

Multimodal Medical Image Registration Based on an Information-Theory Measure with Histogram Estimation of Continuous Image Representation

Bicao Li, Guanyu Yan, Zhoufeng Liu, Jean Louis Coatrieux, Huazhong Shu

▶ **To cite this version:**

Bicao Li, Guanyu Yan, Zhoufeng Liu, Jean Louis Coatrieux, Huazhong Shu. Multimodal Medical Image Registration Based on an Information-Theory Measure with Histogram Estimation of Continuous Image Representation. Mathematical Problems in Engineering, 2018, 2018, pp.2135453. 10.1155/2018/2135453 . hal-01807120

HAL Id: hal-01807120

https://univ-rennes.hal.science/hal-01807120

Submitted on 9 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 2135453, 12 pages
https://doi.org/10.1155/2018/2135453

Research Article

Multimodal Medical Image Registration Based on an Information-Theory Measure with Histogram Estimation of Continuous Image Representation

Bicao Li,[1] **Guanyu Yang**,[2,3,4] **Zhoufeng Liu**,[1] **Jean Louis Coatrieux**,[2,5,6,7] **and Huazhong Shu**[2,3,4]

[1] *School of Electronic and Information Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China*
[2] *Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing 210096, China*
[3] *Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China*
[4] *Centre de Recherche en Information Médicale Sino-Français (CRIBs), Nanjing 210096, China*
[5] *INSERM, U1099, 35000 Rennes, France*
[6] *LTSI, Université de Rennes 1, 35042 Rennes, France*
[7] *Centre de Recherche en Information Médicale Sino-Français (CRIBs), 35000 Rennes, France*

Correspondence should be addressed to Bicao Li; lbc@zut.edu.cn

Received 27 October 2017; Revised 27 February 2018; Accepted 13 March 2018; Published 29 April 2018

Academic Editor: Bogdan Dumitrescu

Copyright © 2018 Bicao Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work presents a novel method for multimodal medical registration based on histogram estimation of continuous image representation. The proposed method, regarded as "fast continuous histogram estimation," employs continuous image representation to estimate the joint histogram of two images to be registered. The Jensen–Arimoto (JA) divergence is a similarity measure to measure the statistical dependence between medical images from different modalities. The estimated joint histogram is exploited to calculate the JA divergence in multimodal medical image registration. In addition, to reduce the grid effect caused by the grid-aligning transformations between two images and improve the implementation speed of the registration method, random samples instead of all pixels are extracted from the images to be registered. The goal of the registration is to optimize the JA divergence, which would be maximal when two misregistered images are perfectly aligned using the downhill simplex method, and thus to get the optimal geometric transformation. Experiments are conducted on an affine registration of 2D and 3D medical images. Results demonstrate the superior performance of the proposed method compared to standard histogram, Parzen window estimations, particle filter, and histogram estimation based on continuous image representation without random sampling.

1. Introduction

Image registration is the task of finding the optimal geometric transformation between two images. Image registration is widely used in numerous fields, such as medical imaging, computer vision, and remote sensing. Medical images from different modalities can provide various complementary information. Therefore, the registration of multimodal medical images is valuable in multimodal diagnosis and computer-aided surgery [1–3].

Information-theory-based image registration has become a popular method for multimodal medical images. In these methods, mutual information (MI) is a frequently used similarity measure simultaneously proposed by Collignon et al. [4], Maes et al. [5], Wells III et al. [6], and Viola and Wells III [7]. These approaches do not require any preprocessing and have been found to hold for a range of applications. Many similarity measures based on information theory have been employed for medical image registration, for example, normalized MI [8], MI combined with gradient information

[9], cumulative residual entropy (CRE) [10, 11], and cross-CRE [12].

The critical element for all these information-theory-based approaches for image registration is the estimation of probability density functions (PDFs). Standard histograms have been the most frequently used for estimating probability distributions [4], which is a straightforward algorithm and easily implemented. However, Rajwade et al. [13] pointed out that the number of histogram bins determines the estimation of PDFs, and a relatively small number can provide the preferable PDFs. Standard histogram method also yields discontinuous probability estimates, and there are no principle techniques to select the suitable number of bins for calculating the PDFs of images to be registered. The Parzen window (PZW) [7] has been exploited for probability density estimation in image registration. This method was developed further by Thévenaz and Unser [14]. However, the optimal selection of kernel function and kernel width is a challenging task when utilizing PZW to estimate PDFs. Nonparametric (NP) windows [15] were introduced to compute PDFs more accurately than PZW. However, the computational and implementation complexity of NP windows limits the method's actual application. Joshi et al. [16] simplified the implementation of NP windows by using planar interpolation. Standard histogram and PZW methods possess one common property; that is, both approaches regard an image as a set of pixels or samples [17]. Discrete histogram transform (DHT) was introduced in [18]. This method averages the estimations from several multivariate histograms to alleviate the discontinuities in the boundaries of histogram bins. However, the computational cost of DHT increases quickly as the number of histograms grows. López-Rubio and Muñoz-Pérez [19] proposed the use of multivariate frequency polygons (MFP) to reduce computational complexity. By contrast, probability estimations based on DHT and MFP also rely on training samples and bin width. Maes et al. [5] introduced the partial volume estimation (PVE) to calculate probability distributions. Chen and Varshney proposed the generalized PVE [20] technique, which applies a general kernel function to determine fractional votes of each intensity pair. The estimated methods based on partial volume construct joint probability distribution through continuous image representation (CIR), whereas the PVE approach still substantially requires the choice of kernel function [17].

In our earlier work [21], a novel divergence measure called the Jensen–Arimoto (JA) divergence was presented as registration criterion for 2D-2D rigid medical registration. To calculate the JA divergence, the joint probability distributions between two images to be registered need to be estimated using the standard/simple histogram (SH) algorithm, PVE, or PZW method. Whereas the critical element for SH based estimation methods is selecting the suitable number of bins, PVE and PZW both belong to the continuous estimation of densities. A comparison between these two methods to density estimation has been introduced in [22]. Darkner and Sporring [23] also concluded that PZW is algorithmically a superior estimator for NMI compared to GPV. Nevertheless, these two approaches both need to select the appropriate kernel function. In a word, the number of bins for SH and selection of kernel function for PVE and PZW are the current challenges and limitations in estimation of densities. To overcome these problems, we introduce a CIR for estimating the joint histogram of two images to be registered. Different from the aforementioned approaches, our method avoids binning problem and the choice of kernel function. The most straightforward strategy to estimate joint histograms by CIR is to use all voxels from aligned images. However, this way is time consuming for the images with large sizes. Additionally, the grid effect caused by grid-aligning transformations and the discontinuities of similarity measures are two major problems associated with these information-theory-based methods [24]. Therefore, to solve these issues, we choose a certain number of random samples from the images to be registered, instead of all pixel points, to estimate the joint histogram. The computation time for histogram estimation is then reduced by random sampling strategy. Our proposed histogram estimation method is regarded as "fast continuous histogram estimation" (FCHE). We apply the FCHE algorithm to estimate joint probability and then calculate the JA divergence for multimodality medical registration. Affine transformation models are employed, and the simplex Nelder–Mead method [25] is used to search the maximum of the JA divergence. To evaluate the performance of our registration technique based on FCHE, we implement experiments on 2D and 3D medical images. The results demonstrate that FCHE method can provide better registration accuracy and improve the implementation efficiency.

