Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Multi-sensor image registration using multi-resolution shape analysis*

YUAN Zhen-ming[1,2], WU Fei[‡1], ZHUANG Yue-ting[1]

([1]*School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*)

([2]*School of Information Engineering, Hangzhou Teacher's College, Hangzhou 310036, China*)

E-mail: zmyuan@cs.zju.edu.cn; wufei@cs.zju.edu.cn; yzhuang@cs.zju.edu.cn

Received Feb. 10, 2005; revision accepted June 3, 2005

Abstract: Multi-sensor image registration has been widely used in remote sensing and medical image field, but registration performance is degenerated when heterogeneous images are involved. An image registration method based on multi-resolution shape analysis is proposed in this paper, to deal with the problem that the shape of similar objects is always invariant. The contours of shapes are first detected as visual features using an extended contour search algorithm in order to reduce effects of noise, and the multi-resolution shape descriptor is constructed through Fourier curvature representation of the contour's chain code. Then a minimum distance function is used to judge the similarity between two contours. To avoid the effect of different resolution and intensity distribution, suitable resolution of each image is selected by maximizing the consistency of its pyramid shapes. Finally, the transformation parameters are estimated based on the matched control-point pairs which are the centers of gravity of the closed contours. Multi-sensor Landsat TM imagery and infrared imagery have been used as experimental data for comparison with the classical contour-based registration. Our results have been shown to be superior to the classical ones.

Key words: Image registration, Shape descriptor, Feature matching, Multi-resolution representation
doi:10.1631/jzus.2006.A0549 **Document code:** A **CLC number:** TP391

INTRODUCTION

Image registration plays an important role in the fusion of remote sensing, video panorama, medical image processing, etc. (Li *et al*., 1995; 2004; Brown, 1992; Belongie *et al*., 2001; Rui *et al*., 2001). More different than video panorama, the registration of multi-sensor image needs to align the multi-sensor images of the same scene from different sources (Rignot *et al*., 1991). Multi-sensor image registration is widely used in the fusion of panchromatic images with infrared images, or the fusion of CT with MRI, but registration performance is degenerated when heterogeneous images are involved. The possible reasons are: (1) Different cameras with different properties cause different view angle and focus of each image; (2) The resolutions of images from different cameras are also different; (3) The intensity of multi-sensor images may be different, sometimes even opposite, or the objects of the same scene in different images are not invariant and therefore the correspondence of images cannot be found.

Current image registration methods can be categorized into three groups: (1) The method global registration through image space correlation method (Brown, 1992) which cannot be used for registration for images with different resolutions and characteristics; (2) The transform based registration method (Brown, 1992; Reddy and Chatterji, 1996; Viola and Wells, 1997) which transforms original images into logarithmic polar coordinates systems, and finds their correlation in the frequency domain. This method is fit for homogeneous images, but it has difficulty in

‡Corresponding author
*Project supported by the National Natural Science Foundation of China (No. 60272031), the Hi-Tech Research and Development Program (863) of China (No. 2003AA131032-2), and the Natural Science Foundation of Zhejiang Province (No. M603202), China

heterogeneous ones; (3) The feature based registration method which first detects the visual features like points and contours, and implements transformation of images according to correspondence of features (Li *et al.*, 1995; Dai and Khorram, 1999; Fonseca and Manjunath, 1996; Irani and Anandan, 1998; Li and Zhou, 1996). The feature based image registration method can not only be used for homogeneous image registration but also for heterogeneous images registration. However, selection of well defined invariant feature is extremely important in this method. Since the shapes of main objects in images are basically invariant, and not affected by the heterogeneous images, it is better to use shapes for heterogeneous images registration.

