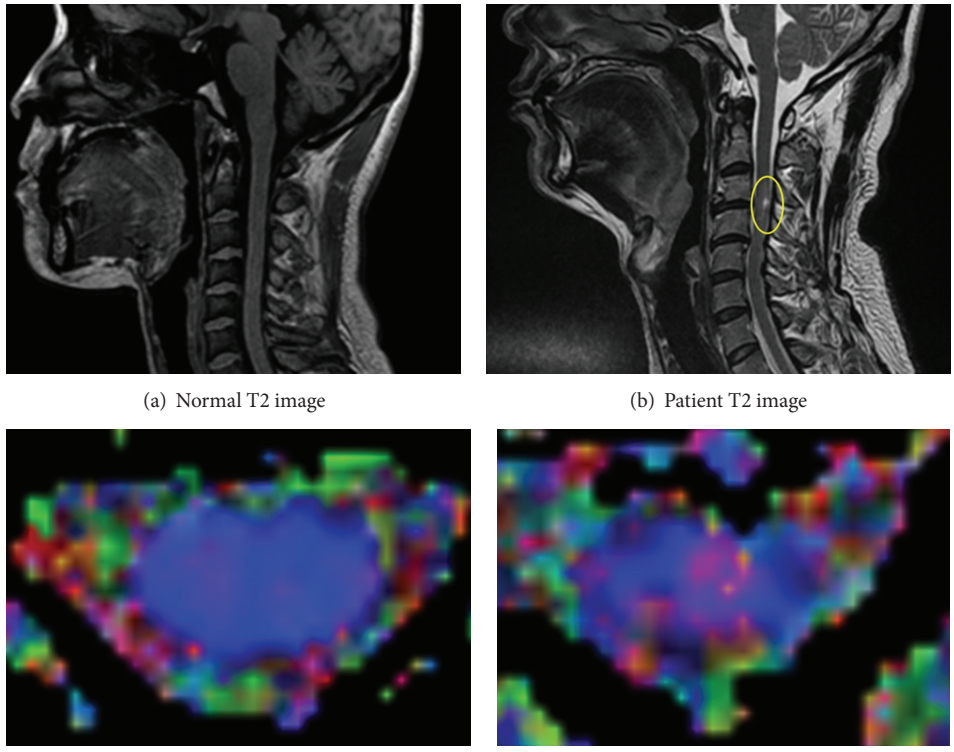
Academic Editor: Rong Chen

Diffusion Tensor Imaging (DTI) uses in vivo images that describe extracellular structures by measuring the diffusion of water molecules. These images capture axonal movement and orientation using echo-planar imaging and provide critical information for evaluating lesions and structural damage in the central nervous system. This information can be used for prediction of Spinal Cord Injuries (SCIs) and for assessment of patients who are recovering from such injuries. In this paper, we propose a classification scheme for identifying healthy individuals and patients. In the proposed scheme, a dataset is first constructed from DTI images, after which the constructed dataset undergoes feature selection and classification. The experiment results show that the proposed scheme aids in the diagnosis of SCIs.

1. Introduction

The Spinal Cord (SC) is a major pathway for motor and sensory signals traveling between the brain and the peripheral nervous system. The SC, along with the brain, comprises the central nervous system. It is tubular in shape and contains white matter (spinal tracks) and gray matter (neuronal cell bodies). When a Spinal Cord Injury (SCI) occurs, the spinal tracks, which convey sensory, motor, and autonomic signals between the brain and organs, are disrupted. An SCI may cause patients to become paralyzed or stop organs from functioning properly. The International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI) motor and sensory scores enable us to assess SCI patients more precisely. These scores were developed by the American Spinal Injury Association (ASIA) in 1982 and updated several times [1–3]. They are correlated with functional status, and they are essential to arriving at a prognosis for SCI patients in clinical rehabilitation units [4]. However, clinical assessment based on ISNCSCI scores has limitations for accurate diagnosis of SCIs. It includes unclear information when there are concomitant injuries in other organs, and it is also somewhat subjective as it relies on information relayed by the patient.

Conventional Magnetic Resonance Imaging (MRI) is widely used to diagnose SCI. MRI is a medical imaging technology that produces high-quality images of organs and tissues at the macroscopic level. It utilizes a black and white contrast image to differentiate between soft and hard tissues. Diffusion Tensor Imaging (DTI) is an advanced technology that utilizes echo-planar images obtained from MRI. It maps the diffusion of water molecules in the brain and SC tissue according to their tissue structure and architecture (Figure 1). DTI is used in the study of diseases and pathological conditions such as multiple sclerosis, brain trauma, brain tumors, and hypertensive encephalopathy. It provides numerical information about the magnitude and orientation of each individual tissue in a three-dimensional (3D) space. This is termed as diffusion anisotropy. The diffusion tensor consists of many diffusion ellipsoids floating in space. The orientation of each diffusion ellipsoid is described via a set of orientation vectors, also known as eigenvectors. When an eigenvector changes either its length or direction, it produces a different result corresponding to its eigenvalue. In DTI, diffusion anisotropy can be expressed as Fractional Anisotropy (FA). FA is sensitive to the number of fibers with directionality within each voxel and is widely used to measure fiber integrity



(a) Normal T2 image (b) Patient T2 image

(c) Normal FA colored map (d) Patient FA colored map

Figure 1: Spinal cord image. In (b), yellow circle shows injured area. In (c) and (d), the color of each pixel means direction of water molecules in SC tissues. In the patient FA map, the direction is irregular because it includes disconnected part of nerves and water molecules cannot flow straight way.

with a range from 0 to 1. The FA value indicates the degree of water diffusion anisotropic motion, with a higher FA value indicating a higher degree, as represented by the following equations:

$$\mathrm{FA} = \sqrt{\frac{3}{2}} \frac{\sqrt{\left(\lambda_1 - \overline{\lambda}\right)^2 + \left(\lambda_2 - \overline{\lambda}\right)^2 + \left(\lambda_3 - \overline{\lambda}\right)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}, \quad (1)$$

$$\overline{\lambda} = \frac{(\lambda_1 + \lambda_2 + \lambda_3)}{3}.$$

In the equations, the corresponding eigenvalues λ give the magnitude of the peak in that direction.

In this paper, we propose a novel SC analysis scheme that uses a machine learning technique. Our goal is to provide support for diagnosis of SCI by human experts. At present, human experts analyze DTI images and FA values and then make decisions based on their experience. We believe that if we can provide them with more objective information, they will be able to make more accurate diagnoses. In machine learning, classification schemes are used for prediction or diagnosis. However, training data is required for classification tasks. Therefore, in our scheme, we construct a training dataset using four FA values from DTI slice images taken from patients and normal controls. We then expand the base dataset to a higher dimensional dataset comprising 15 features and abstract the target dataset using a feature selection algorithm to improve classification accuracy. The resulting dataset has a prediction accuracy of more than 90%. Our contributions are (1) a training dataset from the DTI image construction process and (2) application of a classification scheme to the prediction of SCIs. The size of the DTI data associated with an individual is over 200 MB. Raw DTI data is large and difficult to use in computer-aided diagnosis. In our scheme, we extract useful numeric data from the raw DTI data and then use it for diagnosis. Our approach can be applied in any area in which DTI is utilized for diagnosis.