The remainder of this paper is organized as follows. In Section 2, the FCHE method for joint histogram estimations and a review of the JA divergence measure are introduced. The subsequent image registration technique is detailed in Section 3. Section 4 provides our experimental results on 2D and 3D medical images. This section also compares these findings with those of the registration strategies based on continuous histogram estimation (CHE) without random sampling, SH, PF, and PZW. Concluding remarks and perspectives are presented in Section 5.

2. Theory

2.1. FCHEs. In information-theory-based registration approaches, the joint probabilities between two images to be registered are the critical element to successfully calculate the similarity measure. As mentioned above, several algorithms for computing joint probabilities include SH, PZW, and PVE. In this work, a spatially continuous representation of an image is adopted to estimate a joint histogram for obtaining joint probabilities.

2.1.1. Joint Histogram in 2D. Two images, namely, the reference image (R) and the floating image (F), are considered along with their intensity values normalized into a specified number of bins. For every grid point in each image, around which the intensity values of four neighborhood points were estimated by Rajwade et al. [13], here the distances between the grid point and four neighborhood points are all 0.5 pixels

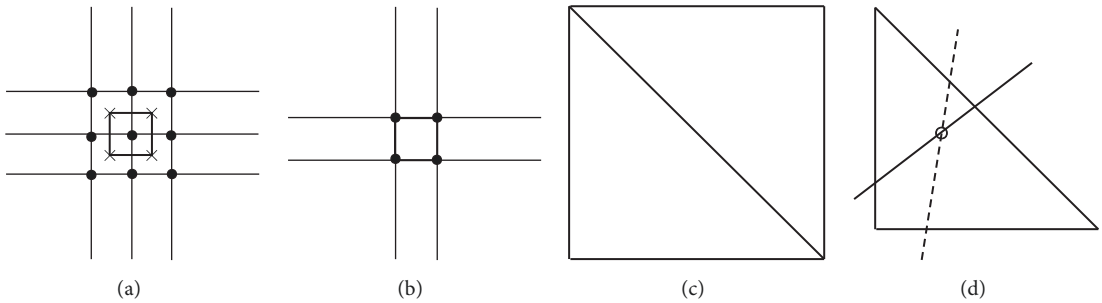


Figure 1: Process of joint histogram estimations in 2D. (a) Four neighborhood points around the central point on nongrid locations (cross) and nine grid points (black dots). (b) Four points on the grid (four dots). (c) The square constituted by four grid points in (b) is divided into two triangles. (d) The intersection point (black circle) of two isointensity lines. The solid line and dashed line represent the respective isocurves of two images.

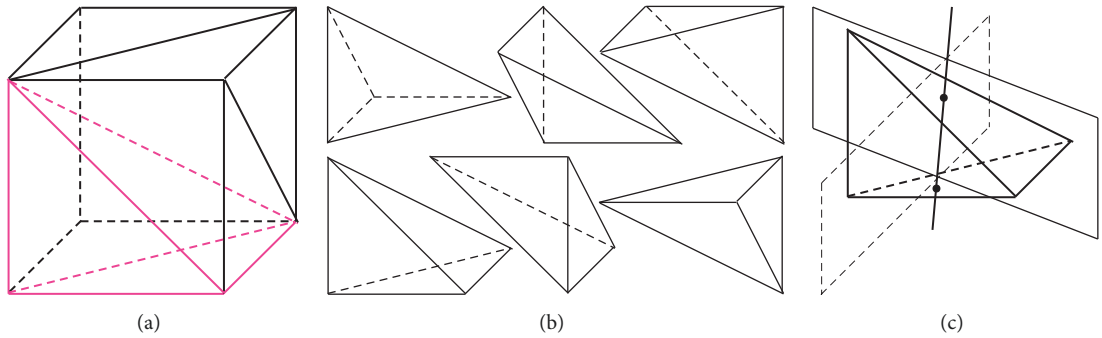


Figure 2: Process of joint histogram estimations in 3D. (a) Cube. (b) Six tetrahedrons divided from the cube. (c) Intersecting line between two isointensity surfaces (plotted by the solid and dashed lines) with the distance of two black dots (two crossing points between the intersecting line and the tetrahedron) of the segment length.

(Figure 1(a)). These intensity values of four neighbors in nongrid points are estimated through interpolation methods, which inevitably increase the computational cost and introduce interpolation error into registration results. To address this problem, four grid points are used to compose a square for histogram estimations (Figure 1(b)), and then the square (a pixel) is split into two triangles (half-pixels).

The intensity values of the three vertices in each triangle are regarded as linear functions of the vertices' respective coordinates. The related equations are as follows:

$$\begin{aligned} g_r\left(x_j, y_j\right) &= A_1 x_j + B_1 y_j + C_1, \\ g_f\left(x_j, y_j\right) &= A_2 x_j + B_2 y_j + C_2, \\ j &= \{1, 2, 3\}, \end{aligned} \tag{1}$$

where (x_j, y_j) denotes the coordinates of three vertices (no. j) of the triangle shown in Figure 1(d) and $g_r(x_j, y_j)$ and $g_f(x_j, y_j)$ represent the corresponding intensity values of images R and F, respectively. Additionally, A_i, B_i, and C_i are the coefficients obtained by solving the equations defined in (1). For the values of the three coefficients, the intensity values of any coordinate (x, y) within each triangle of two images are calculated using $A_i x + B_i y + C_i$. Conversely, given a pair of intensity values (one intensity value in R and one in F, signified as $[r, f]$) within the range of specified bins, we can yield two isointensity lines, $r = A_1 x + B_1 y + C_1$ and $f = A_2 x + B_2 y + C_2$. The joint histogram is updated if the intersection of two lines (x_0, y_0) falls into a pair of corresponding triangles from two images as shown below:

$$\begin{aligned} &h\left(g_1\left(x_0, y_0\right) = r, g_2\left(x_0, y_0\right) = f\right) \\ &\quad = h\left(g_1\left(x_0, y_0\right) = r, g_2\left(x_0, y_0\right) = f\right) + 1, \end{aligned} \tag{2}$$

where h represents the joint histogram of the entry (r, f) between two images R and F. We repeat the aforementioned process and cover all entries in the range of bins to acquire the entire joint histograms.

2.1.2. Joint Histogram in 3D. Similar to the joint histogram in 2D, we still account for the CIR when using 3D images. Eight vertices of a voxel in grid points are selected to constitute one cube, which is divided into 12 or 24 tetrahedrons through the cube center [17]. The intensity values of the eight vertices (grid points) can be directly acquired, whereas that of one center is evaluated through interpolation because of its location on a nongrid point. Nevertheless, we partition the cube (or one voxel) into six equal volumes using the eight vertices of this cube. Hence, we obtain six tetrahedrons (Figure 2(b)) without the need of the cube's center. Similar to the estimation in 2D, the intensity values of the four vertices

Input: Two images R and F.
Stage 1: The coordinates of sample points of R and F are obtained by random sampling strategy.
Stage 2: Each sample point is partitioned into two triangles (2D) or six tetrahedrons (3D case).
Stage 3: The isointensity lines or isosurfaces of the corresponding sample points from two images are constructed.
Stage 4: The joint histogram between R and F is updated using the isointensity lines or isosurfaces. Go to **stage 2**, unless all sample points are exhausted.
Output: The joint histogram between R and F.

