

Deformable Registration Using Scale Space Keypoints

Mehdi Moradi^a, Purang Abolmaesoumi^{a,b} and Parvin Mousavi^a

^aSchool of Computing, Queen's University, Kingston, Ontario, Canada K7L 3N6;

^b Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6

ABSTRACT

In this paper, we describe a new methodology for keypoint-based affine and deformable medical image registration. This fast and computationally efficient method is automatic and does not rely on segmentation of images. The keypoint pixels used in this technique are extreme points in the scale space and are characterized by descriptor vectors which summarize the intensity gradient profile of the surrounding pixels. For each of the keypoints in the scene image*, a corresponding keypoint is identified in the model image using the feature space nearest neighbor criteria. For deformable registration, B-splines are used to extrapolate a regular deformation grid for all of the pixels in the scene image based on the relative displacement vectors of the corresponding pairs. This approach results in a fast and accurate registration in the brain MRI images (an average target registration error of less than 2mm was acquired). We have also studied the affine registration problem in the liver ultrasound and brain MRI images and have acquired acceptable registrations using a mean square solution for affine parameters based on only around 30 corresponding keypoint pairs.

Keywords: Deformable registration, MRI, ultrasound, medical image processing, scale space keypoints, SIFT, B-splines.

1. INTRODUCTION

In most of the computer-aided intervention techniques, medical images are crucial for treatment planning and guidance during the surgery. However, images acquired in different stages of diagnosis and treatment are affected by the position of the patient, deformation of soft tissue and the orientation of the imaging device. Therefore, computer aided interventions often involve a registration step in which coordinate systems of two images (of the same or different modalities) are matched. As a result, corresponding pixels in two images represent the same anatomical regions of the tissue being imaged.

In recent years, different methodologies for medical image registration have been proposed in the literature (see surveys by Maintz² and Hill³). Three general categories of these techniques are: segmentation-based, landmark-based and intensity-based methods[†]. In segmentation based methods, the borders of an anatomical structure is delineated in the “model” and in the “scene”. The surface bounded by this border in the scene image is deformed to fit the corresponding surface in the model. In landmark-based methods, anatomical or geometrical structures are used to determine the appropriate deformation or transformation. Geometric landmarks can be sites in the image with specific geometric properties, such as the corner points, line intersection points, turning points in the

Further author information: (Send correspondence to P. Abolmaesoumi)

M. Moradi: E-mail: moradi@cs.queensu.ca

P. Abolmaesoumi: E-mail: purang@cs.queensu.ca, Telephone: (613)533 2767

P. Mousavi: E-mail: pmousavi@cs.queensu.ca

*The terminology used in image registration literature is not compatible in all sources. Here we follow the common one (also used by Besl and McKay¹) which uses the term “model” for the fixed image and the term “scene” for the moving image (which is transformed to match the model).

[†]Different classifications of medical image registration methodologies have been proposed in the literature. The categories discussed above are based on Maintz's² classification. In another classification proposed by Bankman,⁴ registration algorithms are considered to be either based on features extracted from the images (feature-based approaches) or on the optimization of a similarity measure between the images (similarity-based approaches).

curvatures on contours or even artificially placed markers. In intensity-based image registration techniques, the image grey values are directly utilized. Some of the intensity-based techniques rely on designing transformations that optimize similarity or dissimilarity measures (to provide just a few examples: cross-correlation,⁵ mutual information⁶ and variance of grey values⁷).

Intensity-based registration methods have the advantage of not involving manual or automatic segmentation or landmark selection processes. However, statistical distribution of the grey values is highly affected by noise in some of the image modalities, particularly in ultrasound images. In these instances, methods relying on global intensity features usually fail. One possible solution is to determine the desired transformation based on a set of keypoints with distinctive and stable local intensity-related features. The methodology proposed in the current paper falls into this category and is inspired by image template matching problem, an important research area in computer vision.⁸ Recently, a new development in that field was introduced by Lowe.⁹ He has proposed a novel method, called Scale Invariant Feature Transform (SIFT), for extracting distinctive features that characterize a set of keypoints for an image. The keypoints are the extreme points in the scale space of the image generated by applying “difference of Gaussian functions” to the image. These keypoints are shown to be scale and orientation invariant. Every keypoint is characterized with a 128 dimensional vector which forms a profile of the distribution of local image gradients around that point. This detailed profile allows for matching the keypoints with deformed instances of the image.