2. Related Work

T1- and T2-weighted imaging (Figure 1) have been performed in sagittal and axial panels in order to evaluate SCI individually. The neurological level of injury and the severity were determined by clinical assessments, and conventional MRI was used to detect the level of signal changes within the injured spinal cord. The signal change level was correlated with clinical findings [5]. We used a classification scheme to automatically identify the image slices of injured patients. The scheme provides speedy and accurate results, and it works effortlessly with pattern recognition mechanisms. Classification accuracy is a significant measure of the quality of a prediction scheme. Many researchers have attempted to increase classification accuracy by manipulating the dataset

```
k-Nearest Neighbor
Classify (X, Y, x) // X: training data, Y: class labels of X, x: unknown sample
for i = 1 to m do
  Compute distance d(X_i, x)
end for
Compute set I containing indices for the k smallest distances d(X_i, x).
return majority label for {Y_i where i ∈ I}
```

Algorithm 1: Pseudocode for KNN classification.

or by improving the classification algorithms. As stated in Section 1, our objective is to abstract FA values from DTI images and thereby aid human experts in ascertaining which part is injured. Even though an FA value is a reliable measure for the specification of SCI, human experts rely heavily on their individual experiences gained from previous analyses of T1- or T2-weighted images. We believe that an automated method that can pinpoint the injured part and provide related information would be invaluable for SCI diagnosis.

Machine learning covers a wide ranging area and mostly uses computation models in a variety of real-world domains. In machine learning, the focus is on the design and development of algorithms that can make decisions as humans do, based on information from a database. It is used in tasks such as recognition, diagnosis, planning, robot control, and prediction. Machine learning can also be used to predict tissue toxicity [6] and to examine neuroimaging data [7]. Both tasks use pattern recognition, which involves the detection of multiple variables of interest. Nowadays, everything we do involves massive amounts of data that must be managed, analyzed, and utilized by researchers. Machine learning can be used to extract important relationships and correlations that may be hidden within large piles of data. It improves the efficiency of systems and the design of machines. In recent years, image analysis has become a major area of application for machine learning in computer-aided diagnosis, medical image analysis, and lesion segmentation.

Classification is a key theme in biological studies, and machine learning plays an important role in the classification process. Machine learning is used to predict unknown sample data among an already known dataset. It can discriminate between two or more contrary objects, group together similar objects, or separate different objects. In the classification process, the specific properties of objects are used to categorize them, and each object is assigned a class label (e.g., "patient," "normal," etc.) that describes the particular group to which it belongs. However, classification mainly operates by making predictions based on known sample data that has been learned from training data. In classification, term training and testing are used to predict the unknown data's labels. We train the classifier with known sample data in a training dataset and check its performance by examining the test dataset, which consists of the unknown sample used to predict its class label. K-Nearest Neighbor (KNN) [8] and Support Vector Machine (SVM) [9, 10] are well-known classification algorithms. KNN is an instance-based learning classifier that performs classification based on the closest data point in feature space. The KNN algorithm is outlined in Algorithm 1.

Feature selection preserves classification accuracy and reduces the number of features that are irrelevant and redundant. It is a statistical machine learning technique used to build a strong and stable predictor and is often viewed as a preprocessing step for classification. Feature selection is not only used to handle noisy data but also to cope with very large datasets. Due to its dimensional reduction of the data, feature selection enables a classification algorithm to be faster and more effective. Feature seLection (FeaLect) [11], Feature Selection based on a Distance Discriminant (FSDD) [12], ReliefF [13], Clearness-Based Feature Selection (CBFS) [14], and R-value-based Feature Selection (RFS) [15] are well-known feature selection algorithms. RFS is based on the R-value [16], which is a measure used to capture the congestion area among classes in a feature. The basis of the FSDD algorithm is to identify the features that result in good class separability between classes and ensure that samples in the same classes are as close as possible. ReliefF is regarded as one of the most successful feature selection algorithms. The basic idea underlying ReliefF is to iteratively estimate feature weights according to their ability to discriminate between neighboring instances. CBFS is a very efficient feature selection algorithm based on "CScore" measures. CScore calculates the degree of samples located in the correct class region. Based on the Lasso, FeaLect introduces an alternative algorithm for feature selection. This algorithm measures the "quality" of each individual feature by defining a scoring scheme. Training data is used to generate several samples, select high relevance-ordering of features for each sample, and finally combine the highly relevant features in each relevance ordering together. In this study, we mainly used feature selection to evaluate the discriminative power of each feature and to find the most accurate feature subset.

3. Materials and Methods

Figure 2 summarizes our proposed approach. The left side of Figure 2 illustrates the training dataset (target dataset) production process, while the right side shows the production of the test data in which it is not known whether the subject is a normal subject or a patient. After building the training dataset, we perform prediction analysis using the test data. The steps used to prepare the test data are the same as those used to build the training dataset. To build the training dataset, we acquired DTI image data

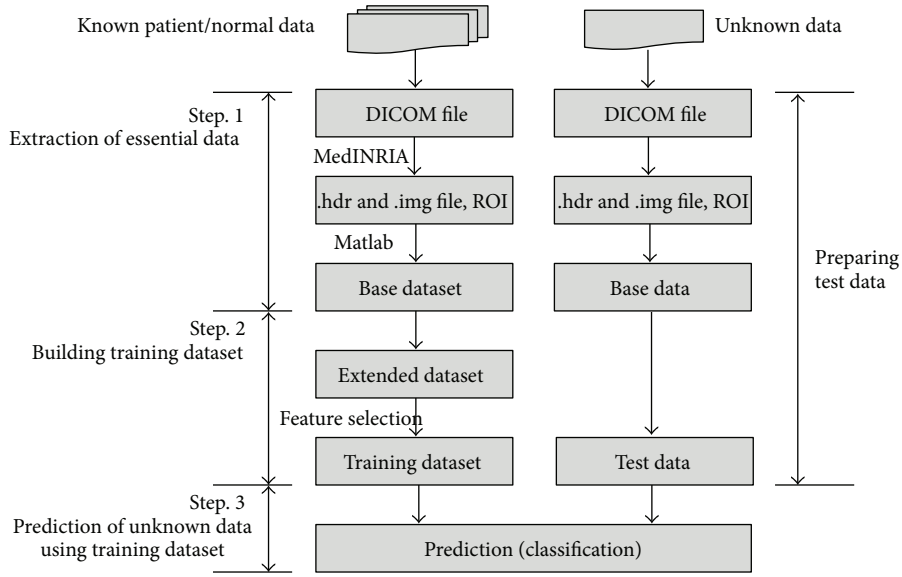


Figure 2: Preview of the classification analysis procedure for the new dataset.