Algorithm 1: Fast continuous histogram estimation.

of each tetrahedron are deemed as linear functions of the vertices' respective coordinates as shown below:

$$g_i\left(x_j, y_j, z_j\right) = A_i x_j + B_i y_j + C_i z_j + D_i \quad i = \{1, 2\}, \; j = \{1, 2, 3, 4\}. \tag{3}$$

Likewise, (x_j, y_j, z_j) and $g_i(x_j, y_j, z_j)$ are the coordinates and intensity values of the four vertices of each tetrahedron, along with i of the two images and j as the four vertices. In addition, A_i, B_i, C_i, and D_i are the coefficients that determine the linear equations in (3). The values of these coefficients are computed through these equations. On the basis of the coefficients, the two isointensity planes $r = A_1x + B_1y + C_1z + D_1$ and $f = A_2x + B_2y + C_2z + D_2$ are established. For the 3D case, we first consider the segment of the intersecting line of two planes. Then, the joint histogram of (r, f) is accumulated by measuring the length of the segment within each pair of homologous tetrahedrons from two respective images. The related equation is as follows:

$$h\left(g_1(x, y, z) = r, g_2(x, y, z) = f\right) = h\left(g_1(x, y, z) = r, g_2(x, y, z) = f\right) + S_{\text{length}}, \tag{4}$$

where S_{length} and h denote the segment length and the joint histogram, respectively. Considering one tetrahedron, the calculation process for the joint histogram in 3D is shown in Figure 2(c).

2.1.3. Random Sample. Information-theory-based similarity measures rely on the computation of a joint histogram between two images or volumes to be registered. The straightforward strategy for calculating the joint histogram is applying all pixels or voxels. However, the high implementation cost impedes the application of this strategy for large images or 3D images. An alternative approach is selecting a subset of pixels or voxels from an entire image. Generally, the subset is chosen from a uniform grid or random coordinate. Notably, random samples can be obtained from the grid points or the nongrid coordinates [25]. Bhagalia et al. [26] proposed selecting samples that lie on notable image features.

Press et al. [25] demonstrated that random samples on nongrid locations can improve the smoothness of MI. They also compared their experimental results to those obtained from several sample methods provided by Klein et al. [27]. The random sampling off the grid can reduce the grid effect and discontinuities of information-theory-based similarity measures. Accordingly, while applying our registration method, we adopt a subset of samples selected from the nongrid coordinates in the fixed image to estimate the joint histogram in 2D or 3D. Finally, under FCHE (Algorithm 1), the execution time of histogram estimations is substantially shortened while the registration accuracy is preserved. For instance, estimating the joint histogram between two 3D images with dimensions of $41 \times 41 \times 41$ voxels approximately lasts for 21.7 s without the need of a random sample. By contrast, the evaluated time under the FCHE strategy is only about 1.5 s when choosing 5000 random samples from the fixed image.

2.2. JA Divergence. The Arimoto entropy [28], a generalization of Shannon entropy, was introduced by Arimoto and further developed by Liese and Vajda [29]. Its definition is given by

$$A_\alpha(X) = \frac{\alpha}{\alpha - 1}\left[1 - \left(\sum_{i=1}^{M} p_i^\alpha\right)^{1/\alpha}\right] \quad \alpha > 0, \; \alpha \neq 1. \tag{5}$$

Unlike Shannon entropy, Arimoto entropy is a nonextensive entropy. The parameter α reflects the degree of nonextensivity. The nonextensivity gradually weakens as the parameter approaches 1, and the limit of the Arimoto entropy equals the Shannon entropy when $\alpha \to 1$.

We then follow the derivation of the Jensen–Shannon divergence reported by Lin [30]. In our previous work [21], a divergence measure based on the Arimoto entropy called the JA divergence was introduced. The JA divergence was used to measure the distance of two random variables and showed a superior performance in registering medical images to those of other information-theory measures.

Given a random variable X $(x_1, x_2, \ldots, x_M)$ with probability distributions of P $(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_M)$, we define the JA divergence [21] as

$$\mathrm{JA}_\alpha\left(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_M\right) = A_\alpha\left(\sum_{i=1}^{M} \omega_i \mathbf{p}_i\right) - \sum_{i=1}^{M} \omega_i A_\alpha\left(\mathbf{p}_i\right), \tag{6}$$

where $A_\alpha(\cdot)$ denotes the Arimoto entropy for $\alpha > 0$, $\alpha \neq 1$, and ω_i represent weight factors, such that $\sum_{i=1}^{M} \omega_i = 1$ with $\omega_i \geq 0$. When the parameter α approaches 1, the Arimoto entropy converges to the standard Shannon entropy, and the limitation of the JA divergence in (6) resembles that of the traditional MI obtained by L'Hopital's rule.

Input: Reference image R, floating image F
Stage 1: Given the initialized transformation T_μ, Output is the transformed floating image $F(T_\mu)$.
Stage2: Random sampling strategy was used for R and $F(T_\mu)$, and then output are the coordinates of sample points.
Stage 3: The joint histogram between R and $F(T_\mu)$ was estimated using these sample points. The output is the joint probability distribution of R and $F(T_\mu)$.
Stage 4: The joint probability distribution was adopted to calculate the JA similarity measure between R and $F(T_\mu)$. The output of this stage is the value of JA between R and $F(T_\mu)$.
Stage 5: The simplex optimization method was employed. If the stop criteria of optimization method are satisfied, output is the optimal transformation T^*, otherwise go to **stage 1**.

Algorithm 2: Multimodel medical image registration.

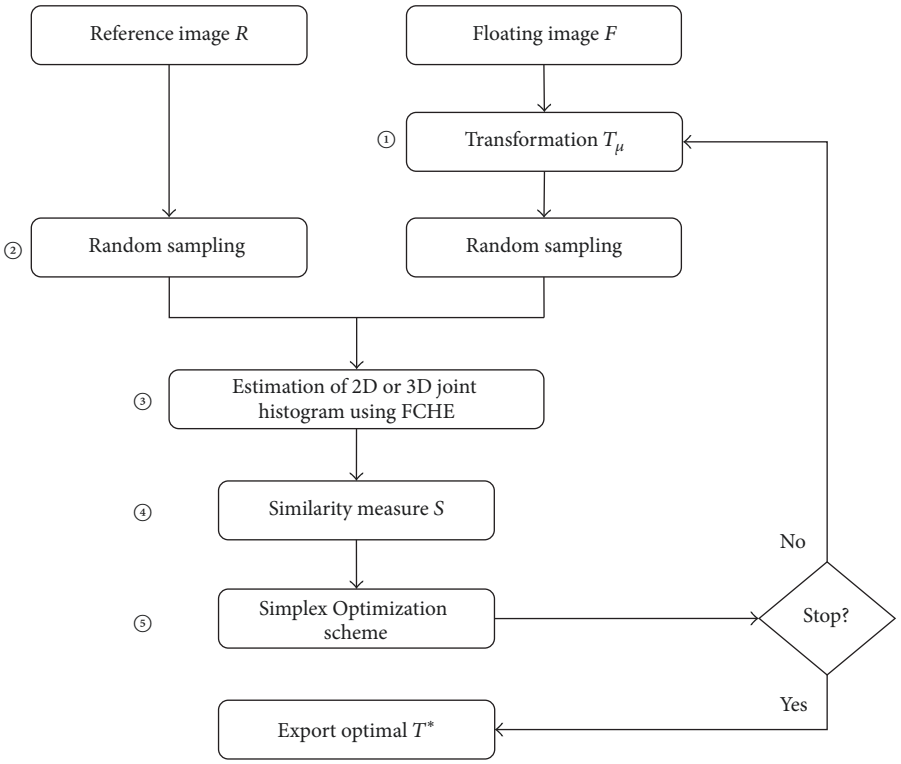


Figure 3: Block diagram of our registration algorithm.

