Deformable multi-modal registration using 3D-FAST conditioned mutual information

Xueli Liu\textsuperscript{a,b}, Zhixian Tang\textsuperscript{a,b}, Manning Wang\textsuperscript{a,b} and Zhijian Song\textsuperscript{a,b}
\textsuperscript{a}Digital Medical Research Center, Fudan University, Shanghai, China; \textsuperscript{b}Shanghai Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention, Shanghai, China

ABSTRACT

Purpose: Mutual information (MI) has been a preferred choice of similarity measure for multi-modal image registration, but the accuracy and robustness of MI are not satisfactory as MI only considers the global intensity correlation while ignoring local and structural information. To address this problem, we combine MI with local and structural information.

Method: We bring structural information extracted by a modified Accelerated Segment Test (FAST) algorithm into MI. Traditional FAST is transferred into 3D for the first time, and the 3D-FAST based structural information is added into MI as another channel, thereby incorporating spatial and geometric information with intensity information in the registration.

Result: The robustness and accuracy of the proposed method were demonstrated in three experiments. The average registration errors of our method were 1.17, 1.33 and 1.20 compared to 1.47, 1.63 and 1.40 of LMI in T1-T2, T1-PD and T2-PD registration respectively.

Discussion: In this paper, we use the structural similarity computed by 3D-FAST as the conditional information to encode spatial and geometric cues into LMI. In all of these three experiments, our method shows to be more robust and accurate than common registration methods based on information theory.

KEYWORDS

Non-rigid; Multi-modal; Mutual information; FAST; MR; clustering

Introduction

Multi-modal medical image registration is a challenging problem commonly faced in image-guided interventions and image fusion [1]. The main difficulty of the multi-modal image registration comes from the great variability of tissue or organ appearance when imaged with different physical principles, which results in the lack of a general rule to establish structure correspondence. Therefore, efforts to tackle this problem focus mainly on the design of similarity measures. Existing similarity measures in multi-modal image registration can be roughly divided into three categories: (1) similarity measures by matching sparse structural points, (2) transferring different modalities into the same modality and then registering by simple unimodal similarity measures, and (3) similarity measures based on information theory.

In the first category, structural points, such as corners and edges, are extracted from the two images to be registered, then point set registration methods like the Iterative Closest Point (ICP) [2] method, Gaussian Mixture Model (GMM) [3] based point set registration method and Thin Plate Spline-Robust Point Matching (TPS-RPM) [4] can align these structures and register images at the same time. However, it is difficult to extract a sufficient amount of corresponding structural points in some scenarios of the deformable registration of medical images, and the errors between the extracted corresponding points may influence the accuracy of registration.

In the second category, multi-modal images are transferred into the same modality, then a simple similarity measure such as Sum of Absolute Difference (SAD) can register the unimodal images. This category can be further divided into two sub-categories. In the first sub-category, one image is transformed into the modality of the other image. For example, Ultrasound (US) images are simulated from MR/CT images according to the US imaging physics, and then registered to real US images in [5,6]. In the second sub-category, the multi-modal images are transformed into images of an intermediate modality. For example, Wachinger...
et al. [7] transformed multi-modal images into entropy or Laplacian images. Heinrich et al. [8,9] and Jiang et al. [10,11] transferred multi-modal images to unimodal images with local structure descriptor named ‘Modality Independent Neighborhood Descriptor’ (MIND), ‘Self-Similarity Context’ (SSC), ‘discriminative Local Derivative Pattern’ (dLDP) and ‘modality independent Local Binary Pattern’ (mLBP). Although methods in this category have higher registration speed, they may not be applicable for some tissues or organs with less structural information and the modality transformation may lose a large amount of image information.