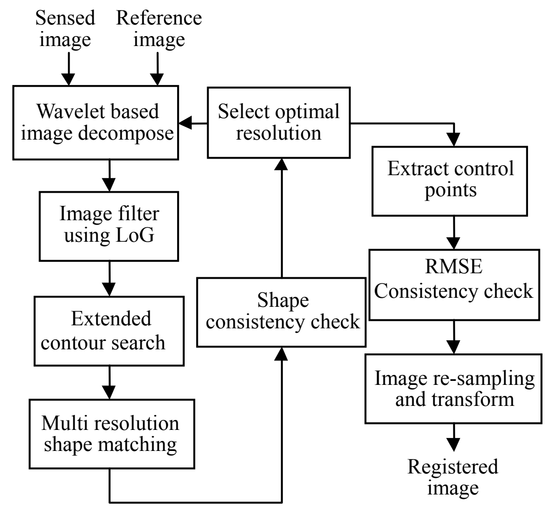


Fig.1 Framework of multi-resolution image registration

MULTI-RESOLUTION IMAGE REGISTRATION

The multi-resolution shape matching method described below implements the multi-sensor image registration, which can be classified as the feature based image registration. Multi-resolution shape detection, description and matching are used to implement the coarse-to-fine registration.

An image pyramid constructed by wavelet pyramid decomposition is constructed to extract the contours under different image resolutions, which can yield both coarse and detailed contours. The contours are detected by extended contour search algorithm in order to acquire more closed contours. In order to avoid differences of resolution between multi-sensor images and the problem of more details in high resolution images, suitable resolutions for each image are decided by maximizing the consistency between the shapes extracted from multi-resolution images. During the feature matching, multi-resolution shape descriptors are constructed by Fourier transformation, and are independent of translation, rotation and scaling of shapes. A minimum distance function is used to compare similar shapes. The centroids of the contour pairs are regarded as the control points to estimate the parameters of the transform model aligning the target image with the reference image (Fig.1).

CONTOURS SHAPE DETECTION

In traditional feature-based registration, such as that in (Li *et al.*, 1995), many important contours are missed because of the noise and the low resolution of images, and the result of registration is affected greatly by the detected contours. Here we decompose each image into low frequency image with coarse contours and high frequency image with detailed contours using wavelet transform together, and select optimal resolution iteratively to reduce the influence of image resolution. The key point of contour detection is to filter the image by LoG operator, which preserves the prominent characters under different resolution. And an extended contour search algorithm is used to obtain relatively integrated contours.

LoG based image filtering

Laplacian of Gaussian (LoG) is an edge detection operator and can smooth the image by using Gaussian filter. Generally the image noise is regarded as Gaussian noise. The 2D Gaussian filter is given as Eq.(1):

$$G(x,y)=\frac{1}{2\pi\sigma^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{1}$$

By calculating the second differential of Eq.(1), the Gaussian smooth filter and the Laplacian operator can be combined into an LoG operator:

$$\nabla^2 G(x,y)=\frac{\partial^2 G(x,y)}{\partial^2 x}+\frac{\partial^2 G(x,y)}{\partial^2 y}$$

$$=\frac{1}{2\pi\sigma^4}\left(\frac{x^2+y^2}{\sigma^2}-2\right)\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}. \quad (2)$$

After convolving the image by LoG, the zero-crossing points can be confirmed by the change of signs. Practically the LoG operator can be simplified into two 1D convolutions along the X and Y directions to accomplish rapid computation of 2D convolutions (Li *et al.*, 1995):

$$\nabla^2 G * \boldsymbol{I}(x,y) = \sum_{j=-W}^{W}[C(x-j,y)K_1(j)+D(x-j,y)K_2(j)] \quad (3)$$

where $\boldsymbol{I}(x,y)$ is the image matrix, K_1 and K_2 are the coefficients of LoG simplified along X and Y directions, the 1D convolutions as follows:

$$C(x-j,y)=\sum_{j=-W}^{W}\boldsymbol{I}(x-j,y-i)K_2(i), \quad (4)$$

$$D(x-j,y)=\sum_{j=-W}^{W}\boldsymbol{I}(x-j,y-i)K_1(i). \quad (5)$$

After filtering the image by LoG, the image intensity is obtained to support the extended contour searching.

Extended contour searching

The contours can be detected by searching the zero-crossing points or changed sign points in the filtered image. However, such edge points are often discontinuous when noises occur. In order to obtain the continuous contour, we usually set up two predefined thresholds T_1 and T_2 (Li *et al.*, 1995; Worring, 1993), and the convolved value of pixels larger than T_1 and smaller than T_2 are selected as the candidate edge points. But this method cannot preserve the whole distribution of shapes in the filtered image practically. Extended contour searching using one threshold is adopted here.