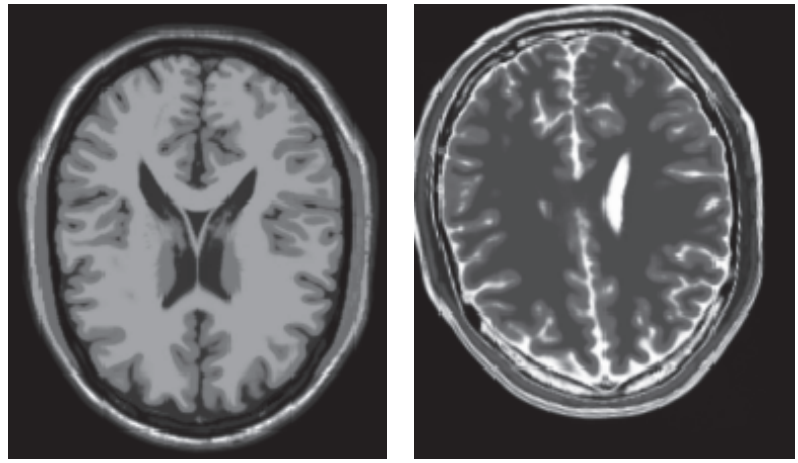
3. Description of the Proposed Method

Herein, we described our medical registration method. First, the FCHE method was applied to estimate the joint histogram of two images to be registered and calculate their joint probability distributions. Then, the JA divergence used as similarity measure was derived in an affine transformation space. Finally, the simplex optimization scheme was employed to search for the maximum of the JA divergence, along with the optimal transformation being obtained. A block diagram of our registration algorithm (Algorithm 2) consisting of five stages is displayed in Figure 3, where the sequence numbers ①–⑤ denote these stages, respectively.

3.1. Registration Algorithm. For two misaligned images to be registered, one is selected as the floating image F and the other is considered as the reference image R. As an example, Figure 4 describes the corresponding slices in T1-weighted magnetic resonance (MR) and T2-weighted MR volumes ((a) denotes the reference image, and (b) denotes the floating image). The goal of image registration is to search the optimal transformation between F and R. We denote x and x' by an arbitrary point in R and its corresponding point in F, respectively. The spatial transformation relation between the corresponding points in R and F can be formulated as

$$x' = T_\mu(x), \tag{7}$$

where T_μ represents the mapping function and μ is a vector of parameters.



(a) (b)

Figure 4: Two corresponding slices in two registered volumes. (a) Central slice in MR T1 volume. (b) Corresponding slice in the transformed MR T2 volume registered by a known 3D rigid transformation method.

Image registration aimed to align the floating image to the reference image by maximizing a similarity measure between F and R. In other words, the registration problem is formulated as the following optimization problem:

$$T^* = \arg\max_{\mu} S\left(F\left(T_\mu(x)\right), R(x)\right), \tag{8}$$

where S denotes a suitable similarity measure, which achieves its maximum when the reference image and floating image are completely registered, and $F(T_\mu(x))$ stands for the floating image transformed by the space mapping T_μ.

3.2. Transformation Model. Rigid, affine, perspective, and curve (elastic) transformation models were reported to simulate geometric transformations of images [3]. Rajwade et al. [17] reported that the CHEs are not suitable for deformation registration because the derivatives of histograms acquired by CHE are not analytically derived. Therefore, in this paper, we restrict the transformation T_μ to rigid cases. The parameter μ is the vector set of 6 degrees of freedom (DoFs) for 3D rigid transformation. These 6 DoFs include three displacements in the x, y, and z directions and three rotation angles around the x-, y-, and z-axes. Hence, we model the transformation T_μ and rewrite (7) as follows:

$$x' = R \cdot x + t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\cos\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot x + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}, \tag{9}$$

where x' and x hold the same denotations as those in Section 3.1, t_x, t_y, and t_z are the translation parameters, and α, β, and γ denote the rotation parameters. In the following section, we adopt the transformation model defined by (9).

3.3. Similarity Measure. In image registration, the value of the similarity measure for two misaligned images is expected to be maximum when the images are perfectly aligned. Hence, choosing an appropriate similarity measure is an important task in image registration. As mentioned in Section 2.2, the JA divergence can reach the maximum when the difference between two random variables is minimal. Therefore, the JA divergence measure in (6) is applied as a similarity measure, and S is given as follows:

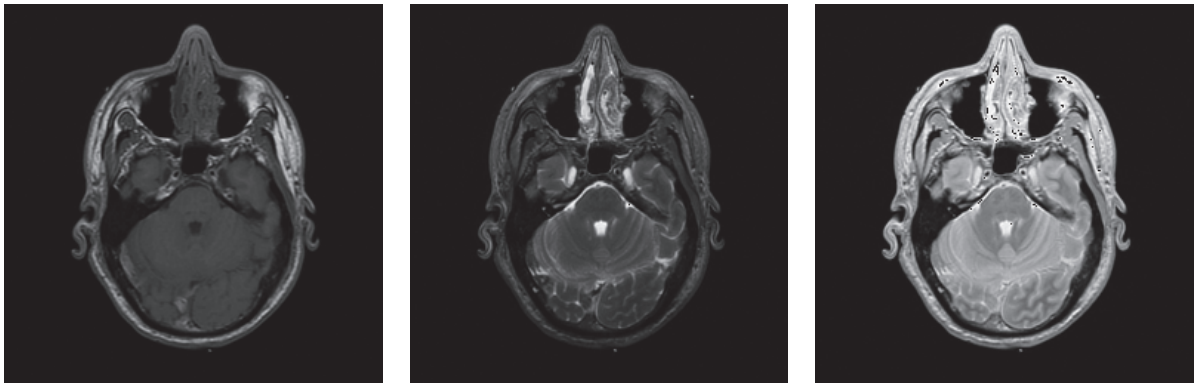
$$S\left(F\left(T_\mu(x)\right), R(x)\right) = \mathrm{JA}_\alpha\left(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_M\right). \tag{10}$$

By combining (8) and (10), we formulate the optimal spatial mapping between $F(T_\mu(x))$ and $R(x)$ as

$$\begin{aligned} \mu^* &= \arg\max_{\mu} S\left(F\left(T_\mu(x)\right), R(x)\right) \\ &= \arg\max_{\mu} \left\{\mathrm{JA}_\alpha\left(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_M\right)\right\}, \end{aligned} \tag{11}$$

where μ^* denotes the set of estimated parameters and $\mathbf{p}_i = p_i(F(T_\mu(x)) \mid R(x))$, $1 \le i \le M$, represents the conditional probability distribution of the transformed floating image given the reference image.