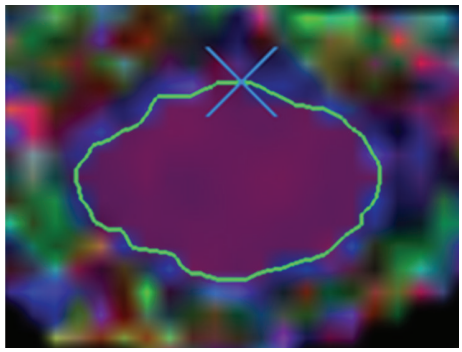


Figure 3: Example of ROI in colored FA.

from the Dankook University Hospital (Cheonan, Republic of Korea). The SCI was then clinically evaluated by applying ISNCSCI standards [17]. The DTI raw data was in the Digital Imaging and Communications in Medicine (DICOM) format. We extracted FA values from the DICOM file using MedINRIA Ver. 1.8.0 (available free at http://www-sop.inria.fr/asclepios/software/MedINRIA/) and Matlab Ver. 2007a (http://www.mathworks.com/products/matlab/), and we used these values to build a base dataset with four features. We then extended the base dataset and processed the target dataset through feature selection. Finally, we conducted classification analysis.

3.1. Extraction of Essential Data. The clinical data acquired from Dankook University Hospital was from 14 individuals (9 patients and 5 normal controls). The nine patients had SC problems near the neck (see Figure 3). To generate one DTI file, an aggregate image of approximately 26 to 28 slices was captured from the midbrain to SC level T1 or T2. For simplicity, we eliminated the slices above the C1-2 level and below the C7 level that may cause some artifacts. We placed the rest of the slices into one of two categories: "injured" or "normal." T1 or T2-weighted sagittal images were then used to check the level of those exhibiting abnormal signal intensities within the cervical SC. We matched the axial images with the sagittal section to check which axial slices were abnormal. These axial images, which were obtained with conventional T1 or T2 MRI, were also compared to the axial section in Echo Planar Imaging (EPI). The ANALYZE image (.hdr and .img files) and Region of Interest (ROI) were generated using the dedicated software MedINRIA. To select the ROI, we paid special attention to the void region of gray matter, blank areas, motion artifacts, and CerebroSpinal Fluid (CSF) partial volume effects. ROIs were defined in every slice in the axial plane (Figure 3).

We took the diffusion tensor measure (FA) from DTI and Fiber Tracking [18]. The gradient table must also be validated with the DTI tool, which describes magnetic field information to render the diffusion image in a particular direction. We imported the ROIs into the Matlab program as a binary mask to filter the SC regions using the Neuroimaging Informatics Technology Initiative (NIfTI) and ANALYZE image tools [19]. The imported ROIs were then rotated by 90° to generate the DTI. The ROIs were maintained at the same coordinates and direction.

After the SC region was filtered, we subdivided the selected SC region into four different subregions: posterior, anterior, left, and right. We then applied the result individually on each axial plane and computed the subdivision as the cumulative intensity along the X- and Y-axes. The center point on the spine area (ROI as mask) was chosen in both the X and Y directions. To create four separated segments, we rotated the line joining the X and Y center points by 45°; the intersection and ROI were considered as bounding lines to separate the regions into anterior, posterior, left, and right subregions (Figure 4) [20]. In the division process, we balanced each voxel in each region on the central point.

Table 1: Example of a base dataset.

Class label	Posterior	Anterior	Left	Right
0	0.37388	0.331959	0.401053	0.511554
0	0.071062	0.336189	0.101144	0.434287
0	0.154871	0.532321	0.338796	0.30359
0	0.51665	0.710171	0.475486	0.482532
⋮	⋮	⋮	⋮	⋮
1	0.728015	0.553007	0.673909	0.484319
1	0.569538	0.538663	0.715141	0.427902
1	0.720117	0.527224	0.642175	0.535785
1	0.617641	0.482621	0.573918	0.424463
⋮	⋮	⋮	⋮	⋮

Figure 4: Divided spinal cord regions (red: posterior, green: anterior, blue: left, yellow: right).

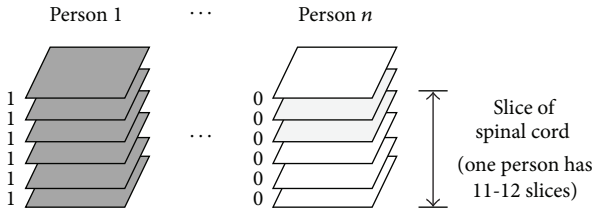


Figure 5: Class labeling for each slice image (gray: injured slice, white: normal slice).

For the features of the base dataset, we derived the mean value of each region in every slice. This is a simple way to acquire SC information from the FA value; it is clear, fast, and distinguishable between normal and abnormal slices. In the final step, we had four features of each slice that described the condition of the SC through numerical data. To prepare for classification analysis, we labeled each slice, which had been examined by experts at Dankook University Hospital, as either class 0 (patient) or class 1 (normal control) (Figure 5). This provided our base dataset. Table 1 presents an example of a base dataset. Our dataset had 164 data samples comprising 106 injured and 58 normal slices.

3.2. Building the Target Training Dataset. In this section, we outline the method used to expand the base dataset to a target training dataset in order to obtain high classification accuracy. This approach is useful due to the maximization of the base dataset potential. We have the original four features: posterior (P), anterior (A), left (L), and right (R). Hence, the feature set S^B of the base dataset is

$$S^B = \{P, A, L, R\}. \tag{2}$$

Based on S^B, we expand the base dataset to feature set S^T of the target dataset by taking all combinations of the four features. S^T is as follows:

$$\begin{aligned} S^T = \{&P, A, L, R, (P, A), (P, L), (P, R), (A, L), (A, R), \\ &(L, R), (P, A, L), (A, L, R), (P, A, R), \\ &(P, L, R), (P, A, L, R)\}. \end{aligned} \tag{3}$$

The feature value of (P, A) is the average of P and A, and the feature value of (P, A, L) is the average of P, A, and L. The total number of combinations from n features is $2^n - 1$. Therefore, the number of features in S^T is given by

$$n\left(S^T\right) = 2^n - 1 = 2^4 - 1 = 15. \tag{4}$$

The first four features are the original features from the base dataset (P, A, L, R), while the remaining features are generated from the average values of the base dataset features.

To evaluate the quality of the new features, we performed t-tests for each feature in S^T. The hypothesis H_0 states that "the FA value of a normal slice is the same as the FA value of an injured slice." The p value of t-test for each feature is presented in Table 2. All the p values are lower than 0.05, which indicates that hypothesis H_0 should be rejected. In other words, the FA value of a normal slice is significantly different from the FA value of a patient slice in every feature. It also implies that the 15 given features are sufficient for classification analysis.

The extended dataset has 15 features to which we can apply the feature selection scheme to improve classification accuracy. Various feature selection schemes, with a varying number of features needed for optimal classification accuracy, are available for use. In our experiment, we tested four feature selection algorithms: FSDD, ReliefF, CBFS, and RFS. Each feature selection algorithm has a feature evaluation function. After evaluating each feature in the dataset, we chose n features that had the best score. The next step was classification analysis using these n features.

3.3. Classification Analysis. To acquire the most accurate results from the extended dataset, we evaluated the quality of each feature using various feature selection algorithms. We then used a set of high-qualified features for classification analysis. The KNN and SVM supervised learning algorithms were used as classifiers for the target dataset.

4. Results and Discussion

4.1. Classification of Patients and Normal, Healthy Individuals. To classify patients and normal, healthy controls, we modified the extended dataset because a single individual has 14-15 slice data points and direct comparison with other individuals is

Table 2: Results of t-tests for each feature in the target dataset.

