Rapid and robust medical image elastic registration using mean shift algorithm

Xuan Yang (杨 烜)[1] and Jihong Pei (裴继红)[2]

[1] *College of Information Engineering, Shenzhen University, Shenzhen 518060*

[2] *Intelligent Information Processing Laboratory, Shenzhen University, Shenzhen 518060*

Received February 18, 2008

In landmark-based image registration, estimating the landmark correspondence plays an important role. In this letter, a novel landmark correspondence estimation technique using mean shift algorithm is proposed. Image corner points are detected as landmarks and mean shift iterations are adopted to find the most probable corresponding point positions in two images. Mutual information between intensity of two local regions is computed to eliminate mis-matching points. Multi-level estimation (MLE) technique is proposed to improve the stability of corresponding estimation. Experiments show that the precision in location of correspondence landmarks is exact. The proposed technique is shown to be feasible and rapid in the experiments of various mono-modal medical images.

OCIS codes: 100.6950, 100.0100.

doi: 10.3788/COL20080612.0950.

For elastic registration based on landmarks, a challenging work is to estimate corresponding landmarks automatically from images[1−3]. In general, landmarks could be automatically derived from images. But the difficult problem is establishing correspondence and rejecting non-homologies as outliers[4]. There are three major types of methods to determine the right correspondence: feature-based methods, modal matching methods, and structural description based methods. Rangarajan proposed the softassign Procrustes matching algorithm which iteratively establishes correspondence, rejects non-homologies as outliers between two point sets automatically[4]. However, this method converges slowly and cannot preserve the spatial relation between two objects shapes.

In this letter, we present a landmark correspondence estimation method based on local similarity using mean shift algorithm. Mean shift could search corresponding local regions with similar intensity distribution and converge rapidly. It could be used to estimate corresponding points. We combine the mean shift with the mutual information and select optimal initial position of mean shift to improve the correspondence precision. In our method, multi-level correspondence estimation is introduced in detail also. Finally, thin-plate spline interpolation is done on the corresponding landmarks to obtain registration results.

Given a set $\{x_i \,|i=1,2,\cdots,n\}$ of n points in the d-dimensional space R^d, the sample mean shift with kernel $G(x)$ is defined as

$$M_h(x) = \sum\nolimits_{i=1}^{n} x_i G\left(\left\|\frac{x_i - x}{h}\right\|^2\right) \Bigg/ \sum\nolimits_{i=1}^{n} G\left(\left\|\frac{x_i - x}{h}\right\|^2\right) - x, \qquad (1)$$

where $x_i \in S_h$, S_h is the d-dimensional sphere with radius of h. The mean shift vector is an estimate of the normalized density gradient. The density estimate with kernel $K(x)$ and bandwidth h computed in the point x is given by

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \qquad (2)$$

and the estimate of the density gradient is defined by the gradient of the kernel density estimate as

$$\hat{\nabla} f(x) \equiv \nabla \hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} \nabla K\left(\frac{x - x_i}{h}\right). \qquad (3)$$

The most popular kernel function is Epanechnikov kernel[5], which minimizes the average global error between the estimate and the true density. The estimate of the density gradient using Epanechnikov kernel is

$$\hat{\nabla} f_K(x) = \frac{2}{nh^{d+2}}\left[\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)\right] \times \left[\frac{\sum_{i=1}^{n} x_i g\left(\|(x - x_i)/h\|^2\right)}{\sum_{i=1}^{n} g\left(\|(x - x_i)/h\|^2\right)} - x\right], \qquad (4)$$

where $g(x) = -k'(x)$, $k(x)$ is the Epanechnikov profile[5]. Based on the sample mean shift definition, Eq. (4) contains the sample mean shift vector

$$M_h(x) = \frac{\sum_{i=1}^{n} x_i g\left(\|(x - x_i)/h\|^2\right)}{\sum_{i=1}^{n} g\left(\|(x - x_i)/h\|^2\right)} - x = \frac{h^2}{2/C}\frac{\hat{\nabla} f_K(x)}{\hat{f}_G(x)}, \qquad (5)$$