Considering the character of single point width at the zero-crossing point in filtered image by LoG, and the remarkable visual characters, the statistically majority zero-crossing points are selected as the contour points. Assign the first minimum value in the gray histogram function $H(i)$ as the threshold:

$$\frac{\mathrm{d}H(i)}{\mathrm{d}i}=0 \quad (6)$$

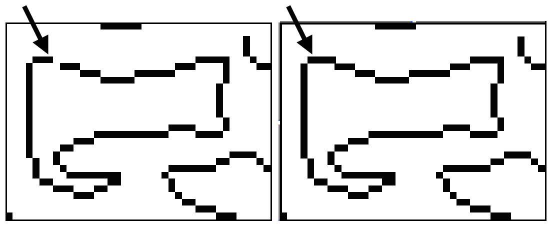
where $H(i)$ is the ith bin of histogram ($0\le i\le 255$). Then the threshold is used to segment the image into the edge image, and the contour searching is processed on the edge image.

Normal contour searching algorithm gets the 8-neighborhood chain code by searching along the 8 neighbourhoods sequentially (Haralick and Shapiro, 1993). 8-neighborhood chain code starts from the current pixel, defines the pixel in the east as 0, and constructs codes of contour along the counter-clockwise direction. The searching process selects the upper left pixel as the start point, and checks the 3×3 neighbourhood of the current pixel along the counter-clockwise direction, then takes the first nonzero point as the next contour point until there is no nonzero point in the neighbourhood. However, because of the noise, such searching is difficult to yield many significant closed contours with more contribution to better registration. Such contours are always continuous visually, but discontinuous in the edge image. So the contour searching process is extended by tracking the 4×4 neighbourhood when there is no contour point in the 3×3 neighbourhood. If the extended searching meets the starting point, then it connects with the starting point to form the closed contour. Or if another non-zero point is met, then it is connected by point interpolation and the searching goes on continuously. Such extension can yield additional visually closed contours and longer open contours, which will improve the efficiency of the matching result (Fig.2).

The rules of extended contour searching process are defined as Fig.3.

6	5	4	3	2
7	3	2	1	1
8	4		0	0
9	5	6	7	15
10	11	12	13	14

Fig.2 The diagram of 4×4 extended contour searching



(a) (b)

Fig.3 Contours searching results

(a) Contours from two threshold algorithm, the middle contour is an opened contour disconnected at the left corner; (b) Contour from Extended Search Algorithm, the middle contour is a closed contour

In 3×3 neighborhood, the next searching direction $d1$ is:

$$d1 = (d1+i) \bmod 8, \tag{7}$$

and in 4×4 neighborhood, the next searching direction $d2$ is:

$$d2 = (d2+i) \bmod 16, \tag{8}$$

and the direction from $d1$ to $d2$ is:

$$(d2 - d1) \bmod 8. \tag{9}$$

MULTI-RESOLUTION SHAPE MATCHING

Shape is the important representation of image contents (Dudek and Tsotsos, 1997). The contours extracted from the above searching are used to present both shape and visual distribution of the object in the image. In image registration, the features will be represented by shape descriptors, which must be independent of scale, rotation, and translation. Considering the periodicity of the closed contour and the Fourier transformation, the closed contours can be represented by the coefficients of Fourier transformation on the contours. Fourier transform is a multi-resolution method in which the low-frequency of the transformation domain expresses the coarse shape of the contour, and the high-frequency expresses the details of the contours. Such transform also accords with the visual characters in shape matching: human being always compares the main shapes between the contours, and then aligns details between two similar contours.

Shape descriptor extracting

Suppose C is the closed contour in complex plane, let the starting point b be on the contour and move anticlockwise along the contour at uniform velocity, then we obtain a function $z(t)$, where t is time. Let the period of $z(t)$ be 2π, then $z(t)$ can be represented by the following Fourier coefficient:

$$z(t) = \int C_n \mathrm{e}^{\mathrm{i}nt}, \tag{10}$$

where C_n is the Fourier descriptors of C. Let L be the contour length, and l be the arc length along with the contours, then we have

$$t = \frac{2\pi l}{L}. \tag{11}$$

The Fourier coefficients can be obtained by:

$$C_n = \frac{1}{N}\int_0^N z(t)\mathrm{e}^{-\mathrm{i}nt}\mathrm{d}t = \frac{2\pi}{L}\int_0^L z(l)\mathrm{e}^{-\mathrm{i}(2\pi/t)nl}\mathrm{d}l \tag{12}$$

where $z(t)$ is the curvature function. Suppose contour C is composed of N points, (x_l, y_l), $0 \le l \le N-1$, then the curvature of the contour is defined as the differential of $\theta(l)$:

$$z(l) = \frac{\mathrm{d}}{\mathrm{d}l}\theta(l), \tag{13}$$

where $\theta(l)$ is the cutting angle function:

$$\theta(l) = \arctan(\mathrm{d}x/\mathrm{d}y) = \arctan\left(\frac{\mathrm{d}x/\mathrm{d}l}{\mathrm{d}y/\mathrm{d}l}\right) = \arctan\left(\frac{x_l'}{y_l'}\right). \tag{14}$$

The intrinsic characteristic of the curvature function guarantees the invariant of shape and the invariability of rotation by using the module of the Fourier coefficients without the phase. In order to keep the scale invariant, each module of Fourier coefficients is divided by DC components. Then each shape descriptor of the contour can be represented as follows, only using a few low-order coefficients F_i, (in our experiment, i=5):

$$\boldsymbol{F}_k = (|F_1|, |F_2|, |F_3|, |F_4|, |F_5|). \tag{15}$$

Minimum distance classification based shape matching

Regarding the shape descriptor of each closed contour as the eigenvector, the similarity between two contours can be transformed to the correlation problem between two eigenvectors. Suppose m closed contours are extracted from the reference image, and n closed contours are extracted from the sensed image, then the distance of two eigenvectors can be defined as:

$$d_{ij}=\sum_k |\boldsymbol{F}_{ik}-\boldsymbol{F}_{jk}|^2, \quad (1\le i\le m, 1\le j\le m). \tag{16}$$

The distance matrix of the features, $|d_{ij}|$, can be obtained by calculating the distances between all shape pairs of two images. According to the principle of minimum distance classification, the shape pair (C_i,C_j) satisfying the following conditions may be considered similar:

$$\min_i |d_{ij}| = \min_j |d_{ij}|, \tag{17}$$

$$m=\min_i |d_{ij}| < t_s, \tag{18}$$

where t_s is a threshold. After getting the pairs of contours, the centroids of contours are used as the control points for the image registration.

TRANSFORMATION MODEL ESTIMATION

To register two images, the coordinate transformation between two images must be found. In a majority of multi-sensor spatial images, the distance of camera to objects approximates infinitude relative to the size of objects. So the coordinate transformation between two images can be regarded as the rigid planar models, which is composed of scaling, rotation and translation. The pixel (x_1,y_1) in image I_1 is mapped to the pixel (x_2,y_2) in image I_2 by Eq.(19) (Brown, 1992; Rignot *et al.*, 1991):

$$\begin{aligned}\begin{bmatrix}x_2\\y_2\end{bmatrix}&=s\boldsymbol{R}\begin{bmatrix}x_1\\y_1\end{bmatrix}+\boldsymbol{D}\\&=s\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}x_1\\y_1\end{bmatrix}+\begin{bmatrix}d_x\\d_y\end{bmatrix},\end{aligned} \tag{19}$$

where $(x_1,y_1)^{\mathrm{T}}$, $(x_2,y_2)^{\mathrm{T}}$ is the position of the corresponding points in two images respectively. s, θ, and $(d_x, d_y)^{\mathrm{T}}$ are respectively the parameter of scale, rotation, and translation. Then the four parameters are estimated by the set of control points $\{(x, y), (x',y')\}$.