	Feature	p value = P $(T \leq t)$	AVG (patient)	AVG (normal)
1	P	$7.69 * 10^{-9}$	0.427	0.535
2	A	$8.05 * 10^{-31}$	0.451	0.684
3	L	$1.89 * 10^{-20}$	0.475	0.645
4	R	$2.39 * 10^{-26}$	0.419	0.610
5	(P, A)	$3.03 * 10^{-23}$	0.439	0.609
6	(P, L)	$8.37 * 10^{-18}$	0.451	0.590
7	(P, R)	$5.61 * 10^{-19}$	0.423	0.573
8	(A, L)	$8.99 * 10^{-31}$	0.463	0.665
9	(A, R)	$2.74 * 10^{-33}$	0.435	0.647
10	(L, R)	$4.35 * 10^{-31}$	0.447	0.628
11	(P, A, L)	$4.95 * 10^{-26}$	0.451	0.621
12	(A, L, R)	$2.05 * 10^{-34}$	0.448	0.647
13	(P, A, R)	$8.57 * 10^{-27}$	0.432	0.610
14	(P, L, R)	$3.3 * 10^{-24}$	0.440	0.597
15	(P, A, L, R)	$4.62 * 10^{-29}$	0.443	0.619

H_0: the FA value of a normal slice is the same as the FA value of a patient slice; AVG: average of FA value.

Table 3: Classification results of patients and normal individuals.

	Diagnosis result	1-NN result using LOOCV
Person 1	Patient	Patient
Person 2	Patient	Patient
Person 3	Patient	Patient
Person 4	Patient	Patient
Person 5	Patient	Patient
Person 6	Patient	Patient
Person 7	Patient	Patient
Person 8	Patient	Patient
Person 9	Patient	Patient
Person 10	Normal	Normal
Person 11	Normal	Normal
Person 12	Normal	Normal
Person 13	Normal	Normal
Person 14	Normal	Normal

difficult. The average FA values of entire slices were calculated for each individual and a modified dataset was constructed. We used the KNN classifier with $k = 1$ and performed the Leave-One-Out Cross-Validation (LOOCV) test [21] because the data sample size of the modified dataset was small. Table 3 presents the classification results. These results show 100% classification accuracy, indicating that the proposed dataset is useful in the classification of normal, healthy individuals and SCI patients.

4.2. Classification of Normal and Patient Slices. The test outlined in this section was performed to predict which part of the SC is injured. This test is more difficult than the diagnosis of injured patients in Section 4.1. To classify normal individuals and patients' slices, we used an entire extended dataset and the KNN classifier. We tested the influence of feature selection using RFS, FSDD, ReliefF, and CBFS. From the experimental results, we found that RFS selects the best features that produce high classification accuracy compared with the other feature selection algorithms. Figure 6 shows the classification accuracy for the entire feature list. The

Table 4: Results of feature evaluation using the RFS algorithms.

Ranking	Index of feature	Feature	R-value
1	11	(A, L, R)	0.060976
2	8	(A, R)	0.067073
3	9	(L, R)	0.085366
4	14	(P, A, L, R)	0.085366
5	13	(P, L, R)	0.085366
6	12	(P, A, R)	0.085366
7	7	(A, L)	0.091463
8	3	R	0.091463
9	1	A	0.097561
10	2	L	0.103659
11	10	(P, A, L)	0.109756
12	4	(P, A)	0.115854
13	6	(P, R)	0.146341
14	5	(P, L)	0.146341
15	0	P	0.158537

Table 5: The best accuracy of the base dataset, extended dataset, and feature selection applied dataset using KNN and SVM.

	Base dataset (4 features)	Extended dataset (15 features)	FS applied dataset (n feature)
KNN	0.901	0.914	0.938 ($n = 2$)
SVM	0.932	0.927	0.933 ($n = 8$)

Table 6: The experimental variables for the best accuracies in Table 3.

	Base dataset (user-defined value)	Modified dataset (user-defined value)	Best FS dataset (FS/no. of feature/user-defined value)
KNN	$k = 15$	$k = 3$	RFS/2/$k = 5$
SVM	Radial basis function	Linear	CBFS (FSDD)/8/linear kernel

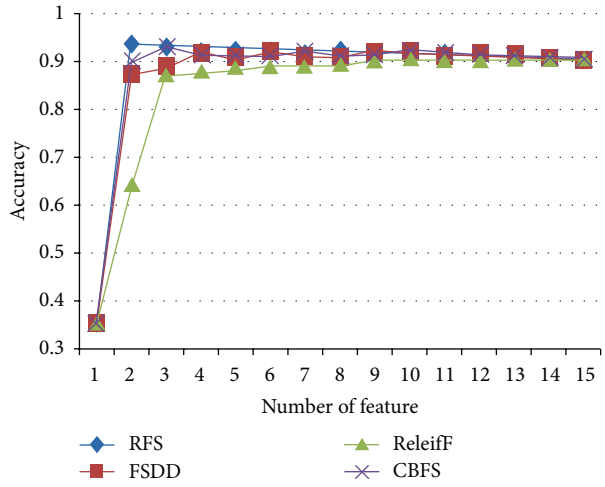


Figure 6: The graph of the accuracy by each feature using KNN algorithms with $k = 5$.

classification accuracy is obtained by applying the four feature selection algorithms with KNN and SVM classifications. In KNN, the choice of k affects the performance of the algorithm. We set $k = 5$ for KNN and used threefold cross-validation to avoid overfitting problems. In Figure 6, the number of features refers to the number of selected features for the classification test. For example, if the number of features is six, then the best six features evaluated by the four feature selection algorithms are used for the classification test. It can be seen that two features from the RFS algorithm give a superior accuracy of 93.8%. The accuracy measurements of FSDD, ReliefF, and CBFS (87.7%, 64.2%, and 90.1%, resp.) are not competitive with that of RFS. From the results depicted in Figure 6, we know that two or three is the optimal number of features for achieving the best classification accuracy. Table 4 shows the features and R-values for RFS. The R-value scores are ranked in ascending order; a lower R-value score indicates better quality features. Table 5 presents a comparison of the best classes among the base dataset, extended dataset, and the feature selection dataset. We adopted RFS for feature selection. Table 6 describes the experiment variables presented in Table 5. The applied dataset produced better classification accuracy than the base and extended datasets, while the proposed modified dataset that was collected by the RFS algorithm proved to be better than the original base dataset (Table 5). These results indicate that both the FA value and the SC region are important factors for SCI prediction; the latter is one of the main aspects influencing the classification result.

Table 7 shows the sensitivity and specificity analysis results, which is another way to evaluate classification results

Table 7: Accuracy, sensitivity, and specificity.

	Prediction	
	Injured	Normal
Fact		
Injured	100 (TP)	5 (FN)
Normal	5 (FP)	52 (TN)

Sensitivity = 0.952, specificity = 0.912, accuracy = 0.938.

apart from accuracy. The results of a classification are generally assessed using the following measures: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Accuracy (i.e., the proportion of correctly classified samples) is defined by

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}. \quad (5)$$

Sensitivity (i.e., the proportion of correctly classified positive samples) is defined by

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}. \quad (6)$$

Specificity (i.e., the proportion of correctly classified negative samples) is defined by

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})}. \quad (7)$$

In Table 7, the normal case specificity is 0.912 and the abnormal case sensitivity is 0.952. These results indicate that the proposed dataset has good predictive power for both normal and injured slices.