The third category explores the statistical dependency between the contents (usually intensity) of the two images. Mutual Information (MI), Cross Correlation (CR) and their variants fall into this category [1] and they are usually more accurate and more generally applicable than voxel-wise matching in the categories above [12]. Traditionally, information theory-based methods only consider the statistical relationship of intensity and ignore the spatial and geometric information. Therefore, they are unable to drive small local deformation and often have some problems with spatially-varying intensity inhomogeneity. Several variants have been proposed to overcome these problems. For example, Klein et al. [13] proposed Local Mutual Information (LMI). Rivas et al. [14] proposed a Robust PaTch-based cOrrelation Ratio (RaPTOR) by blocking the intensity values of the fixed image on the joint probability density distribution graph and computing CR locally to resist spatial intensity inhomogeneity. Staring et al. [15] proposed a graph-based implementation of α-Mutual Information (α-MI) using more features. Rivas et al. [16,17] presented a self-similarity conditioned LMI named Contextual Conditioned Mutual Information (CoCoMI) and a self-similarity conditioned α-MI (SeSaMI). Although α-MI, RaPTOR, SeSaMI and CoCoMI are of higher accuracy and robustness, they all need complex pre-processing and are time-consuming (performing self-similarity estimations on a volume of size \(100^3\) voxels takes about 5h on one core of a 3GHz processor before down-sample strategies in [17]). Therefore, how to improve the accuracy while keeping low time complexity is still an open question in the field of multimodal deformable registration.

In this paper, we propose a structure conditioned mutual information using a novel local structure descriptor. Three aspects are addressed to make our method more feasible for 3D deformable registration. First, we extend the traditional 2D corner detector FAST to a 3D structural descriptor using a fast clustering method, which makes the structure description fairly fast. Second, 3D-FAST can decompose the image into two parts: one part with more structures and the other region with less structures. Then our similarity measure is calculated on the region with more structures instead of the whole volume in traditional MI or the randomly selected region in LMI. Third, the similarity of structure is calculated by comparing 3D-FAST which acts as the conditional information of MI (MI on selected regions in our method). Thereby, the proposed method can incorporate spatial and geometric information with intensity information in the registration.

**Background**

**Local mutual information**

Mutual information of a fixed image \(I_f\) and a moving image \(I_m\) can be defined by Equation (1). The two images are geometrically aligned when MI is maximal.

\[
MI(I_f(x), I_m(T(\phi, x))); \Omega = \sum_{i_m} \sum_{i_f} p_{im}(i_f, i_m) \log \frac{p_{im}(i_f, i_m)}{p_{if}(i_f) \times p_{im}(i_m)}
\]

(1)

Where \(\Omega\) is the image domain, \(\phi\) is a collection of deformation parameters, \(T\) is the transformation modeled by \(\phi\), \(p_{if}\) and \(p_{im}\) are the marginal intensity probabilities of \(I_f\) and \(I_m\), and \(p_{im}\) is the joint intensity probability of \(I_f\) and \(I_m\). As MI only considers the global intensity distribution, it is easy to be disturbed by local deformation or inhomogeneity. Klein et al. [13] proposed LMI by adding spatial information in Equation (2).

\[
LMI(I_f(x), I_m(T(\phi, x))); \Omega_j = \frac{1}{N} \sum_{j=1}^{N} MI(I_f(x), I_m(T(\phi, x))); \Omega_j
\]

(2)

Where \(\Omega_j \subset \Omega\) are spatial neighborhoods, and \(N\) is the number of these neighborhoods which are randomly selected in [13].

**Fast**

FAST is a corner detector [18] proposed by Rosten et al, the advantage of FAST is the higher speed than other well established corner detection algorithms such as Harris, SURF, SUSAN.

Select a pixel \(P\) (indicated by a red star in Figure 1) with grey value of \(I_P\) which is to be identified as an interest point or not. Taking the neighbor circle of the...
In this section, we will elaborate the proposed method denoted as FastMI. FastMI takes structural similarity as the conditional information of MI. The similarity measure (structure conditioned MI) of the registration in our method is described in section 3.1, followed by the modified 3D-FAST in section 3.2, which is used to measure the similarity of structures. Section 3.3 gives the final cost function. Section 3.4 details the optimization of the cost function. The framework of the registration using FastMI is summarized in Table 1.

Table 1. The framework of the registration using FastMI.

| Step | Description |
|------|-------------|
| 1    | Select N structural regions using 3D-FAST. (section 3.2) |
| 2    | Calculate the structural similarity \(p_{\text{ufm}}\) of corresponding voxels in the overlap regions. (section 3.2) |
| 3    | Calculate FastMI as the objective function for the registration. (section 3.1 and section 3.3) |
| 4    | Optimize the cost function using the gradient descent method to obtain the optimal deformation parameters. (section 3.4) |

Structure conditioned MI

When considering the structural similarity as an extra ‘channel’ of intensity-based mutual information between images under registration, information theory can be used to describe the relationship between the three types of information [22]. The three channel mutual information is given by [23] in Equation (3).