where $\hat{f}_G(x) = \frac{C}{nh^d}\sum\limits_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)$ is the density estimate at x computed with kernel $G(x) = Cg\left(\|x\|^2\right)$, C

is the normalization constant. It means that the sample mean shift vector is an estimate of the normalized density gradient obtained with Epanechnikov kernel K[5]. The mean shift vector always points toward the direction of maximum increase in the density. The mean shift procedure is defined recursively by computing the mean shift vector $M_{h,G}(x)$ and modifying the center of kernel G. For the Epanechnikov kernel, the convergence of mean shift procedure is guaranteed.

Bhattacharyya coefficient was adopted as the similarity metric between the target modal and the target candidates to track moving objects[5]:

$$\rho\left[p\left(y\right),\ q\right]=\int\sqrt{p_z\left(y\right)q_z}\mathrm{d}z\approx\sum_{u=1}^{m}\sqrt{\hat{p}_u\left(y\right)\hat{q}_u}, \qquad (6)$$

where $\hat{q}_u$ and $\hat{p}_u\left(y\right)$ are the density estimate of the target modal and the target candidate at a given location y, which can be represented by m-bin histogram of the local regions:

$$\hat{q}_u=C\sum_{i=1}^{n}k\left(\|x_i^*\|^2\right)\delta\left[b\left(x_i^*\right)-u\right], \qquad (7)$$

$$\hat{p}_u\left(y\right)=C_p\sum_{i=1}^{n_p}k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\delta\left[b\left(x_i\right)-u\right], \qquad (8)$$

where δ is the Kronecker delta function and function $b(x)$ maps the intensity x into the histogram bins. The Bhattacharyya coefficient can be approximated as

$$\rho\left[\hat{p}\left(y\right),\hat{q}\right]\approx\frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u\left(y_0\right)\hat{q}_u} +\frac{C_p}{2}\sum_{u=1}^{n_p}w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right), \qquad (9)$$

where $w_i=\sum\limits_{u=1}^{m}\delta\left[b\left(x_i\right)-u\right]\sqrt{\hat{q}_u/\hat{p}_u\left(y_0\right)}$. The Bhattacharyya coefficient can be maximized by using mean shift procedure, because the second term represents the density estimate computed with kernel k at location y[5]. When the Bhattacharyya coefficient is maximized, the intensity distribution similarity between the target modal and the target candidate is maximized.

Similar thoughts can be used to register mono-modality images. When two local regions of two images are similar, the intensity distributions of two regions are similar and the Bhattacharyya coefficient is large correspondingly. The mean shift procedure can be used to maximize the Bhattacharyya coefficient by finding the optimal location. The center points of two regions are the correspondence landmarks.

However, the kernel function is isotropy, which means that the density function has no space information. In other words, when two regions are different in space structure, but are similar in intensity distribution, the density functions are similar, and the similarity metric is also large.

The mutual information between two local regions could be used to overcome this disadvantage of Bhattacharyya coefficient. The mutual information represents the space relationships between two regions. If these two regions are mis-matched, the mutual information will be small; otherwise, it is large. The similar metric combined with mutual information is defined as

$$\mathrm{SM}=\rho\left[\hat{p}\left(y^*\right),\hat{q}\right]\times\mathrm{MI}\left(\mathrm{Loc}_q,\mathrm{Loc}_p\left(y^*\right)\right), \qquad (10)$$

where Loc_q and Loc_p are two matched regions obtained by mean shift procedure in the reference image and source image alternatively, y^* is the optimal location in the source image, which is the center of Loc_p. The centers of two local regions are decided as corresponding landmarks if the following condition is satisfied:

$$\mathrm{SM}>\mathrm{Th}_{\mathrm{SM}} \quad \text{and} \quad \rho\left[\hat{p}\left(y^*\right),\hat{q}\right]>\mathrm{Th}_\rho, \qquad (11)$$

where $\mathrm{Th}_{\mathrm{SM}}$ is the threshold of the similar metric and Th_ρ is the threshold of the Bhattacharyya coefficient. We must emphasize the fact that the combined similarity metric is only suitable to mono-modal medical image registration, because the intensity distributions of mono-modal medial images are most identical. In other words, the similar regions in mono-modal images have identical intensity.