The intensity values in $F(T_\mu(x))$ and $R(x)$ are denoted by $f = (f_1, f_2, \ldots, f_m)$ and $r = (r_1, r_2, \ldots, r_m)$, respectively. M is the number of the chosen intensity bins for estimating the joint probabilities. By substituting f and r into the conditional probability $p_i(F(T_\mu(x)) \mid R(x))$, we can rewrite



(a) (b) (c)

Figure 5: Real 2D brain images. (a) MR T1 image. (b) MR T2 image. (c) MR PD image.

the formula as $\mathbf{p}_i = p(F = f_j \mid R = r_i) = p(f_j \mid r_i)$. We choose the JA divergence with different weights and α values as the similarity measure S shown in (10). In (6), the reference image is adopted as basis for image registration, and the marginal probability distributions of the reference image are chosen as the weights; that is, $\omega_i = p(R(x)) = p(R = r_i) = p(r_i)$. By substituting ω_i and $\mathbf{p}_i$ into (6), we rewrite the similarity measure S as

$$\begin{aligned} S(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m) &= \mathrm{JA}_\alpha(\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m) \\ &= \frac{\alpha}{\alpha - 1} \left\{ \sum_{i=1}^{m} p(r_i) \left[\sum_{j=1}^{m} p(f_j \mid r_i)^\alpha \right]^{1/\alpha} \right. \\ &\quad \left. - \left[\sum_{j=1}^{m} \left[\sum_{i=1}^{m} p(r_i)\, p(f_j \mid r_i) \right]^\alpha \right]^{1/\alpha} \right\}. \end{aligned} \tag{12}$$

The similarity measure defined in (12) is then maximized using an optimization method to search for the optimal transformation parameters.

3.4. Optimization. The optimization algorithm mainly influences the convergence speed of a similarity measure. Hence, an appropriate optimization method is valuable to a registration framework. We do not obtain the derivatives of the probability distributions using FCHE method. Hence, gradient-based optimization techniques are not available herein. In this work, we exploit the downhill simplex method (Nelder–Mead [31] or Amoeba optimization) to maximize the similarity measure defined in (12), considering that this method only requires function evaluations without need for derivatives.

Notably, the downhill simplex optimization scheme is terminated if the difference between the current function value and the best function value is less than 0.01 and the maximum of difference between the current value and the best value for all parameters is less than 0.1. The scheme is also terminated if the number of iterations reaches a previously fixed value NMAX. In our work, NMAX is set to 100.

4. Experiments and Results

To evaluate the FCHE method, two groups of experiments were implemented on 2D and 3D medical images. In Section 4.1, the datasets used are described. The performance of the FCHE method in registration accuracy is exemplified through 2D registration in Section 4.2. Meanwhile, the method addresses the rigid registration of 3D medical images in Section 4.3.

The FCHE method, CHE algorithm without random sampling, SH, and PZW algorithms were coded by Visual C++ and implemented in Visual Studio 2012. The classes and functions were employed in the Insight Toolkit (ITK) [32]. In this work, all experiments were performed on a 3.10 GHz and 8 GB-memory personal computer. Meanwhile, we set the nonextensive parameter in the JA divergence as $\alpha = 1.50$. We also used the downhill simplex optimization to maximize the JA measure in all experiments.

4.1. Test Images. The datasets for 2D tests were obtained from the Visible Human Project [33]. We chose the head subset, including three protocols of MR images, T1-weighted MR, T2-weighted MR, and proton density- (PD-) weighted MR images (Figure 5) to enforce 2D registration. These images were acquired in the Portable Network Graphics (.png) format and aligned originally with one another at dimensions of 256×256 pixels. 3D registration experiments were carried out on simulated 3D datasets. The simulated data were obtained from the Brainweb database of the Montreal Neurological Institute [34]. These data contained 3D brain MR imaging (MRI) volumes simulated through the following three different protocols: T1-weighted, T2-weighted, and PD-weighted with various slice thicknesses, noise levels, and intensity nonuniformities. In the simulated experiments for 3D, the size of the volumes (with voxel coding on 8 bits) was $181 \times 217 \times 181$ voxels at a voxel size of $1\,\mathrm{mm} \times 1\,\mathrm{mm} \times 1\,\mathrm{mm}$. All corresponding slices in these volumes were aligned with each other.

4.2. Registration for 2D Images. The performance of our registration scheme based on FCHE method was first assessed

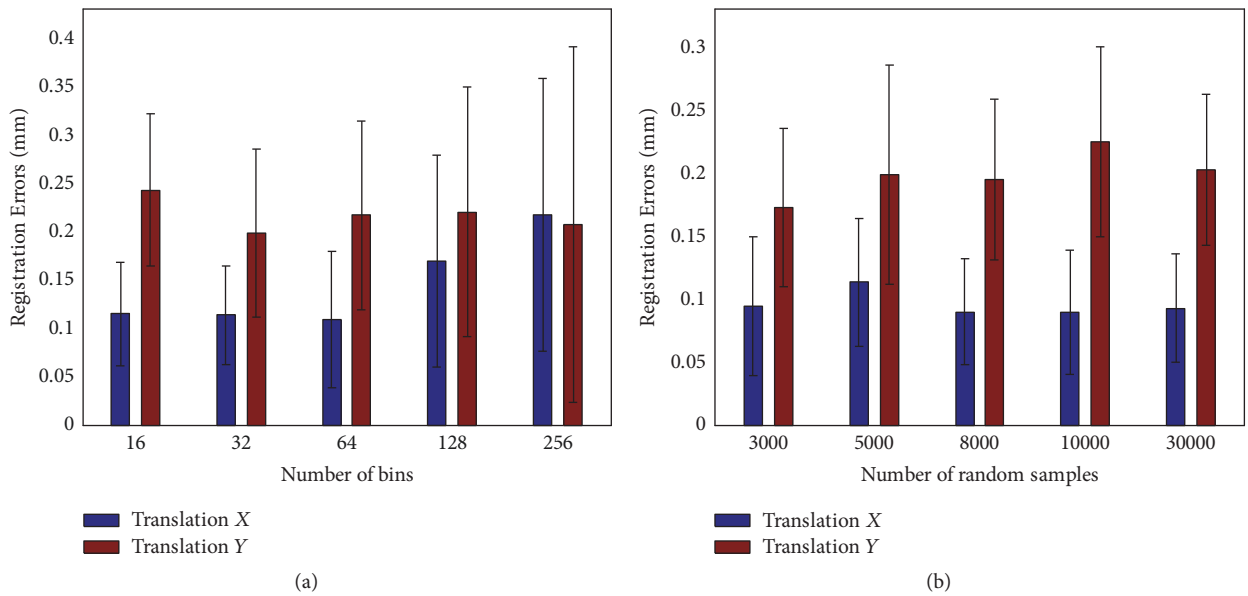


Figure 6: Rigid registration results of the real 2D brain MR T1 and MR PD images. Various values of two parameters m and N were chosen. The transformation parameters were generated randomly from the range 0–10 mm, with the unit of errors also in millimeters. (a) Registration errors for several various m values, $m \in \{16, 32, 64, 128, 256\}$ and $N = 3000$. (b) Results of different numbers of random samples, $N \in \{3000, 5000, 8000, 10000, 30000\}$ with $m = 32$.

using the real 2D brain data. The role of the number of bins and random samples employed in our method was examined and compared with those in the CHE without random sampling, SH, PZW, and particle filter (PF) [35].