\[
\text{CMI} \left( I_f(x), I_m(T(\phi, x)), u_{\text{ufm}}(x) : \Omega \right) = \left[ H(I_f(x)) + H(I_m(T(\phi, x))) + H(u_{\text{ufm}}(x)) \right] - H(I_f(x), I_m(T(\phi, x)), u_{\text{ufm}}(x)) \\
= \sum_{x} \sum_{i_f} \sum_{i_m} p(i_f, i_m, u_{\text{ufm}}(x)) \log \frac{p(i_f, i_m, u_{\text{ufm}}(x))}{p(i_f) \times p(i_m) \times p(u_{\text{ufm}}(x))} 
\]

(3)

Where \(p(i_f, i_m, u_{\text{ufm}}(x))\) is the joint probability of intensities of the fixed image \(I_f\) and the moving image \(I_m\) co-occurring with the similarity of structure (defined by \(u_{\text{ufm}}(x)\) and will be explained in section 3.2), \(i_f\) and \(i_m\) are the bins of intensity distribution, and \(x\) is the voxel of the overlap region.

As \(u_{\text{ufm}}(x)\) is only related to the corresponding voxels in the two images, we can assume that the marginal probability \(p(u_{\text{ufm}}(x))\) is constant at each voxel \(x\). The joint probability \(p(i_f, i_m, u_{\text{ufm}}(x))\) can be replaced by Equation (4) in the form of conditioned probability.

\[
p(i_f, i_m, u_{\text{ufm}}(x)) = p_{\text{ufm}}(i_f, i_m) \cdot p(u_{\text{ufm}}(x)) 
\]

(4)

Simplify Equation (3) with Equation (4), the three channel mutual information is given by Equation (5).

\[
\text{CMI} \left( I_f(x), I_m(T(\phi, x)), u_{\text{ufm}}(x) : \Omega \right) = \sum_{i_m} \sum_{i_f} \sum_{x} p_{\text{ufm}}(i_f, i_m) \log \frac{p_{\text{ufm}}(i_f, i_m)}{p(i_f) \times p(i_m)} 
\]

(5)

Where \(p_{\text{ufm}}(i_f, i_m)\) is the structure conditioned probability, and \(p(u_{\text{ufm}}(x))\) is taken off as a constant. Quantitative-qualitative measure [24] of information is used in the calculation of \(p_{\text{ufm}}(i_f, i_m)\). This measure consists of two aspects of an event: the quantitative one and the qualitative one. The quantitative one is related to the probability of occurrence, and the qualitative one is related to its utility for the fulfillment of the goal. In image registration, images are better aligned when their local structure is more similar.
therefore, structure similarity \( u_{im}(x) \) acts as the qualitative one of the probability of joint intensity. The contribution of voxel \( x \) to point \((l_i, l_m)\) of the structured conditioned joint intensity histogram is defined by Equation (6).

\[
h_{x}(l_i, l_m) = u_{im}(x)\int \left( \frac{l_i - l_i(x)}{\rho_p} \right) \int \left( \frac{l_m - l_m(T(\phi, x))}{\rho_p} \right)
\]

(6)

where \( \chi \) is the Gaussian Parson window, and \( \rho_p \) is the width of the window. The similarity measure (the structure conditioned MI) of FastMI is given by Equation (7) where \( p_{um}(l_i, l_m) \) is calculated in Equation (8).

\[
CMI(l_i(x), l_m(T(\phi, x)), u_{im}(x); \Omega) = \sum_{l_i} \sum_{l_m} p_{um}(l_i, l_m) \log \frac{p_{um}(l_i, l_m)}{p(l_i) \times p(l_m)}
\]

(7)

\[
p_{um}(l_i, l_m) = \frac{1}{\Omega} \sum_{x \in \Omega} u_{im}(x) \int \left( \frac{l_i - l_i(x)}{\rho_p} \right) \int \left( \frac{l_m - l_m(T(\phi, x))}{\rho_p} \right)
\]

(8)

**Modified 3D-FAST**

For the measurement of structural similarity, we proposed the modified 3D-FAST algorithm which extends FAST on 2D images to that on 3D volumes. A fast clustering method [25] was adapted to detect contiguous regions on the spherical surface fulfilling FAST criterions instead of contiguous lines on a circle in traditional 2D FAST.