Some of the properties of this keypoint extraction and description methodology are promising for medical image registration. The local stability and deterministic nature of the keypoints result in extraction of stable keypoints from images and reduces the effects of noise. However, the method has some disadvantages as well. Medical images and images captured for computer vision applications are of different texture and contrast properties. This distinction results in a need for preprocessing procedures without which the scarcity of keypoints in specific parts of medical images will affect the registration. Particularly, this scarcity causes problems in registering ultrasound images with elastic deformations.

The present paper is based on a feasibility study to explore the applicability of a modified version of the SIFT method in designing a new technique for deformable medical image registration. We used a B-spline based method for applying the deformations (displacement vectors) computed based on matching SIFT keypoints to all the pixels of the scene image. In MRI images, sufficient and almost evenly distributed matching keypoints can be acquired and target registration errors are in an acceptable range.

We have also studied the affine registration using SIFT keypoints both in MRI and in ultrasound images. Our results demonstrate that in affine transformations, the matching process is very accurate and a limited number of best matching points (15-30) are sufficient for a precise least square solution to find the transformation parameters (both in MRI and in ultrasound images). The fast and efficient affine registration results in both MRI and ultrasound image modalities, and promising deformable registrations performed in MRI images show that the method is potentially applicable in medical image registration.

2. METHOD

2.1. Data and Preprocessing

We performed our experiments on two sets of images. The first set contains a sequence of 30 liver images acquired using a GE Voluson 730 Expert ultrasound machine. The second dataset is an MRI scan series consisting of 60 DICOM images provided by Robert Bemis and accessible at <http://www.mathworks.com/matlabcentral/>. The first attempts to run the SIFT algorithm on medical images returned no keypoint. Therefore, we applied a histogram equalization process to increase the contrast of the images (in both modalities).

Keypoint extraction from the ultrasound images proved more challenging than from the MRI images. Increasing the sensitivity of the SIFT method to image contrast reduced the number of keypoints to a level that made registration impossible. On the other hand, decreasing the sensitivity generated many keypoints that corresponded to speckle noise instead of any anatomical feature. This problem was partially addressed with the application of a sticks filter¹⁰ to reduce the level of speckle noise in the images[‡]. Designed for increasing the SNR

[‡]Ultrasound speckle noise is uncorrelated at large distance and Sticks filter takes advantage of this fact to reduce the level of noise and enhance the edges.

in ultrasound images, the sticks filter replaces the central pixel of each $N \times N$ neighborhood with the maximum summation value along lines with length N passing through the central pixel ($N = 15$ in our case).

2.2. Keypoint Extraction and Matching

Detection of Keypoints: In computer vision, the idea of analyzing the images at different spatial scales has long been studied. It has been shown¹¹ that Gaussian functions with different values of standard deviation (σ) form a unique family of kernels for decomposition of images into spatial scales. Therefore, the scale-space of an image, $L(x)$, can be defined as the indexed family of Gaussian-filtered images. The keypoints in SIFT method are the extreme points in the scale-space generated based on Equation (1):

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (1)$$

where G is the Gaussian function, I is the image and each value of k defines one scale. As Equation (1) shows, “difference of Gaussian function” is used in the SIFT method: two Gaussian filters are applied to the image and one image subtraction step is performed. Therefore, the implementation is computationally efficient.

A point is selected as a keypoint only if it has the largest or the smallest grey value in its 3×3 neighborhood in three successive scales. The keypoint set is then refined by rejecting points with low contrast (which are sensitive to noise) and also by eliminating the edge response effects (see⁹ for details). To achieve rotation invariant set of descriptor features, one or more orientations based on local image gradient directions are also assigned to the keypoints. The keypoint descriptors (later explained in this section) are extracted separately for each orientation and, therefore, are invariant to rotation.