4.2.1. Parameter Setting. The registration parameters were chosen by trial and error from the real data. To demonstrate the effect of these parameters, we performed the experimental process as follows. Using the pair of MR T1 and MR PD volumes mentioned in Section 4.1, we chose the former volume as reference image. By contrast, the latter volume was randomly transformed and considered as the floating image. Twenty initial transformations were generated, where the x and y coordinates were derived from the range 0–10 mm. To quantitatively evaluate the registration results, we calculated the difference between the true values and the estimated values as the registration error.

Figure 6 displays the mean and standard deviation of the registration errors when varying the number of bins m and the number of random samples N. The lowest mean of the registration error was obtained when $m = 32$ (herein, $N = 3000$) (Figure 6(a)). By contrast, the number of random samples $N = 3000$ achieved the relatively low errors (Figure 6(b)). In the sequel, we used $m = 32$ and $N = 3000$ for 2D images.

4.2.2. Accuracy. To illustrate the registration accuracy of our method, we compared the CHE, SH, PZW, and PR methods with the FCHE in terms of real 2D brain datasets MR T1 and MR PD. In the process, the MR T1 image was also applied as the reference image. We selected 20 affine transformations, which involved displacements along the x and y directions generated randomly from the range 10–20 mm, rotation angles generated randomly from 0° to 20°, and scales generated randomly from 0.9 to 1.1. Consequently, a total of 20 pairs of test images and 100 affine registrations were obtained for the five methods. Table 1 shows the means and standard deviations of the registration errors in all 100 trials that exploited the five algorithms. Notably, the registration errors incurred under the FCHE method are lower than those acquired under the other four methods. Furthermore, the FCHE method improved the execution efficiency in estimating joint distributions compared with the CHE algorithm.

4.2.3. Noise Resistance. This section assesses the performance of our method, as well as the CHE, SH, PZW, and PF algorithms, in the presence of different noise levels. Similar to Section 4.2.1, MR T1 and MR PD images were chosen as the reference and floating images. Three levels of Gaussian noise with zero mean were added to the MR PD image by applying the *imnoise* function of Matlab. The respective variances were 0.01, 0.05, and 0.1. Then, 20 initial affine transformations were obtained to transform each noisy MR PD image. As a result, 20 pairs of test images were acquired for each noise level. Altogether, 300 registrations were carried out for the five approaches.

The resulting registration errors are displayed in Table 2. In each cell, the numbers in round brackets denote the means of the translation errors in the x and y directions. The rotation angles θ with their standard deviations are in square brackets. As shown in Table 2, the registration accuracies of FCHE are superior to those of the other four algorithms for each noise level.

Table 1: Statistics of the registration errors (mm) for 20 trials on the real MR T1 and MR PD brain images shown in Figure 4. The second and fourth columns provide the means of the errors in the x and y directions. The third and fifth columns display the respective standard deviation of the errors incurred under the five methods. The sixth to eleventh columns show the means and standard deviation of rotation and scales errors.

Methods	Errors in the X direction		Errors in the Y direction		Rotation angles		Scales in the X direction		Scales in the Y direction		Average run time (s)
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	
FCHE	0.11	0.05	0.14	0.05	0.07	0.04	0.08	0.06	0.07	0.05	1.58
CHE	0.13	0.07	0.19	0.04	0.08	0.03	0.09	0.05	0.09	0.04	23.6
SH	0.14	0.09	0.20	0.07	0.11	0.07	0.12	0.08	0.13	0.06	1.27
PZW	0.12	0.05	0.22	0.06	0.09	0.05	0.11	0.05	0.12	0.07	1.43
PF	0.11	0.06	0.15	0.05	0.08	0.06	0.09	0.07	0.09	0.05	2.81

Table 2: Mean registration errors (mm) for 20 tests of the real 2D brain MR T1 and MR PD images using three approaches in the presence of different noise levels with their variances $\sigma = 0.01$, 0.05, and 0.1. The numbers in square brackets are the standard deviations of the errors.

Methods	$\sigma = 0.01$	$\sigma = 0.05$	$\sigma = 0.1$
FCHE	$(0.21 \quad 0.29 \quad 0.22)$	$(0.29 \quad 0.31 \quad 0.25)$	$(0.39 \quad 0.54 \quad 0.41)$
	$[0.12 \quad 0.16 \quad 0.19]$	$[0.19 \quad 0.21 \quad 0.15]$	$[0.25 \quad 0.52 \quad 0.32]$
CHE	$(0.20 \quad 0.31 \quad 0.27)$	$(0.34 \quad 0.32 \quad 0.31)$	$(0.39 \quad 0.51 \quad 0.45)$
	$[0.14 \quad 0.17 \quad 0.16]$	$[0.22 \quad 0.20 \quad 0.14]$	$[0.23 \quad 0.50 \quad 0.29]$
SH	$(0.23 \quad 0.32 \quad 0.31)$	$(0.33 \quad 0.39 \quad 0.34)$	$(0.38 \quad 0.71 \quad 0.58)$
	$[0.14 \quad 0.15 \quad 0.17]$	$[0.18 \quad 0.20 \quad 0.15]$	$[0.32 \quad 1.03 \quad 0.55]$
PZW	$(0.24 \quad 0.31 \quad 0.29)$	$(0.31 \quad 0.33 \quad 0.31)$	$(0.52 \quad 0.69 \quad 0.62)$
	$[0.24 \quad 0.14 \quad 0.13]$	$[0.21 \quad 0.22 \quad 0.18]$	$[0.43 \quad 0.44 \quad 0.53]$
PF	$(0.22 \quad 0.31 \quad 0.23)$	$(0.31 \quad 0.32 \quad 0.25)$	$(0.39 \quad 0.57 \quad 0.49)$
	$[0.15 \quad 0.18 \quad 0.16]$	$[0.20 \quad 0.19 \quad 0.16]$	$[0.28 \quad 0.46 \quad 0.38]$

4.3. Registration for 3D Images. To evaluate the performance of the FCHE approach applied to 3D image registration, we compared the method with CHE, SH, and PZW on three groups of simulated 3D brain datasets introduced in Section 4.1, namely, MR T1 and MR T2, MR T1 and MR PD, and MR T2 and MR PD. In each pair of test images, the subvolumes of size $41 \times 41 \times 41$ voxels were exploited to implement the registration experiments along with the former image applied as reference image. Then, the zero-mean Gaussian noise with the variance of 0.05 was added to the latter of each image pair using the *imnoise* function of Matlab. We selected 20 rigid transformations with x, y, and z coordinates generated randomly from the range 0–10 mm to transform the noisy images. Consequently, 20 pairs of test images for each group of data were generated, and a total of 240 3D rigid registrations for the four methods were implemented. Table 3 displays the statistics of the registration errors of the successful experiments in 20 tests conducted on three groups of images using four algorithms. Here, an experiment is regarded as "success" if the registration error in every direction is less than 1 mm. It is observed from Table 3 that the registration accuracies using the FCHE and CHE methods are comparable and better than those employing SH and PZW approaches. Meanwhile, the FCHE method also improved the registration efficiency compared to CHE.