5. Conclusion

Diffusion Tensor Imaging (DTI) is generally used in the analysis of the brain and brain injuries. In recent years, this technique has been applied to Spinal Cord Injury (SCI) analysis. In this paper, we proposed a machine learning scheme for the diagnosis of SCI. We outlined a method of building datasets for diagnosis by observing four FA values from DTI slice images of patients and normal controls. Extension of the base dataset to a 15-feature dataset is one of the main contributions made in this study. Feature selection with the extended dataset resulted in improved prediction accuracy. A limitation of the study is that we did not acquire a large amount of SC data due to legal limitations for medical data. Nevertheless, we believe that if we gather more data, we

can improve our dataset and prediction accuracy. Our dataset is designed for specific cases of SCI in the C4–C6 area. A different area of injury will require a different kind of dataset. However, our approach can be used to build a suitable dataset for any such area. The ultimate goal of our study is to predict injured SC slices using a predefined training dataset. In the future, we plan to extend the current dataset with additional information so as to improve the prediction accuracy.

Conflict of Interests

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

Acknowledgment

The present research was conducted by the research fund of Dankook University in 2011.

References

[1] The American Spinal Injury Association, *Standard for Neurological Classification of Spinal Injured Patients*, ASIA, Chicago, Ill, USA, 1982.

[2] The American Spinal Injury Association/International Medical Society of Paraplegia, *International Standards for Neurological and Functional Classification of Spinal Cord Injury Patients*, ASIA, Chicago, Ill, USA, 1996.

[3] The American Spinal Injury Association/International Medical Society of Paraplegia, *International Standards for Neurological and Functional Classification of Spinal Cord Injury Patients*, ASIA, Chicago, Ill, USA, 2000.

[4] J. D. Steeves, D. Lammertse, A. Curt et al., "Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: clinical trial outcome measures," *Spinal Cord*, vol. 45, no. 3, pp. 206–221, 2007.

[5] F. Miyanji, J. C. Furlan, B. Aarabi, P. M. Arnold, and M. G. Fehlings, "Acute cervical traumatic spinal cord injury: MR imaging findings correlated with neurologic outcome. Prospective study with 100 consecutive patients," *Radiology*, vol. 243, no. 3, pp. 820–827, 2007.

[6] T. Lin, R. Li, X. Tang, J. G. Dy, and S. B. Jiang, "Markerless gating for lung cancer radiotherapy based on machine learning techniques," *Physics in Medicine and Biology*, vol. 54, no. 6, pp. 1555–1563, 2009.

[7] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fMRI data," *Trends in Cognitive Sciences*, vol. 10, no. 9, pp. 424–430, 2006.

[8] V. Athitsos, J. Alon, and S. Sclaroff, "Efficient nearest neighbor classification using a cascade of approximate similarity measures," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 486–493, June 2005.

[9] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.

[10] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, The MIT Press, Cambridge, Mass, USA, 2002.

[11] H. Zare, "FeaLect: Feature seLection by computing statistical scores," The R Journal, 2010, http://cran.rakanu.com .

[12] J. Liang, S. Yang, and A. Winstanley, "Invariant optimal feature selection: a distance discriminant and feature ranking based solution," *Pattern Recognition*, vol. 41, no. 5, pp. 1429–1439, 2008.

[13] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of reliefF and RReliefF," *Machine Learning*, vol. 53, no. 1-2, pp. 23–69, 2003.

[14] M. Seo and S. Oh, "High performance feature selection algorithm based on feature clearness (CBFS)," *PLoS ONE*, vol. 7, no. 7, Article ID e40419, 2012.

[15] J. Lee, B. Nomin, and S. Oh, "RFS :efficient feature selection method based on R-value," *Computers in Biology and Medicine*, vol. 43, no. 2, pp. 91–99, 2012.

[16] S. Oh, "A new dataset evaluation method based on category overlap," *Computers in Biology and Medicine*, vol. 41, no. 2, pp. 115–122, 2011.

[17] American Spinal Injury Association/International Medical Society of Paraplegia, *International Standards for Neurological and Functional Classification of Spinal Cord Injury Patients*, ASIA, Chicago, Ill, USA, 2003.

[18] K. Dirk-Jan, 2008, http://www.mathworks.com/matlabcentral/fileexchange/21130-dti-and-fiber-tracking.

[19] S. Jimmy, 2005, http://www.mathworks.com/matlabcentral/fileexchange/8797.

[20] J. Cohen-Adad, H. Leblond, H. Delivet-Mongrain, M. Martinez, H. Benali, and S. Rossignol, "Wallerian degeneration after spinal cord lesions in cats detected with diffusion tensor imaging," *NeuroImage*, vol. 57, no. 3, pp. 1068–1076, 2011.

[21] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistics Surveys*, vol. 4, pp. 40–79, 2010.

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 370849, 5 pages
http://dx.doi.org/10.1155/2014/370849

Research Article

Fully Automated Detection of Corticospinal Tract Damage in Chronic Stroke Patients

Ming Yang,[1] Ya-ru Yang,[1] Hui-jun Li,[1] Xue-song Lu,[2] Yong-mei Shi,[3] Bin Liu,[1] Hua-jun Chen,[1] and Gao-jun Teng[1]

[1] *Department of Radiology, Zhongda Hospital, Southeast University, Nanjing 210009, China*
[2] *Department of Rehabilitation, Zhongda Hospital, Southeast University, Nanjing 210009, China*
[3] *Department of Neurology, Affiliated Zhongda Hospital of Southeast University, Nanjing 210009, China*

Correspondence should be addressed to Ming Yang; yangming19710217@163.com

Received 8 November 2013; Accepted 24 December 2013; Published 15 January 2014

Academic Editor: Rong Chen

Structural integrity of the corticospinal tract (CST) after stroke is closely linked to the degree of motor impairment. However, current methods for measurement of fractional atrophy (FA) of CST based on region of interest (ROI) are time-consuming and open to bias. Here, we used tract-based spatial statistics (TBSS) together with a CST template with healthy volunteers to quantify structural integrity of CST automatically. Two groups of patients after ischemic stroke were enrolled, group 1 (10 patients, 7 men, and Fugl-Meyer assessment (FMA) scores $\leqslant 50$) and group 2 (12 patients, 12 men, and FMA scores = 100). CST of FA_{ipsi}, FA_{contra}, and FA_{ratio} was compared between the two groups. Relative to group 2, FA was decreased in group 1 in the ipsilesional CST ($P < 0.01$), as well as the FA_{ratio} ($P < 0.01$). There was no significant difference between the two subgroups in the contralesional CST ($P = 0.23$). Compared with contralesional CST, FA of ipsilesional CST decreased in group 1 ($P < 0.01$). These results suggest that the automated method used in our study could detect a surrogate biomarker to quantify the CST after stroke, which would facilitate implementation of clinical practice.