Keypoint Descriptor Vector: The next step is to compute a descriptor vector for each keypoint. This vector contains a set of highly distinctive local features which are as invariant as possible to alterations in the intensity and 3D view point. A keypoint descriptor is created by computing the gradient magnitude and orientation at each image pixel in a region around the keypoint location (see Figure 1). These samples are then accumulated into orientation histograms each summarizing the gradient directions in one sub-region of the neighborhood (see Figure 2). Based on the size of the neighborhood and the sub-regions defined in each neighborhood, different number of descriptor parameters can be computed. We have used the 128-dimensional feature vector suggested by Lowe.⁹

Keypoint Matching Process: The match for each keypoint in the scene is found by identifying its nearest neighbor in the database of keypoints extracted from the model image using an exhaustive search[§] in the 128-dimensional feature space. The nearest neighbor is defined based on the Euclidean distance between the 128-dimensional descriptor feature vector representing each keypoint. For this particular application, due to the low SNR of the ultrasound images, we also applied an additional thresholding on the matched keypoint set. Pairs with Euclidean distance higher than the median of the distances in the matched dataset were discarded. This thresholding provides us with a more reliable matching set. In addition, we restricted our search to a maximum spatial distance in the image domain: For each keypoint in the scene a neighborhood of possible matches was defined (this value is experimentally set to a 10×10 window).

In case of affine transformation, there is an almost trivial least square solution to find the optimal transformation parameters. An affine transformation maps a pixel at $[x_i, y_i]$ to its matching point at $[z_i, w_i]$ according to:

$$\begin{bmatrix} z_i \\ w_i \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \times \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (2)$$

where a_1, a_2, a_3, a_4, t_x and t_y are the affine transformation parameters. Equation (2) can be collectively written for the entire matching keypoint dataset as:

[§]The algorithms suggested for nearest neighbor search (for example the k-d tree algorithm) provide negligible increase in speed compared to exhaustive search for larger than 10-dimensional spaces.¹²

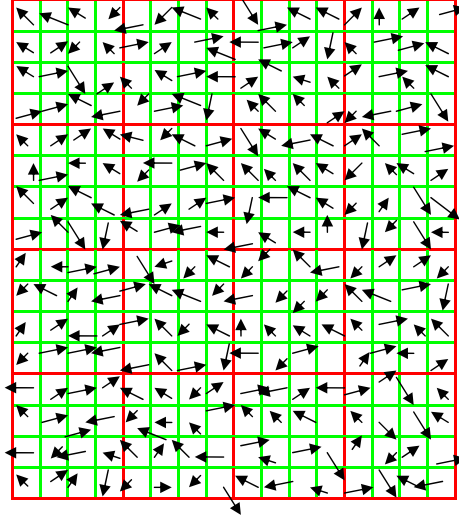


Figure 1. In SIFT algorithm (as used in this work) gradient vectors in a 16×16 neighborhood around a keypoint are used for feature generation.

$$\begin{bmatrix} \dots & & & & & \\ x_i & y_i & 0 & 0 & 1 & 0 \\ 0 & 0 & x_i & y_i & 0 & 1 \\ \dots & & & & & \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} \vdots \\ z_i \\ w_i \\ \vdots \end{bmatrix} \quad (3)$$

Equation (3) can be written with the notations of linear systems as:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (4)$$

The least-square solution for the parameters of this system is well known¹³:

$$\mathbf{x} = [\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{b} \quad (5)$$

From the theoretical point of view, three exact matches are sufficient to solve Equation (3). However, in practice more points are needed due to noise. We have found that 20 to 30 most reliable matching keypoints (with reliability defined as having the minimum Euclidean distance between corresponding feature vectors) are sufficient. Quantitative results are presented in Section 3.

2.3. Deformable Registration: B-Spline Based Deformation

We have followed the same keypoint extraction and matching process discussed earlier to solve the deformable registration problem. However, in this case we needed **a larger number of matching keypoints**. This is due to the fact that in deformable registration the number of unknowns is practically equal to the number of pixels in the image. To acquire the sufficient number of keypoints, we had to reduce the contrast threshold of Lowe's SIFT method. In case of ultrasound images, the resulting set of keypoints often contained several invalid keypoints (originated from speckle). However, sufficient number of valid keypoints (1000-3000) were acquired on MRI images.