The intensity level of the reference image and source image should be decreased to estimate the m-bin histogram of the local regions, and the mapping function $b(x)$ in Eqs. (7) and (8). Moreover, $b(x)$ is of great benefit to the robustness of mutual information. We adopt two-level histogram fuzzy restriction clustering method (TFRCM)[6] to classify intensity of medical images. By TFRCM, anatomical information is represented to discard the redundant noise, and the image intensity quantification is decreased to reduce the number of bins for the joint probability of mutual information, which ensures the robustness of mutual information.

In order to improve the global convergence ability of mean shift, choosing an appropriate initial position of mean shift iteration is important. Suppose $\{q_i\,|i=1,2,\cdots,m\}$ and $\{p_i\,|i=1,2,\cdots,n\}$ are the corner points of two images extracted by Harris corner detection. For each point q_i, all points p_j that are located in the local regions centered at q_i are selected as the candidate initial positions. Mean shift iteration is done to search the convergence points of all these initial positions. The convergence location with the maximum similarity metric is chosen as the corresponding point to q_i.

When the corner points are dense, the points corresponding estimation is sensitive to space distribution of surrounding points. We propose the multi-level estimation (MLE) technique to improve the stability of corresponding estimation. MLE estimates the landmark correspondence by performing iterative processing, which lowers the decision threshold $\mathrm{Th}_{\mathrm{SM}}$ in deterministic annealing. $\mathrm{Th}_{\mathrm{SM}}$ is reduced gradually as $\mathrm{Th}_{\mathrm{SM}}^k=\mathrm{Th}_{\mathrm{SM}}^{k-1}\cdot r$, where r is the annealing rate. At each iteration step, the corresponding landmarks are estimated using different threshold.

Suppose $\{q_i\,|i=1,2,\cdots,m\}$ is the reference corner point. The procedure works as follows.

1) Initialize the threshold $\mathrm{Th}_{\mathrm{SM}}^0$ and the minimum value of the threshold $\mathrm{Th}_{\min}$, Let $\left\{q_i^0\right\}$ =

$\{q_i\,|i=1,2,\cdots,m\}$, $R_q^0=\{q_i^0\}$, $k=1$.

2) Do landmark correspondence estimation for R_q^{k-1} with $\mathrm{Th}_{\mathrm{SM}}^{k-1}$. The corresponding points of $\{q_i^k\,|i=1,2,\cdots,m_k\}$ are estimated.

3) Get the residual point set $R_q^k=\{q_i^{k-1}\}-\{q_i^k\}$.

4) Derive the new threshold $\mathrm{Th}_{\mathrm{SM}}^k=\mathrm{Th}_{\mathrm{SM}}^{k-1}\cdot r$.

5) While $\mathrm{Th}_{\mathrm{SM}}^k>\mathrm{Th}_{\min}$, do $k=k+1$.

6) Go to Step 2).

MLE extracts the corresponding landmarks with different importance at each iteration step. $\{q_i^0\}$ represents all feature points whose surrounding local regions are similar to the corresponding regions in the source image to a great extent. After deleting $\{q_i^0\}$ from $\{q_i\,|i=1,2,\cdots,m\}$, the residual points are sparse, which is benefit to choose candidate initial position points in the next iteration. After all corresponding landmarks are selected automatically, thin-plate spline function[7] is adopted to deform the source image, and elastic image registration is done.