1. Introduction

Diffusion tensor imaging (DTI) can delineate anatomic connectivity of white matter and evaluate tract disruption in vivo, which is increasingly used in stroke-related research [1–5]. DTI-derived parameter such as fractional anisotropy (FA) has been found to reliably reflect the microstructural status of corticospinal tract (CST) in patients with stroke [6–8]. Greater gains in motor function were related to higher FA values of ipsilesional CST, and slice-by-slice analysis of FA values along the CST demonstrated that the more the ipsilesional FA profiles of patients resembled those of healthy controls, the greater their functional improvement was [6]. Meanwhile the reverse is also true that greater loss of structural integrity of the ipsilesional CST is associated with poorer motor outcomes in patients with hemiparetic stroke [7, 8].

Despite these advances, some factors impede the uptake of these approaches. CST tracking in individual stroke is often difficult due to interruption of fibers by the infarct which can result in the unreliable morphology of the tracts. Moreover, manual placement of regions of interest (ROI) in individual patients is also problematic because of operator bias, and manual labeling is time-consuming. For these reasons, its feasibility is limited. Therefore, a fully automated method of evaluating CST is urgently needed to satisfy the translational potential of CST injury quantification to clinical practice.

Tract-based spatial statistics (TBSS) is a new approach which is fully automated to investigate the whole brain without prespecification of tracts of interest [9]. Meanwhile the method does not require smoothing and performs alignment across participants and has a high sensitivity for identifying white-matter (WM) deficits using nonlinear registration and tract projection. Recently this method has been applied to evaluate white-matter changes of schizophrenia [10, 11] and Parkinson's disease [12]. Yet, few studies have reported stroke-related changes in the WM structural networks using TBSS [13–15]. Most of these studies mirrored lesion in one

hemisphere to another across the midsagittal axis. The goal is to increase patients' homogeneity and control for the location of the lesion to be able to interpret voxel-wise analysis [13, 14]. However, this method does not use a symmetric cerebral template and may cause unreliable results. Only one study did not adopt mirror about midline due to enrolling lesions in the same hemisphere [15]. However the lesion locations are not in accordance with clinical practice and greatly limit the subject sample.

The aim of our study was to contrast DTI-derived fully automated detection of corticospinal tract damage between two subgroups of chronic stroke patients with different recovery and meanwhile investigate the possibility of the translational potential of CST injury quantification to clinical practice. Given the diversity issue in the stroke population, a skeletonized mean FA image came from TBSS was not used to contrast between two subgroups and only individual FA images were evaluated in our study. Each participant's (aligned) FA image was acquired by filling the skeleton with FA values from the nearest relevant tract center. Then individual CSTs were acquired by overlaying individual FA images with CST digital WM atlas JHU [16]. This was done to fully automate the process of obtaining ipsilesional and contralesional CST individually with no need to mirror the lesion from one hemisphere to another.

2. Methods

2.1. Subjects. The study group consisted of 22 patients, all of whom had ischemic strokes in the middle cerebral artery territory at least 6 months before entering the study. The patients were divided into 2 subgroups according to their Fugl-Meyer assessment (FMA) scores: group 1 (10 patients, 7 men, and FMA scores $\leqslant 50$) and group 2 (12 patients, 12 men, and FMA scores = 100). The FMA is a 50-item motor function assessment with scores ranging from 0 to 100. FMA scores $\leqslant 50$ belong to severe motor impairment, and FMA scores = 100 means no motor impairment [17]. It is one of the most frequently used clinical motor impairment tests in stroke rehabilitation research. All patients in this study required inpatient rehabilitation, which consisted of standard physical and occupational therapy. Exclusion criteria were as follows: (1) prior or subsequent symptomatic stroke; (2) bihemispheric or brain stem infarcts, primary intracerebral hemorrhages, and other disorders that impaired motor function of the stroke-affected hand and leg; (3) other concomitant neurological or psychiatric diseases; and (4) contraindication to MRI. Patient clinical characteristics and demographic data are summarized in Table 1. This study was approved by the Ethics Committee of Southeast University and a signed informed-consent form was obtained from every subject prior to the experiment.

2.2. MRI Procedures. MR images were obtained using a Siemens Verio Tim 3.0 T MR scanner (Siemens Medical Solutions, Erlangen, Germany). DTI images were acquired using a 12-channel phased-array head coil with the implementation of the parallel imaging scheme GRAPPA (GeneRalized Autocalibrating Partially Parallel Acquisitions) and an acceleration factor of 2. A single-shot echo planar imaging (EPI) sequence was used in DTI data including 30 nonlinear diffusion directions with $b = 1000\,\text{s/mm}^2$ and an additional volume with $b = 0\,\text{s/mm}^2$. The detailed parameters are as follows: number of axial sections, 70; slice thickness, 2 mm; gap, none; acquisition matrix, 128×124; TR, 10,900 ms; TE, 95 ms; field of view, 256 mm × 256 mm; and average, 2. We also acquired 3D high resolution brain structural images (voxel size = $1\,\text{mm}^3$, isotropic) using a T1-weighted MPRAGE sequence for each subject. The sequence parameters were TR/TE = 1900 ms/2.48 ms, inversion time (TI) = 900 ms, flip angle = 9°, FOV = 256 mm × 256 mm, slice thickness = 1 mm, and 176 sagittal slices covering the whole brain. To identify the location and size of the lesion, we acquired fluid attenuated inversion recovery (FLAIR) images: number of axial sections, 20; slice thickness, 5 mm; gap, none; acquisition matrix, 128×124; TR, 8,500 ms; TE, 94 ms; field of view, 230 mm × 208 mm; flip angle, 150°. All subjects were scanned using the same MR scanner.

2.3. Lesion Mapping. We used MRIcron to define the lesions to obtain a VOI (volume of interest) in the raw T1-weighted images (T1WI). For each patient, we manually outlined the lesion area on each slice while using the FLAIR images as an additional guide to confirm the extent of the lesion. Lesions were defined as hypointensity area with clear border in T1WI. The whole lesion volumes were determined by integration across all relevant slices. Lesions were drawn by an experienced rater who was blinded to the patients' FMA scores.

2.4. DTI Data Processing. DTI processing was performed using the FSL 5 (FMRIB Software Library, http://www.fmrib.ox.ac.uk/fsl). Then, a statistical analysis of FA data was carried out using TBSS, part of FSL [9]. First, for each participant, DTI images were registered to the corresponding $b = 0$ images with an affine transformation to correct eddy-current distortion and head motion by using FMRIB's Diffusion Toolbox (FDT, part of FSL). Next, no brain tissues were removed using the Brain Extraction Tool (BET, part of FSL) [18]. These were referred to as the preprocessing stages. For the next step, the FA data were aligned into 1 mm × 1 mm × 1 mm Montreal Neurological Institute (MNI) 152 space using the FMRIB's Nonlinear Image Registration Tool (FNIRT). Then, the mean FA image (threshold of 0.2) was created and thinned to create a mean FA skeleton which represents the centers of all tracts common to group 1 and group 2 (Figure 1). Each subject's aligned FA data was then projected onto this skeleton and the skeletonized FA images were obtained.

2.5. ROI Analysis. Our goal is to calculate regional FA for fiber tracts defined in the JHU white-matter tract template [16]. JHU white-matter tract template is defined in the MNI152 space. For each skeletonized FA map, we used the

Table 1: Clinical and demographic data of 22 patients with ischemic stroke.