Our deformable registration solution is based on finding the displacement vectors for all pixels in the image. For the matching set of keypoints, the displacement vectors can be acquired immediately by subtracting the

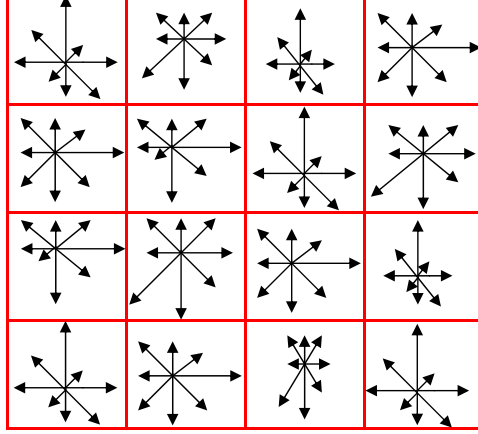


Figure 2. The pixels in the 16×16 neighborhood are divided to 4×4 windows. In each window, the gradient vectors are accumulated into 8 bins. This gives rise to 128 descriptors for each keypoint.

corresponding coordinates in the two images. However, the SIFT method does not result in evenly distributed set of keypoints. Therefore, an intermediate step of generating a regular grid from the scattered keypoint displacement vectors is needed. This problem can be stated as an interpolation process in which the function $\vec{d} = F(x, y)$ (where \vec{d} represents the displacement vector of the pixel located at (x, y)) with known values at irregularly distributed $x - y$ locations should be mapped to a regular grid. For this purpose, we have used Lee's method for scattered data interpolation with multilevel B-splines.¹⁴ We have considered a regular grid of $1/40$ of the original image size and have filled it in by using the uniform cubic B-spline basis functions. In this approach, displacement vector of each keypoint affects a neighborhood of 4×4 grid nodes. This regular grid is further interpolated (using a linear approach) to generate the original image size grid. The B-spline deformation tool in ITK[¶] software is used for applying the deformation field to the scene and for generating the registration result.

3. RESULTS

In order to create the scenes needed for implementing the affine registration experiments on both ultrasound and MRI images, we used MATLAB to apply a known affine transformation to the model. For the deformable registration, in ultrasound images, we used liver images with a typical time delay of 5 frames between the scene and model images (the patient was allowed to breath during imaging which caused elastic deformations in consequent images). For deformable registration of brain MRI images, the deformed scene was generated by applying synthetic distortions using ITK's B-spline deformation tool. The image dimensions were 440×400 in ultrasound dataset and 256×256 in MRI dataset. The scale factor of the images was 0.4mm (per pixel) for ultrasound and 0.9mm for MRI images.

Deformable Registration Results: In MRI images, particularly inside the borders of the brain tissue, the SIFT method successfully extracted almost evenly distributed keypoints. In our experiments, we extracted around 2000 keypoints from these images and used the top 700 matching keypoints for generating the B-spline regular grid. SIFT-based registration process effectively compensated the deformations of anatomical structures inside brain tissue. Target registration error (averaged over 10 images for each target) was 2 pixels or 1.8mm in MRI images (while the average error was around 20mm for the same targets before registration). Figure 3