As 3D-FAST can split the image into a set of non-overlapping regions with more or less structural information, we perform the registration on spatially meaningful regions with more structural information instead of the randomly selected region in Equation (2). This modified 3D-FAST algorithm has two goals: (1) calculate the structural similarity of two voxels, and (2) find local neighborhoods with more structural information for the computation of FastMI. 3D-FAST achieves these two goals by the following steps:

1. For each voxel \( x \) in overlap regions of the two images, we select 90 blocks weighted by Gaussian kernel with center points on the spherical surface surrounding \( x \). Set a threshold intensity value \( M \), and then \( x \) can be encoded as a vector \( b \) composed of 90 elements in Equation (9). \( b_n (n = 1, 2, \ldots, 90) \) in Equation (9) is calculated by Equation (10), where \( a_n \) is the absolute difference between the \( n \)th patch’s Gaussian weighted intensity and the intensity of the central voxel \( x \).

\[
b = \{b_1, b_2, \ldots, b_{90}\} \quad (9)
\]

\[
b_n = \begin{cases} 0 & (a_n < M) \\ 1 & (a_n \geq M) \end{cases} \quad (10)
\]

2. Select six elements of vector \( b \) on the sagittal, coro- 

dinary and vertical axis of voxel \( x \) and form a new vector \( c \) in Equation (11). Voxel \( x \) is detected to be with more structural information when satisfying Equation (12) where \( K \) is an integer from one to six set by users.

\[
c = \{c_1, c_2, \cdots, c_6\} \quad (11)
\]

\[
\sum_{i=1}^{6} c_i > K \quad (12)
\]

3. Build a lookup table \( Q \) for finding the spatial distance between these blocks according to their indexes in vector \( b \). \( Q \) is independent of images and can be calculated before registration.

4. Find the indexes of nonzero elements in \( b \) and form a distance matrix \( D \) of these indexes using the lookup table \( Q \). The matrix \( D \) is then used to get the maximal contiguous region of the spherical surface surrounding \( x \) with a fast clustering method on Science 2014 [25].

5. Count the number of blocks in the maximal contiguous region as the structural value of \( x \). The structural similarity \( u_{im}(x) \) of \( l_i \) and \( l_m \) at voxel \( x \) is defined by the Gaussian distance of their structural value in Equation (13), Where \( F_{i}(x) \) and \( F_{m}(x) \) are the structural value of \( l_i \) and \( l_m \) at voxel \( x \), \( \alpha \) is the weight of structural information.

\[
u_{im}(x) = 1 - \alpha \left( 1 - \chi \left( \frac{F_i(x) - F_m(x)}{\rho_u} \right) \right) \quad (13)
\]

**Structure conditioned MI using modified 3D-FAST (FastMI)**

The objective function for the registration is given by Equation (14). The first term FastMI is the similarity measure. The second term \( \omega \) is the regularization penalty item with \( \nabla \) being the gradient operator of transformation field \( T \).

\[
C(l_i(x), l_m(T(\phi, x))) = \text{FastMI}(l_i(x), l_m(T(\phi, x))) + \omega(\nabla T^T \nabla T) \quad (14)
\]

In this paper, cubic B-spline functions are used to model the transformation in Equation (15) where \( i, j \) and \( k \) are the indexes of the control nodes, and \( B \) represents the B-spline basis functions (see [26] for more details).
\[ I_m(T(\phi, x)) = \sum_{a=0}^{3} \sum_{b=0}^{3} \sum_{c=0}^{3} B_a(x)B_b(\beta)B_c(\gamma)i_{a+b+c} \tag{15} \]

Image blocks \( \Omega_j \subset \Omega \) are regions with more structural information detected by 3 D-FAST are the neighborhoods for FastMI calculation as in Equation (16) where \( N \) is the number of this kind of neighborhoods.

\[
\text{FastMI}(I_I(x), I_m(T(\phi, x))) = \frac{1}{N} \sum_{j=1}^{N} \text{CMI}(I_I(x), I_m(T(\phi, x)), u_{im}(x); \Omega_j) \tag{16}
\]

**Optimization**

Gradient descent method [27] is adopted in this paper for the optimization of the objective function. The update equation of the gradient descent method is shown in Equation (17) where \( a_t \) is the step size and \( \nabla_{\Delta \phi} C \) is the gradient of the objective function w.r.t. \( \phi \) at the current position \( \phi_t \).