Patient no.	Sex	Age (y)	Dominant hand	Location of lesion	Time after stroke (months)	Education (years)	Lesion volume (mm^3)	FMA Score
				Group 1				
01	M	67	R	R, PLIC	9.7	9	551.2	42
02	F	64	R	L, CR, temporal lobe	6.5	8	859.3	31
03	M	63	R	R, PLIC, parietal lobe	23.3	6	1914.0	10
04	M	37	R	R, PLIC, BG	6.9	17	1429.6	30
05	M	51	R	R, PLIC, BG	59	12	1960.8	47
06	M	65	R	R, CR, BG	54.4	9	1339.0	42
07	M	60	R	R, CR, BG	42.2	9	5133.6	47
08	F	69	R	L, CR, BG	12.6	8	1280.8	20
09	M	62	R	L, BG, CR	26.2	6	1592.6	17
10	F	73	R	R, BG, CR	40.4	0	845.0	25
				Group 2				
01	M	56	R	R, TH	12.2	8	25.7	100
02	M	52	R	L, CR, CS	7.2	12	100.1	100
03	M	60	R	L, CR, temporal lobe	11	12	26071.5	100
04	M	43	R	L, CR, BG	16	9	84.9	100
05	M	70	R	L, TH, BG	43	12	624.7	100
06	M	70	R	L, TH	25.3	6	371.0	100
07	M	56	R	L, BG	24	12	61.0	100
08	M	59	R	L, CR, BG	53.4	12	2073.3	100
09	M	60	R	L, BG	10.6	12	29.6	100
10	M	61	R	R, TH	12.7	0	96.3	100
11	M	68	R	L, CR, BG	9.7	6	772.5	100
12	M	53	R	R, BG	15.2	15	164.0	100

M: male; F: female; L: left; R: right; IC: internal capsule; PLIC: posterior limb of IC; BG: basal ganglia; CR: corona radiata; TH: thalamus; CS: centrum semiovale.

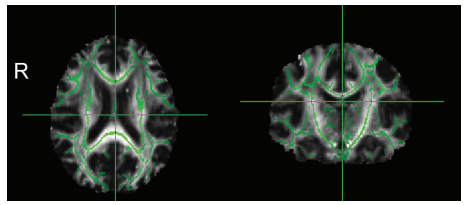


Figure 1: Axial and coronal views of mean FA skeleton which represent the centers of all tracts common to group 1 and group 2.

JHU WM tract template to delineate bilateral CSTs. Ipsilateral FA (FA_{ipsi}), contralateral FA (FA_{contra}), and FA_{ratio} were computed in terms of three measures over the whole CST, FA_{ratio} was computed as a ratio (FA_{ipsi}/FA_{contra}).

2.6. Statistical Analysis. Statistical analysis of the demographic and clinical data was performed using the SPSS 16.0 (SPSS Inc., Chicago, IL, USA). The threshold for statistical significance was $P < 0.05$. The differences between the two groups in age, education, and the time after stroke were tested with two sample t-tests, as well as FA_{ipsi}, FA_{contra}, and FA_{ratio} of CST. Fisher's exact test was used to assess gender difference and side of lesion. Data were normally distributed according to Kolmogorov-Smirnov test.

3. Results

3.1. Clinical Data. No significant differences between the two subgroups were observed in age (group 1: mean ± SD, 61.10 ± 10.28 years; group 2: mean ± SD, 59.00 ± 7.91 years; $P = 0.59$), time after stroke (group 1: mean ± SD, 28.12 ± 19.78 months; group 2: mean ± SD, 20.03 ± 14.40 months; $P = 0.28$), education (group 1: mean ± SD, 8.40 ± 4.35 years; group 2: mean ± SD, 9.67 ± 4.10 years; $P = 0.49$), lesion volume (group 1: mean ± SD, 1690.6 ± 1293.2 mm^3; group 2: mean ± SD, 2539.6 ± 7433.4 mm^3; $P = 0.72$), sex (group 1: M/F, 7/3; group 2: M/F, 12/0; $P = 0.08$), and side (group 1: L/R, 3/7; group 2: L/R, 9/3; $P = 0.08$). We found a significant difference of the FMA score between the 2 subgroups (group 1: mean ± SD, 31.10 ± 13.12; group 2: mean ± SD, 100 ± 0 months; $P < 0.01$).

3.2. ROI Analysis of FA of CST. FA of CST between the two subgroups can be seen in Table 2. Compared with group 2, FA was decreased in group 1 in the ipsilesional CST ($P < 0.01$), as well as the FA_{ratio} ($P < 0.01$). There was no significant difference between the two groups in the contralesional CST

Table 2: FA of CST between the two subgroups.

	FA_{ipsi}	FA_{contra}	FA_{ratio}	P value (FA_{ipsi}–FA_{contra})
Group 1	0.14 ± 0.01	0.17 ± 0.01	0.83 ± 0.07	<0.01
Group 2	0.17 ± 0.007	0.17 ± 0.009	0.98 ± 0.04	0.24
P value	<0.01	0.23	<0.01	

Note: data are means ± SD.

($P = 0.23$). Compared with contralesional CST, FA of ipsilesional CST decreased in group 1 ($P < 0.01$). There was no significant difference between bilateral CSTs in group 2.

4. Discussion

To the best of our knowledge, this study is the first that combines TBSS and CST templates to quantify CST in subgroups of patients who have had a stroke with different recoveries. It is well established that quantitative detection of CST based on DTI plays an important role in recovery after stroke. Greater gains in motor function were related to higher FA values of ipsilesional CST [6], and greater loss in structural integrity of the ipsilesional CST is associated with poorer motor outcomes in patients with hemiparetic stroke [7, 8]. In our study, we enrolled two groups of patients with poor (group 1, Figure 2) and good recoveries (group 2, Figure 3) to validate the automated method in evaluating injury of CST. In accordance with previous studies, we found that FA of the ipsilesional CST significantly decreased in group 1 compared with group 2. And FA_{ratio} also decreased in group 1 compared with group 2.

In addition, we found that there is no significant change of FA in the contralesional CST between the two subgroups. This result was in accordance with the other studies, where FA values of contralateral pyramidal tracts (PT) did not differ from the values contrasted with control [1, 6]. However, there were some inconsistent findings. Puig et al. showed that mean FA values in unaffected CST increased at day 30 in line with increasing motor deficit [2]. Another study found that the FA value of CST in the unaffected hemisphere in patients who could not walk independently was significantly lower than that of the control group ($P < 0.05$) [19]. This discrepancy may be due to the differences in the detailed methodology including the variable subjects and sample, function recovery standard scale, and data processing. The exact FA value change of contralesional CST needs further evaluation based on a larger sample size, a longitudinal evaluation, and more accurate measurement.

Our study adopted a method of both TBSS and a CST template to quantify CST fully automatically in subgroup stroke patients. It has the advantage of TBSS, which eliminates the need for smoothing and alignment across participants, as well as prespecification of tracts of interest. Meanwhile a CST template from healthy volunteers was referenced and bilateral CSTs of individual patients can be fully automatically obtained through projecting onto this skeleton. The method is automatic and objective, improving efficiency and eliminating bias, and would facilitate implementation in clinical practice.