[¶]Insight Segmentation and Registration Toolkit

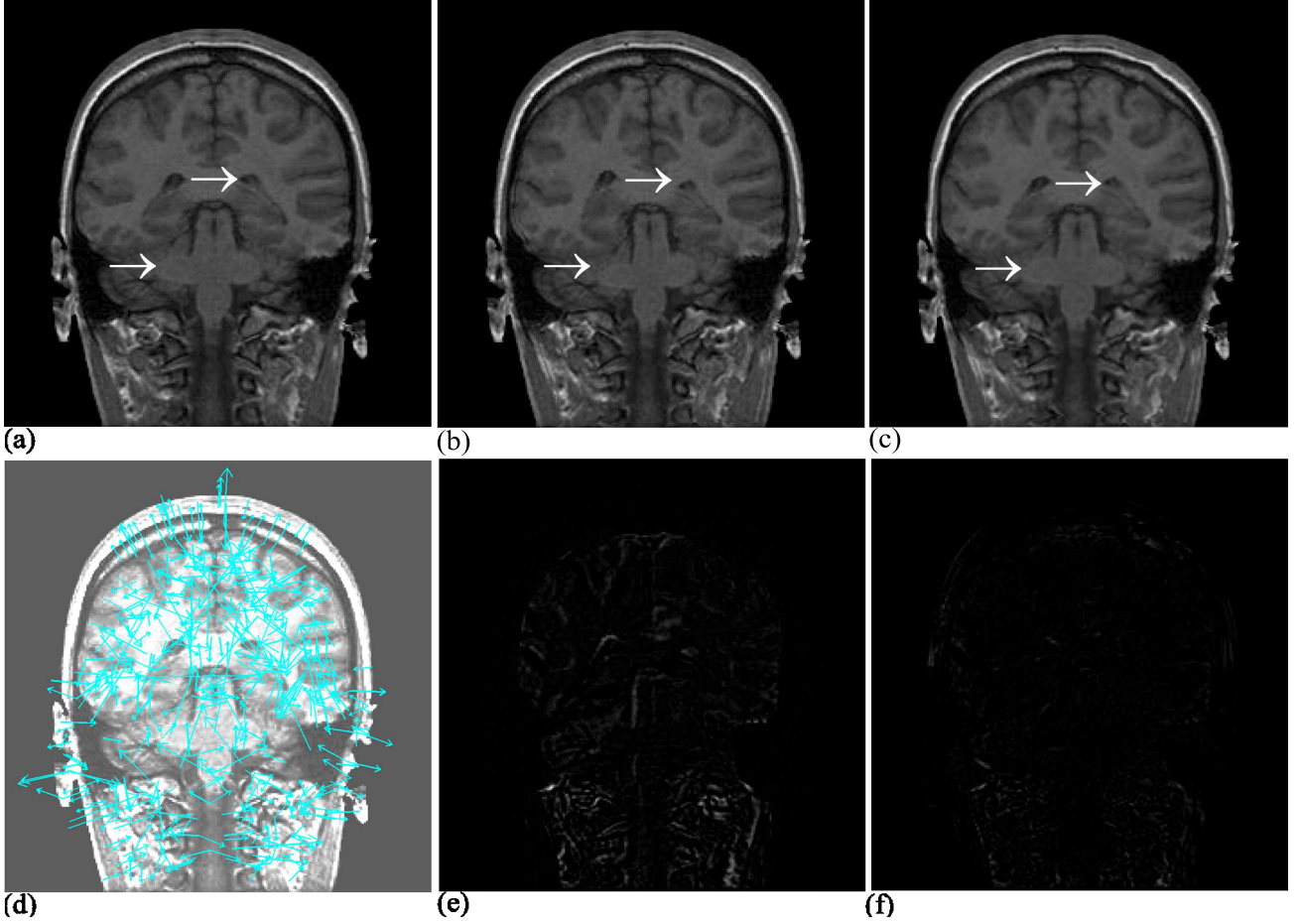


Figure 3. A typical deformable registration result in brain MRI images. The arrows in the three top images point to the same x-y coordinates. (a) Original model image; (b) Scene image; (c) The registration result - the position of anatomical structures in the brain tissue is considerably accurate; (d) Distribution of keypoints extracted from the original image (almost uniform); (e) Difference between the model and the scene; (f) Difference between the registered image and the model.

illustrates a typical outcome. The performance of the registration process is also evident through comparing the difference images (Figure 3-e and 3-f).

The task of deformable registration in ultrasound images proved challenging. Elastic deformations in some regions of the images were not compensated and in some cases the pixels had even moved in a wrong direction. The reasons might be: I) In ultrasound images, there often exist regions of very low contrast with high anatomical information content. The SIFT method does not extract keypoints from these regions; II) Ultrasound images acquired at separate times are dominated by uncorrelated speckle patterns. Although the sticks filter reduces the speckle effect, it also dramatically reduces the number of available keypoints in the image. The deformable registration results were only partially acceptable in many ultrasound images due to these two limitations.