4.4. Discussion of Experimental Results. Experiments were conducted on 2D and 3D datasets; the former used real brain data (MR T1, MR T2, and MR PD images), and the latter included simulated 3D images (MR T1, MR T2, and MR PD volumes). In 2D registration experiments, to compare the registration accuracy of our method with those of other four algorithms, 100 trials were carried out based on affine transformations using parameters randomly generated within a certain range. To access the immunity of noise, three levels of Gaussian noise were added to the floating images and 300 trials were performed. The results obtained for the 2D real images showed that our alignment technique provided lower registration errors and a better robustness to noise than those exploiting SH, PZW, and PF strategies and CHE without random sampling.

For 3D case, three groups of simulated 3D brain datasets, MR T1 and MR T2, MR T1 and MR PD, and MR T2 and MR PD, were employed to evaluate the performance of our method. For each group of test images, 20 pairs of test images were generated and totally 240 3D registrations for four methods were implemented. Experimental results on simulated 3D data show a better behavior of FCHE method with low registration errors when compared to the other methods, and the subvoxel registration accuracies are achieved by FCHE method in almost all cases.

Table 3: Mean and standard deviations (Std.) of 20 registrations of the real 3D brain MR T1 and MR T2, MR T1 and MR PD, and MR T2 and MR PD images using three strategies (noise $\sigma = 0.01$). For each group of tests, the translations parameters in the x-, y-, and z-axes were selected from the range 0–10 mm, and the rotations angles around the x-, y-, and z-axes were selected from the range 0–10 degrees.

Datasets	Methods	Mean	Std.	Run time (s)
MR T1 and MR T2	FCHE	(0.24, 0.31, 0.33, 0.41, 0.49, 0.43)	[0.17, 0.14, 0.16, 0.23, 0.19, 0.25]	87
	CHE	(0.25, 0.32, 0.38, 0.43, 0.52, 0.46)	[0.25, 0.15, 0.19, 0.28, 0.18, 0.22]	1236
	SH	(0.28, 0.37, 0.40, 0.48, 0.56, 0.47)	[0.11, 0.10, 0.13, 0.15, 0.16, 0.21]	1.3
	PZW	(0.26, 0.32, 0.36, 0.46, 0.52, 0.45)	[0.10, 0.09, 0.15, 0.28, 0.21, 0.29]	1.43
MR T1 and MR PD	FCHE	(0.37, 0.35, 0.29, 0.47, 0.51, 0.50)	[0.22, 0.15, 0.14, 0.31, 0.24, 0.23]	86
	CHE	(0.38, 0.36, 0.28, 0.49, 0.55, 0.53)	[0.21, 0.16, 0.17, 0.19, 0.25, 0.34]	1203
	SH	(0.44, 0.41, 0.36, 0.54, 0.59, 0.54)	[0.63, 0.59, 0.21, 0.27, 0.31, 0.38]	1.25
	PZW	(0.42, 0.39, 0.33, 0.52, 0.56, 0.55)	[0.27, 0.19, 0.65, 0.21, 0.36, 0.33]	1.37
MR T2 and MR PD	FCHE	(0.35, 0.33, 0.34, 0.38, 0.41, 0.45)	[0.21, 0.19, 0.36, 0.17, 0.22, 0.19]	98
	CHE	(0.38, 0.36, 0.37, 0.42, 0.44, 0.46)	[0.25, 0.34, 0.26, 0.23, 0.29, 0.25]	1329
	SH	(0.39, 0.42, 0.33, 0.46, 0.52, 0.53)	[0.73, 0.55, 0.18, 0.19, 0.27, 0.30]	1.24
	PZW	(0.37, 0.39, 0.41, 0.41, 0.43, 0.48)	[0.36, 0.18, 0.89, 0.27, 0.31, 0.37]	1.4

5. Conclusion

This study introduced the FCHEs to estimate the joint probability distribution and accelerate the registration process by employing CIR and random samples on nongrid points. Estimation methods for 2D and 3D joint histograms were also elaborated. This work reduced the grid effect caused by grid-aligning transformations between two images to be registered. The reduction of such effect was achieved when a joint histogram was calculated from a subset of pixels randomly selected from nongrid coordinates of the fixed image instead of entire images. Then, a similarity measure based on the Arimoto entropy, called the JA divergence, was introduced for medical image registration. This similarity measure is a generalization of the well-known Jensen–Shannon divergence measure. We employed affine transformation as the parameter space, along with the JA divergence as the registration criteria, to carry out the registration experiments. To search the maxima of this similarity measure, we used the downhill simplex optimization approach. Then, we applied the FCHE estimation algorithm to estimate the joint probability distribution of two images to be registered.

As mentioned above, the similarity measure computed by adopting the FCHE estimation algorithm was not differentiable, which is an important property in nonrigid registrations. As a result, gradient-based optimization schemes were not available for the FCHE-based registration method. In the future, we intend to extend our application to other multimodal medical images, such as ultrasound and MR images, and to other organs.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper. They also solidly confirm that the mentioned received grants in Acknowledgments did not lead to any conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of Henan Province under Grant 162300410338, by the National Natural Science Foundation of China under Grant 61772576, by Key Scientific Research Projects of Henan Province under Grant 17A510006, by Scientific and Technological Program Projects of Henan Province under Grant 172102210071, by the project sponsored by SRF for ROCS, SEM, by Natural Science Foundation of Jiangsu Province under Grants DZXX-031 and BY2014127-11, and by the Qing Lan Project.

References

[1] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[2] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine and Biology*, vol. 46, no. 3, pp. R1–R45, 2001.

[3] J. P. W. Pluim, J. B. A. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[4] A. Collignon, F. Maes, D. Delaere et al., "Automated multi-modality image registration based on information theory," *Information Processing in Medical Imaging*, vol. 3, no. 6, pp. 263–274, 1995.

[5] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[6] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.

[7] P. Viola and W. M. Wells III, "Alignment by Maximization of Mutual Information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.

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

[9] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information," *IEEE Transactions on Medical Imaging*, vol. 19, no. 8, pp. 809–814, 2000.

[10] F. Wang, B. C. Vemuri, M. Rao et al., "Cumulative residual entropy, a new measure of information & its application to image alignment," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pp. 548–553, IEEE, Nice, France, 2003.

[11] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang, "Cumulative residual entropy: a new measure of information," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 50, no. 6, pp. 1220–1228, 2004.

[12] F. Wang and B. C. Vemuri, "Non-rigid multi-modal image registration using cross-cumulative residual entropy," *International Journal of Computer Vision*, vol. 74, no. 2, pp. 201–215, 2007.