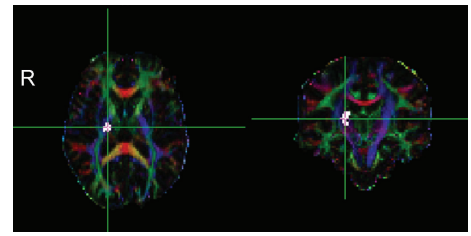


Figure 2: Axial and coronal views of individual color-coded FA map of subject 01 from group 1 together with lesion map. The patient's FMA score was 42 and the time after stroke was 9.7 months. Colors indicate direction of fiber tracts (red, left-right; blue, craniocaudal; green, anterior-posterior). Crosshairs are centered on the infarct (white color) which shows that right posterior limb of internal capsule (part of CST) was interrupted by the lesion.

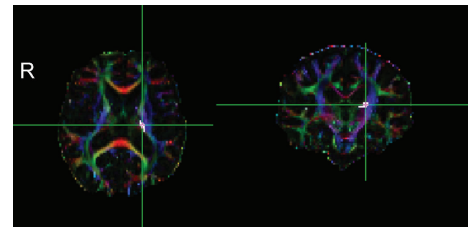


Figure 3: Axial and coronal views of individual color-coded FA map of subject 05 from group 2 together with lesion map. The patient's FMA score was 100 and the time after stroke was 43 months. Colors indicate direction of fiber tracts (red, left-right; blue, craniocaudal; green, anterior-posterior). Crosshairs are centered on the infarct (white color) which shows that the lesion is close to the left posterior limb of internal capsule and CST is intact.

Our study had some limitations. First, the spatial accuracy of TBSS analysis is limited by the skeleton and, therefore, does not provide significant details in some locations. However we carefully examined the result of normalization for each participant to confirm the accuracy and validity of our postprocessing. Second, this is a cross-sectional study with relatively small sample size, and a large scale, longitudinal evaluation is needed to further confirm the reliability of our method, as well as the occurrence and development roles of CST. Finally, we did not consider the extent of leukoaraiosis, which will be involved in a future study.

In conclusion, our research demonstrated the possibility to automatically quantify CST, opening the avenue for large-scale studies of the utility of unbiased assessment of CST integrity after stroke in predicting motor outcomes.

Conflict of Interests

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

Authors' Contribution

Ming Yang and Ya-ru Yang have contributed equally to this work.

Acknowledgments

The study was supported by the Excellent Teacher Project of Jiangsu Provincial Department of Education and was partly supported by the National Science Foundation of China (Grant no. 30970808). The authors also thank Rui Dong for her revisions of the paper.

References

[1] R. Lindenberg, V. Renga, L. L. Zhu, F. Betzler, D. Alsop, and G. Schlaug, "Structural integrity of corticospinal motor fibers predicts motor impairment in chronic stroke," *Neurology*, vol. 74, no. 4, pp. 280–287, 2010.

[2] J. Puig, G. Blasco, J. Daunis-I-Estadella et al., "Decreased corticospinal tract fractional anisotropy predicts long-term motor outcome after stroke," *Stroke*, vol. 44, no. 7, pp. 2016–2018, 2013.

[3] L. K. Sztriha, R. L. O'Gorman, M. Modo, G. J. Barker, S. C. R. Williams, and L. Kalra, "Monitoring brain repair in stroke using advanced magnetic resonance imaging," *Stroke*, vol. 43, no. 11, pp. 3124–3614, 2012.

[4] N. Kou, C. -h. Park, M. L. Seghier, A. P. Leff, and N. S. Ward, "Can fully automated detection of corticospinal tract damage be used in stroke patients?" *Neurology*, vol. 80, no. 24, pp. 2242–2245, 2013.

[5] L. L. Zhu, R. Lindenberg, M. P. Alexander, and G. Schlaug, "Lesion load of the corticospinal tract predicts motor impairment in chronic stroke," *Stroke*, vol. 41, no. 5, pp. 910–915, 2010.

[6] R. Lindenberg, L. L. Zhu, T. Rüber, and G. Schlaug, "Predicting functional motor potential in chronic stroke patients using diffusion tensor imaging," *Human Brain Mapping*, vol. 33, no. 5, pp. 1040–1051, 2012.

[7] J. D. Schaechter, Z. P. Fricker, K. L. Perdue et al., "Microstructural status of ipsilesional and contralesional corticospinal tract correlates with motor skill in chronic stroke patients," *Human Brain Mapping*, vol. 30, no. 11, pp. 3461–3474, 2009.

[8] R. Schulz, C. H. Park, M. H. Boudrias, C. Gerloff, F. C. Hummel, and N. S. Ward, "Assessing the integrity of corticospinal pathways from primary and secondary cortical motor areas after stroke," *Stroke*, vol. 43, no. 8, pp. 2248–2251, 2012.

[9] S. M. Smith, M. Jenkinson, H. Johansen-Berg et al., "Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data," *NeuroImage*, vol. 31, no. 4, pp. 1487–1505, 2006.

[10] T. Nickl-Jockschat, T. Stöcker, V. Markov et al., "The impact of a Dysbindin schizophrenia susceptibility variant on fiber tract integrity in healthy individuals: a TBSS-based diffusion tensor imaging study," *NeuroImage*, vol. 60, no. 2, pp. 847–853, 2012.

[11] X. Liu, Y. Lai, X. Wang et al., "Reduced white matter integrity and cognitive deficit in never-medicated chronic schizophrenia: a diffusion tensor study using TBSS," *Behavioural Brain Research*, vol. 252, pp. 157–163, 2013.

[12] H. J. Kim, S. J. Kim, H. S. Kim et al., "Alterations of mean diffusivity in brain white matter and deep gray matter in Parkinson's disease," *Neuroscience Letters*, vol. 550, pp. 64–68, 2013.

[13] G. Jayaram, C. J. Stagg, P. Esser, U. Kischka, J. Stinear, and H. Johansen-Berg, "Relationships between functional and structural corticospinal tract integrity and walking post stroke," *Clinical Neurophysiology*, vol. 123, no. 12, pp. 2422–2428, 2012.

[14] L. E. Wang, M. Tittgemeyer, D. Imperati et al., "Degeneration of corpus callosum and recovery of motor function after stroke: a multimodal magnetic resonance imaging study," *Human Brain Mapping*, vol. 33, no. 12, pp. 2941–2956, 2012.

[15] D. Yin, X. Yan, M. Fan et al., "Secondary degeneration detected by combining voxel-based morphometry and tract-based spatial statistics in subcortical strokes with different outcomes in hand function," *American Journal of Neuroradiology*, vol. 34, no. 7, pp. 1341–1347, 2013.

[16] K. Hua, J. Zhang, S. Wakana et al., "Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification," *NeuroImage*, vol. 39, no. 1, pp. 336–347, 2008.

[17] A. R. Fugl-Meyer, "Post-stroke hemiplegia: assessment of physical properties," *Scandinavian Journal of Rehabilitation Medicine*, vol. 12, no. 7, pp. 85–93, 1980.

[18] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002.

[19] S. H. Jang, C. H. Chang, J. Lee, C. S. Kim, J. P. Seo, and S. S. Yeo, "Functional role of the corticoreticular pathway in chronic stroke patients," *Stroke*, vol. 44, no. 4, pp. 1099–1104, 2013.