Affine Registration Results: For each affine registration experiment, we extracted two sets of keypoints (from the preprocessed models and scenes) and their corresponding descriptors. These keypoints were used in Equation (5) for computing the affine parameters that optimally registered the scene onto the model image. We report the average results acquired using two sets of affine parameters each applied to ten pairs of images (separate experiments were performed on MRI and ultrasound images). With a moderate affine transformation, ($[a_1 = 1.2, a_2 = -0.1, a_3 = 0.2, a_4 = 1.4, t_x = 5, t_y = 6]$), the registration in both MRI and ultrasound datasets resulted in comparable and very small errors (Figures 4 and 5). The average rotational error was less than 0.5° and the average difference of the distance between two selected landmarks in the registered image and in the model (target registration error) was less than one pixel. With a larger affine transformation

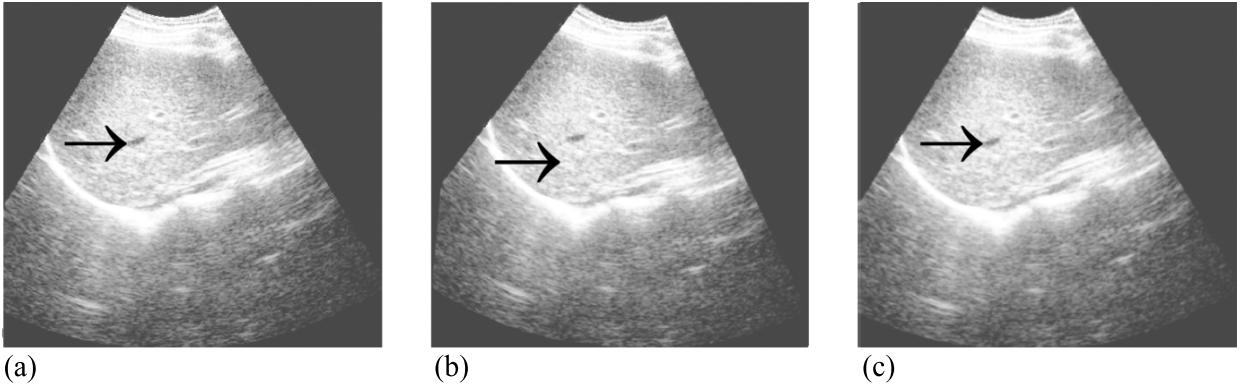


Figure 4. A typical affine registration result in ultrasound images. 30 top matching keypoints were used for solving Equation (4). Arrows in all images point to the same x-y coordinates (before cropping). (a) Model image, (b) scene with affine parameters: $[a_1 = 1.2, a_2 = -0.1, a_3 = 0.2, a_4 = 1.4, t_x = 5, t_y = 6]$, (c) Registered image.

($[a_1 = 1.4, a_2 = -0.3, a_3 = 0.7, a_4 = 2, t_x = 15, t_y = 10]$) the rotational error increased to approximately 1.4° in MRI images and to 4.3° in the ultrasound images.

4. CONCLUSIONS

In this paper, we presented a new medical image registration technique based on the SIFT keypoint extraction and description method which uses the notion of spatial scale space. Our main achievement was in applying this methodology along with a B-spline based interpolation technique for registering elastically deformed MRI images. The technique showed a promising performance with acceptable error in restoring the position of important anatomical structures in brain tissue (with target registration errors of around $2mm$). However, the task of deformable registration in ultrasound images proved challenging and the set of valid SIFT keypoints were not sufficient.

In case of affine registrations, we found that keypoints extracted from the MRI images are more stable compared to the ultrasound image keypoints. Also, as the values of affine transformation parameters increase, the registration error increases more quickly in ultrasound images than in MRI images. Nevertheless, the suggested registration method is sufficiently accurate in both image modalities when the deformations are affine.

In general, we found the SIFT method a potentially powerful tool for medical image registration (particularly on MRI images). One of the possible future developments of the current work can be in dealing with registering partially overlapping images. In many medical applications, the scene does not completely match the model and the anatomical structure of interest covers only a portion of the model. SIFT is a candidate for solving the problem of registering partially overlapping views in which many intensity based methods fail. Lowe has applied SIFT for identifying partially obscured objects in real world images and the same idea might be applicable to medical applications. The other possible development of the current work can be extending the method to 3D images.¹⁵

ACKNOWLEDGMENTS

The authors would like to thank Mr. R. H. Gong for providing the B-spline deformation code and Dr. D. G. Lowe for providing the SIFT code. This project was supported by the Natural Sciences and Engineering Research Council of Canada and IRIS/Precarn Incorporated.