\[
\phi_{t+1} = \phi_t - a_t \nabla_{\Delta \phi} C \tag{17}
\]

For the gradient of the objective function, we follow work in [28] and derive a Taylor expansion of FastMI as in Equation (18) where \( N \) is the number of parameters.

\[
\text{FastMI}(I_I(x), I_m(T(\phi, x))) = \text{FastMI}(I_I(x), I_m(T(\phi, x))) + \sum_{i=1}^{N} \frac{\partial \text{FastMI}(I_I(x), I_m(T(\phi, x)))}{\partial \phi_i} (\Delta \phi_i) + \cdots \tag{18}
\]

Then we simplify Equation (18) by ignoring all terms above the second-order and found the gradient of FastMI as:

\[
\sum_{i=1}^{N} \frac{\partial \text{FastMI}(I_I(x), I_m(T(\phi, x)))}{\partial \phi_i} = \sum_{i=1}^{N} \sum_{l_i} \sum_{l_m} -\frac{p_{nm}(l_i, l_m)}{\partial \phi} \cdot \log \frac{p_{nm}(l_i, l_m)}{p(l_m)} \tag{19}
\]

A multi-resolution scheme is used to represent coarse-to-fine details of both volumes for fast and robust registration during the optimization.

**Results**

Three experiments were performed to demonstrate the accuracy and robustness of the proposed method compared with other information theory based methods. The algorithm was implemented using MATLAB with mex file compiled from C. All experiments were performed on a workstation with Intel Xeon CPU (clock speed 3.30 GHz) and 32 G memory. As medical images are usually noisy and not normalized, an intensity normalization is applied mostly before registration. Therefore, intensity normalization is applied in all experiments below.

**Accuracy experiment with noise and intensity inhomogeneity**

In this experiment, we compared the performance of FastMI to LMI in the presence of noise and intensity inhomogeneity. Images under registration are shown in Figure 2. For each image pair under registration (Figure 2(a) and Figure 2(b), Figure 2(c) and Figure 2(d)), we moved the right one along the horizontal and the vertical axis and calculated the similarity value between the image pairs using LMI and FastMI. The results are plotted in Figure 3. It can be observed from Figure 3 that FastMI can still work when there are severe noise and intensity inhomogeneity while LMI fails, and FastMI has higher convergence speed than LMI. The edge of the leaf in Figure 2(a) and Figure 2(d) detected by 3 D-FAST (can be used in 2 D images) are shown in Figure 4 which proved that 3 D-FAST was robust to noise and intensity inhomogeneity.

In this experiment, 3 D-FAST was performed using 16 neighborhoods with a 3 x 3 Gaussian filter mask.
and a threshold of \( M = 0.5 \) in Equation (10). The window’s width of structural distance \( \rho_u \) in Equation (13) is set to be 4. The parameters were chosen and tuned empirically to achieve the best performance.

**Robustness experiment with local deformation**

In this experiment, we compared the performance of FastMI and MI in the presence of local deformation. We deformed a T1-MR image with dimension of 100 × 100 on the B-spline nodes in a local region as shown in Figure 5. A random displacement in the range of -5 to 5 in each dimension was added to each node of the local region. The images before and after deformation were used as the moving and reference images respectively (see in Figure 6). Figure 7 shows the normalized horizontal and vertical displacement of pixels in the moving image after registration with MI and FastMI. It can be seen that the deformation field obtained from FastMI is better confined in a local region corresponding to the true deformation.

In this experiment, 3D-FAST was performed using 16 neighborhoods with a 3 × 3 Gaussian filter mask.
and a threshold of $M = 0.2$ in Equation (10). The window’s width of structural distance $\rho_u$ in Equation (13) is set to be 4. The parameters were chosen and tuned empirically to achieve the best performance.

**Accuracy experiment on 3D images**

In this experiment, we compared the accuracy of FastMI with LMI and Normalized Cross Correlation (NCR) in three groups (eight registrations in each group) of pairwise registration between T1-T2, T1-PD and T2-PD images from the BrainWeb database. All images used in this experiment were with 3% noise and 20% intensity inhomogeneity. The moving images in each group were obtained by deforming the original image with randomly generated deformation fields while the images of the other modalities were registered to them. In each registration, the mean Target Registration Error (TRE) is calculated at 10000 points detected to be structural using the 3D-FAST detector in our work as shown in Equation (20) where $\| \cdot \|$ is the square operator, $n$ is the number of these structural points.

$$mTRE = \frac{1}{n} \sum_{i} \| T(x_i) - x'_i \|$$  \hspace{1cm} (20)

In the FastMI registration, 3D-FAST was performed using 90 neighborhoods with a $3 \times 3$ Gaussian filter mask and a threshold of $M = 0.1$ in Equation (10). The window’s width of structural distance $\rho_u$ in Equation (13) is set to be 20. The parameters were chosen and tuned empirically to achieve the best performance.