We now present registration results for different mono-modal medical images using our method. The parameters are $\mathrm{Th}_{\mathrm{SM}}^0=0.5$, $\mathrm{Th}_{\min}=0.4$, $\mathrm{Th}_\rho=0.98$, and $r=0.9$. The local region is set as a circle whose radius is 16. We registered two-dimensional (2D) mono-modal brain images. The results are shown in Fig. 1. The reference images are PD, T2, and T1, alternatively. All images are 256×256 pixels in size. The source images are the reference images deformed manually and globally. The detected corresponding landmarks are labelled with asterisks. It can be seen that the feature points matched accurately. Edge differences between the registration result and the reference image show the outstanding performance of our method. The better the registration results are, the much more the superposition edges are. Obviously, the registration image using our method performs well because of the accuracy of landmark correspondence.

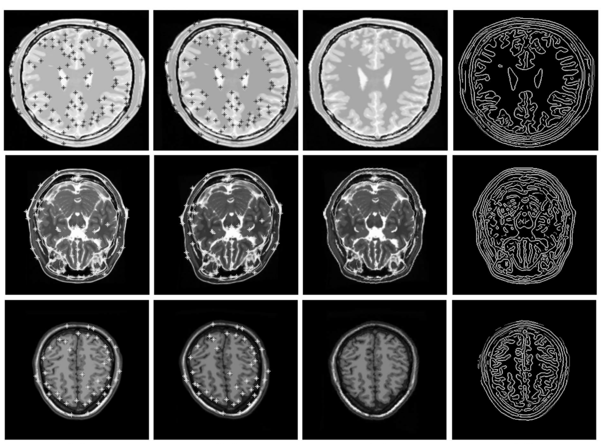

Fig. 1. Mono-modal medical image registration results. The first line is PD, the second line is T2 and the third line is T1. (a) Reference images; (b) deformed source images; (c) registration results of our method; (d) edge differences between the reference images and the registration results.

Our technique runs on a computer with 1.6-GHz CPU and 752-MB memory. The elapsed time of estimation corresponding landmarks is 80.737 for PD, 63.611 s for T2, and 29.242 s for T1. The average elapsed time ratios of mutual information to mean shift is 0.0735 for PD, 0.1006 for T2, and 0.1166 for T1. It can be seen that our method is rapid. This results show the advantage of mean shift, which converges in the direction of the gradient and guarantees hill climbing only for infinitesimal steps[5]. The average number of mean shift iterations is 18.434 for PD and 14.3492 for T2. It shows that our technique performs fast and efficient.

We adopt mean shift procedure to find the corresponding landmarks for mono-modal medical images. Mutual information between two local regions is combined with the similarity metric to deleting mis-matching point pairs. MLE technique is proposed to extract corresponding landmarks using different thresholds, which improves the stability of estimation. Experiments show that our techniques are robust and feasible to register mono-modal medical images. Moreover, the processing time is shortened significantly by adopting mean shift procedure. Our future works involve using mean shift to register multi-modal medical images.

This work was supported by the National Natural Science Foundation of China under Grant No. 60572101. X. Yang's e-mail address is pdwxyang@263.net.

References

1. B. Zitová and J. Flusser, Image and Vision Computing **21,** 977 (2003).
2. X. Yang and J. Pei, Chin. Opt. Lett. **3,** 510 (2005).
3. G. Liu, D. Liu, F. Liu, and Y. Zhou, Acta Opt. Sin. (in Chinese) **28,** 454 (2008).
4. H. Chui and A. Rangarajan, Computer Vision and Image Understanding **89,** 114 (2003).
5. D. Comaniciu, V. Ramesh, and P. Meer, in *Proceedings of IEEE Conf. Computer Vision and Pattern Recognition (CVPR'00)* **2,** 142 (2000).
6. J. Pei, Chin. J. Electron. **15,** 836 (2006).
7. F. L. Bookstein, IEEE Trans. Pattern Anal. Machine Intell. **11,** 567 (1989).