REFERENCES

1. P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), pp. 239–255, 1992.

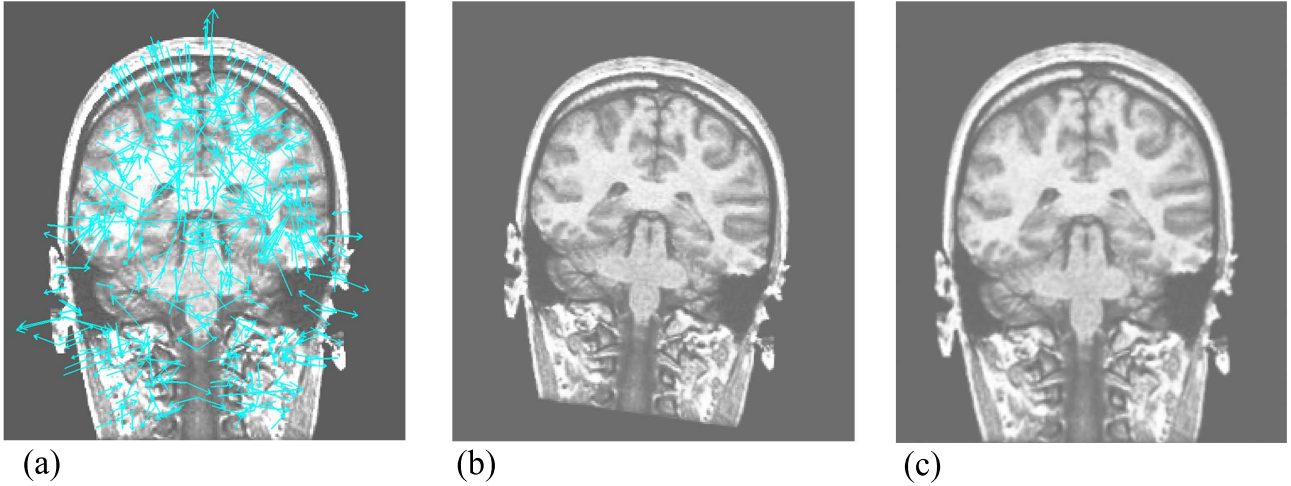


Figure 5. A typical affine registration result in MRI images. 135 selected matching keypoints are used for solving Equation (4). (a) Model with distribution of 900 keypoints illustrated on the image, (b) Scene with affine parameter set of $[a_1 = 1.2, a_2 = -0.1, a_3 = 0.2, a_4 = 1.4, t_x = 5, t_y = 6]$, (c) Registered image.

2. J. B. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis* **2**(1), pp. 1–36, 1998.
3. D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in Medicine and Biology* **46**(3), pp. R1–R45, 2001.
4. I. N. Bankman, *Handbook of Medical Imaging*, Academic Press, CA, 2000.
5. D. L. G. Hill, *Combination of 3D medical images from multiple modalities*. PhD thesis, University of London, 1993.
6. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging* **16**(2), pp. 187–198, 1997.
7. G. S. Cox and G. de Jager, "Automatic registration of temporal image pairs for digital subtraction angiography," in *SPIE Medical Imaging*, **2167**, pp. 188–99, SPIE, 1994.
8. C. Schmid and R. Mohr, "Local gray value invariants for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(5), pp. 530–434, 1997.
9. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**(2), pp. 91–110, 2004.
10. R. N. Czerwinski, D. L. Jones, and W. D. O'Brien, "Detection of lines and boundaries in speckle images Application to medical ultrasound," *IEEE Transactions on Medical Imaging* **18**, pp. 126–136, 1999.
11. J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(1), pp. 26–33, 1986.
12. J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Transactions on Mathematical Software* **3**(3), pp. 209–226, 1977.
13. T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice-Hall, 2000.
14. S. Lee, G. Wolberg, and S. Shin, "Scattered data interpolation with multilevel B-splines," *IEEE Transactions on Visualization and Computer Graphics* **3**(3), pp. 228–244, 1997.
15. C. A. Rockwood, F. Matsen, and M. Wirth, eds., *The Shoulder*, ch. 8, pp. 290–335. WB Saunders Co., Philadelphia, 2nd ed., 1998.