**Figure 8** lists the TRE of each group (eight registrations in each group) by performing registration with CC, LMI and our method, where FastMI shows higher accuracy than other methods with the same terminate conditions. **Figure 9** shows the results of overlaying the 90th slice of the moving image on that of the fixed image before and after registration with different methods in the first registration of T1-T2 image pairs. Some clearly improved alignments at the ventricles and the gyri can be observed with our method (indicated by arrows).

**Discussion**

In the context of deformable multi-modal medical image registration, MI may lead to undesired results. In this paper, we propose a new algorithm named FastMI, which can combine the structural, spatial and intensity information for image registration. In our work, the structural similarity calculated by 3D-FAST was added into MI as the conditional information of intensity correlation. It is also possible to condition other correlation such as entropy and gradients of intensity by this structural similarity.
In our experiments, FastMI achieved better performance than common multi-modal registration when there were noise, intensity inhomogeneity and local deformation. As we can see from Figures 3 and 4 in the first experiment, FastMI is robust to noise and intensity inhomogeneity, and has higher convergence speed. The reason is that 3D-FAST can detect structural voxels and is not easily influenced by noise and inhomogeneity as showed in Figure 4(b). In addition, the structural similarity $u_{fm}$ is more sensitive to the local misalignment than the intensity correlation similarity.

There are two reasons for the better performance of FastMI on local deformation in the second experiment. Firstly, FastMI is calculated on local regions with more structural information. So it performs better than MI which is calculated on the whole image and LMI which is calculated on random selected local regions. Secondly, LMI only considers the statistical intensity relationships between both images and may have the same intensity relationships when the voxels in the images have different intensity, while FastMI are more robust by considering both intensity and structural relationships.

Compared with other variants of MI using structural information, such as SeSaMI and CoCoMI, FastMI is faster in structural detection as it doesn’t need time-consuming preprocessing steps, and 3D-FAST can analyze images in high speed by taking $30 \sim 45$ sec on a volume with dimension of $217 \times 181 \times 181$ on one core of a 3.3 GHz processor (self-similarity estimations on a volume of size $100^3$ takes about 5 h on one core of a 3 GHz processor before down-sample strategies in SeSaMI and CoCoMI).

Additionally, the parameters can influence the performance of registration. How to choose proper parameters is a crucial problem. For example, we may ignore some detailed structure of image if threshold $T$ is too large while the registration might be influenced by noise if $T$ is too small. The window’s width of
structural distance $\rho_u$ cannot be too large because the structural information might be over-smoothed and it cannot be too small because the structural information might be influenced by noise. And so on.

One limitation of our method is the speed. FastMI cannot be real-time. It is possible to speed it up by GPU, parallel computing or more sophisticated optimization schemes which is our future works.

Conclusion
In this paper, we extend traditional 2D-FAST corner detector to a 3D structure descriptor using a fast clustering method, which is the first FAST method used on 3D images to our best knowledge. And then we used the structural similarity computed by 3D-FAST as the conditional information to encode spatial and geometric cues into LMI. The method we proposed is demonstrated to be more robust and accurate than the common registration methods based on information theory.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
This study has been supported by: the Nation Natural Science Foundation of China (projects 60972102), the National Science and Technology Support Program (No. 2015BAK31B01), the National High Technology Research and Development Program (2015AA020507). This study is partly sponsored by Program of Shanghai Academic/Technology Research Leader project 16XD1424900 and project sponsored by Program of Shanghai Academic/Technology Development Program (2015AA020507). This study is partly sponsored by Program of Shanghai Academic/Technology Research Leader project 16XD1424900 and project 15441905500, 17zr1401500 of Science and Technology Research Leader project 16XD1424900 and project sponsored by Program of Shanghai Academic/Technology Development Program (2015AA020507). This study is partly sponsored by Program of Shanghai Academic/Technology Research Leader project 16XD1424900 and project 15441905500, 17zr1401500 of Science and Technology Commission of Shanghai Municipality.