[13] A. Rajwade, A. Banerjee, and A. Rangarajan, "Continuous image representations avoid the histogram binning problem in mutual information based image registration," in *Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 840–843, 2006.

[14] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2083–2099, 2000.

[15] N. Dowson, T. Kadir, and R. Bowden, "Estimating the joint statistics of images using nonparametric windows with application to registration using mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1841–1857, 2008.

[16] N. Joshi, T. Kadir, and S. M. Brady, "Simplified computation for nonparametric windows method of probability density function estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1673–1680, 2011.

[17] A. Rajwade, A. Banerjee, and A. Rangarajan, "Probability density estimation using isocontours and isosurfaces: Applications to information-theoretic image registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 475–491, 2009.

[18] E. López-Rubio, "A histogram transform for probabilitydensity function estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 644–656, 2014.

[19] E. López-Rubio and J. Muñoz-Pérez, "Probability density function estimation with the frequency polygon transform," *Information Sciences*, vol. 298, pp. 136–158, 2015.

[20] H.-M. Chen and P. K. Varshney, "Mutual Information-Based CT-MR Brain Image Registration Using Generalized Partial Volume Joint Histogram Estimation," *IEEE Transactions on Medical Imaging*, vol. 22, no. 9, pp. 1111–1119, 2003.

[21] B. Li, G. Yang, H. Shu, and J. L. Coatrieux, "A new divergence measure based on Arimoto entropy for medical image registration," in *Proceedings of the 22nd International Conference on Pattern Recognition, ICPR 2014*, pp. 3197–3202, Stockholm, Sweden, August 2014.

[22] S. Darkner and J. Sporring, "Locally orderless registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1437–1450, 2013.

[23] S. Darkner and J. Sporring, "Generalized partial volume: An inferior density estimator to parzen windows for normalized mutual information," in *Biennial International Conference on Information Processing in Medical Imaging*, vol. 6801, pp. 436–447, Springer, Berlin, Germany, 2011.

[24] P. Thévenaz, M. Bierlaire, and M. Unser, "Halton sampling for image registration based on mutual information," *Sampling Theory Signal Image Process*, vol. 7, no. 2, pp. 141–171, 2008.

[25] W. H. Press, B. P. Flannery, S. A. Teukolsky et al., *Numerical Recipes in C*, Cambridge University Press, Cambridge, UK, 3rd edition, 2007.

[26] R. Bhagalia, J. A. Fessler, and B. Kim, "Accelerated nonrigid intensity-based image registration using importance sampling," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1208–1216, 2009.

[27] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[28] S. Arimoto, "Information-theoretical considerations on estimation problems," *Information and Control*, vol. 19, no. 3, pp. 181–194, 1971.

[29] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 52, no. 10, pp. 4394–4412, 2006.

[30] J. Lin, "Divergence measures based on the Shannon entropy," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, 1991.

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

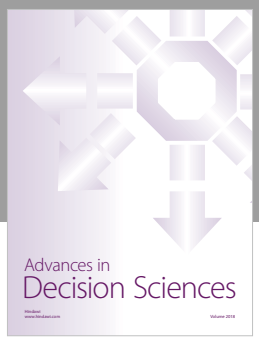
[32] L. Ibß, W. Schroeder, L. Ng, and L. Ibáñez, *The ITK Software Guide Kitware*, 2005.

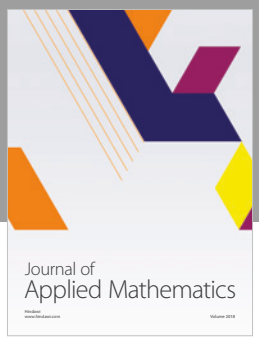
[33] The Visible Human Project, http://www.nlm.nih.gov/research/visible/getting_data.html.

[34] BrainWeb: Simulated Brain Database, http://brainweb.bic.mni.mcgill.ca/brainweb/.

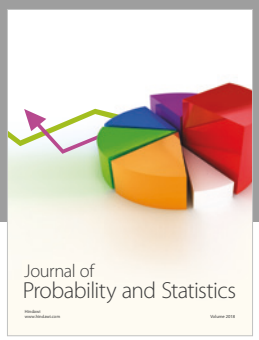
[35] E. R. Arce-Santana, D. U. Campos-Delgado, and A. Alba, "Affine image registration guided by particle filter," *IET Image Processing*, vol. 6, no. 5, pp. 455–462, 2012.

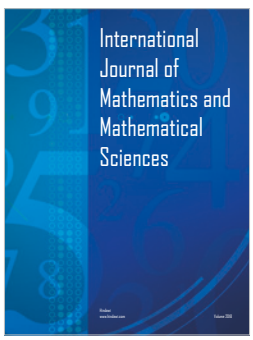








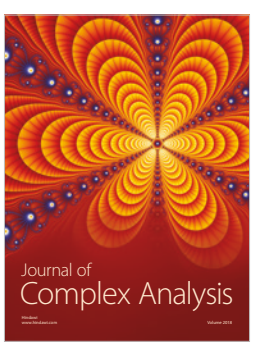


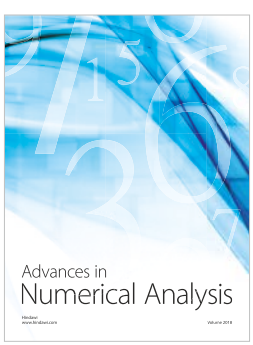


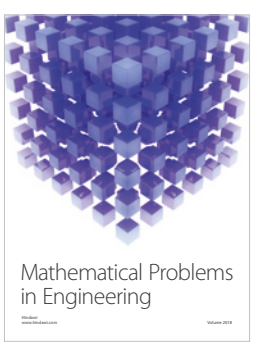




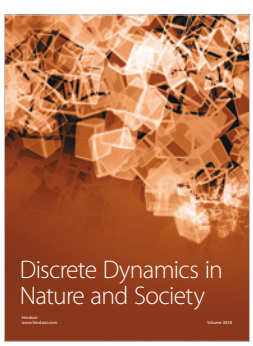




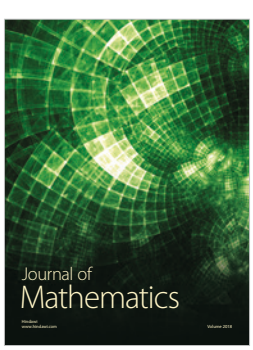




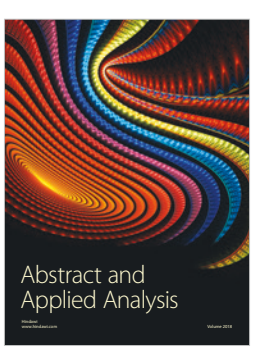


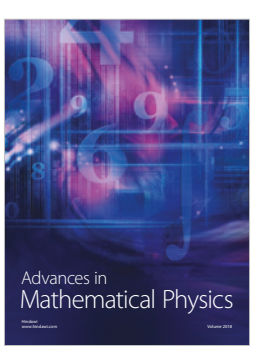












Hindawi

Submit your manuscripts at
www.hindawi.com