References
[1] Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: a survey. IEEE Trans Med Imaging. 2013;32:1153–1190.
[2] Yang J, Li H, Campbell D, et al. Go-ICP: a globally optimal solution to 3D ICP point-set registration. IEEE Trans Pattern Anal Mach Intell. 2016;38:2241–2254.
[3] Jian B, Vemuri BC. Robust point set registration using gaussian mixture models. IEEE Trans Pattern Anal Mach Intell. 2011;33:1633–1645.
[4] Cosse A. Diffeomorphic surface-based registration for MR-US fusion in prostate brachytherapy. 2012; p 903–907.
[5] Wein W, Brunke S, Khamene A, et al. Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis. 2008;12:577.
[6] Fuerst B, Wein W, Müller M, et al. Automatic ultrasound-MRI registration for neurosurgery using the 2D and 3D LC(2) Metric. Med Image Anal. 2014;18:1312–1319.
[7] Wachinger C, Navab N. Entropy and Laplacian images: structural representations for multi-modal registration. Medical Image Analysis. 2012;16:1–17.
[8] Heinrich MP, Jenkinson M, Bhushan M, et al. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. Med Image Anal. 2012;16:1423–1435.
[9] Heinrich MP, Jenkinson M, Papiez BW, et al. Towards realtime multimodal fusion for image-guided interventions using self-similarities. 2013; p 187.
[10] Jiang D, Shi Y, Chen X, et al. Fast and robust multimodal image registration using a local derivative pattern. Med Phys. 2017;44:497.
[11] Jiang D, Shi Y, Yao D, et al. miLBP: a robust and fast modality-independent 3D LBP for multimodal deformable registration. Int J Cars. 2016;11:997–1005.
[12] PWP J, BAM J, AV M. Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging. 2003;22:986–1004.
[13] Klein S, Van-Der-Heide U, Lips I, et al. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med Phys. 2008;35:1407–1417.
[14] Rivaz H, Chen JS, Collins DL. Automatic Deformable MR-Ultrasound Registration for Image-Guided Neurosurgery. IEEE Trans Med Imaging. 2015;34:366–380.
[15] Staring M, VDH U, Klein S, et al. Registration of cervical MRI using multifeature mutual information. IEEE Trans Med Imaging. 2009;28:1412–1421.
[16] Rivaz H, Karimaghaloo Z, Fonov VS, et al. Nonrigid registration of ultrasound and MRI using contextual conditioned mutual information. IEEE Trans Med Imaging. 2014;33:708–725.
[17] Rivaz H, Collins DL. Self-similarity weighted mutual information: a new nonrigid image registration metric. Berlin Heidelberg: Springer; 2012.
[18] Rosten E, Drummond T. Machine learning for high-speed corner detection. 2006; p 430–443.
[19] Rosten E, Porter R, Drummond T. Faster and better: a machine learning approach to corner detection. IEEE Trans Pattern Anal Mach Intell. 2008;30:105–119.
[20] Yu TH, Kim TK, Cipolla R. Real-time action recognition by spatiotemporal semantic and structural forests. 2010; p 1–12.
[21] Koelstra S, Patras I. The fast-3D spatio-temporal interest region detector. 2009; p 242–245.
[22] Reza FM. An introduction to information theory. New York, USA: McGraw-Hill Book Company, Inc; 1961.
[23] Wachinger C, Navab N. Simultaneous Registration of Multiple Images: Similarity Metrics and Efficient Optimization. IEEE Trans Pattern Anal Mach Intell. 2013;35:1221–1233.
[24] Luan H, Qi F, Xue Z, et al. Multimodality image registration by maximization of quantitative-qualitative measure of mutual information. Pattern Recogn. 2008;41:285–298.
Rodriguez A, Laio A. Machine learning. Clustering by fast search and find of density peaks. Science. 2014;344:1492.

Lee S, Wolberg G, Shin SY. Scattered data interpolation with multilevel B-splines. IEEE Trans Visual Comput Graphics. 1997;3:228–244.

Klein S, Staring M, Pluim JP. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. IEEE Trans Image Process. 2007;16:2879.

Studholme C, Drapaca C, Iordanova B, et al. Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. IEEE Trans Med Imaging. 2006;25:626–639.