Let $u=s\cos\theta$, $v=s\sin\theta$, then we have

$$\begin{pmatrix}x'\\y'\end{pmatrix}=\begin{pmatrix}u & -v\\ v & u\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}d_x\\d_y\end{pmatrix}. \tag{20}$$

The unknown quantities u, v, t_x, t_y are estimated by Least-Squares fitting. The registered image is obtained after the consistency checking of control point pairs and the re-sampling of original image. In consistency check, the root mean square error (RMSE) between the matched points provides a measure of registration accuracy and is defined as (Li *et al.*, 1995):

$$\begin{aligned}RMSE=\Big(\sum[(ux_i-vy_i+\Delta x-x_i')^2\\+(uy_i-vx_i+\Delta y-y_i')^2]/m\Big)^{1/2}.\end{aligned} \tag{21}$$

EXPERIMENTAL RESULTS AND CONCLUSIONS

Many pairs of multi-sensor images, such as the multi-sensor Landsat TM images and visual/IR images, were selected as the experimental data for validating our registration approach. The results of traditional feature based registration methods (Li *et al.*, 1995) were compared with the one using our registration based on multi-resolution shape analysis. Fig.4 gives our result. The reference image, with its strength map and the extended contour is listed in the first row from the first column to the third column. The being registered image with its strength map and the extended contour is also listed in the second row. The strength maps and the extended contours reflect obviously the global distribution of the original images, and the outlines of the objects in the original image are well preserved. In the extended contour of the reference image, 27 closed contours were detected, 16 more than that detected by the traditional feature based registration (Li *et al.*, 1995). The fusion image and the transformed image of the to-register image are listed in the last column.

Table 1 shows the error comparison of two registration methods on three pairs of multi-sensor images. It shows more control points are found in our multi-resolution shape matching algorithm than the algorithm using invariant moments in (Li *et al.*, 1995). Consequently our result is better than that obtained by traditional methods.

ACKNOWLEDGEMENT

Thanks to Li H., Manjunath B.S., *et al.*, who provide the experimental data on their website.

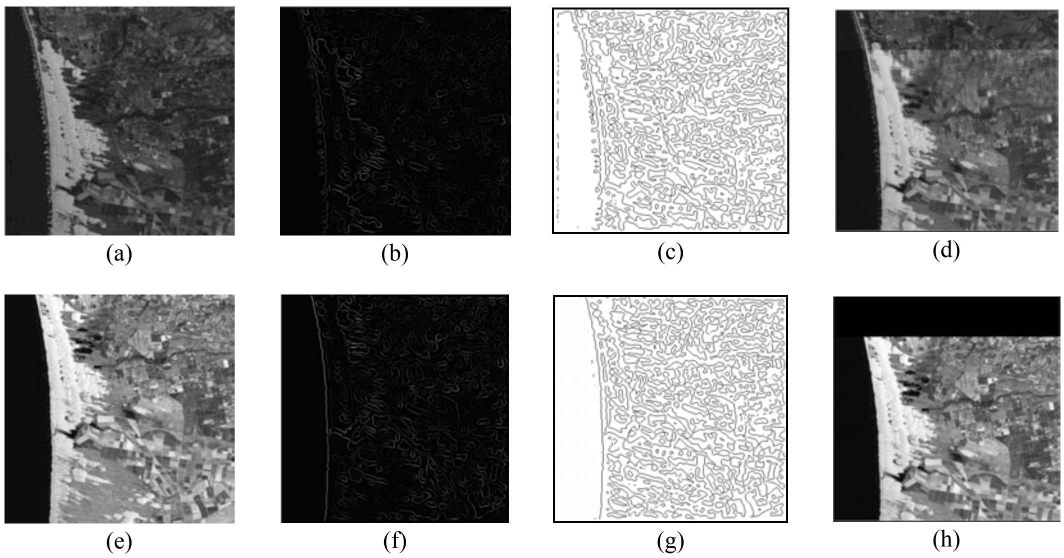


Fig.4 Registration of Dunes883/885 (TM3/TM5)
(a) Reference image; (b) Strength map; (c) Extended contours; (d) Fusion image; (e) To-register image; (f) Strength map; (g) Extended contours; (h) Transformed image

Table 1 Comparisons of different registrations

Image pairs	True transform parameters			Parameters estimated by contour-based method				Parameters estimated by this paper			
	Scale	$\Delta\theta$	$(\Delta x, \Delta y)$	CPs	Scale error	Rotate error	Trans. error	CPs	Scale error	Rotate error	Trans. error
B040/b042	1.005	–29.9413	(–186.606, 275.939)	4 (of 37)	0.0051	0.031	(3,1)	13 (of 56)	0.0049	0.005	(1,2)
Dunes TM3/TM5	1.000	–0.001	(0,78)	3 (of 12)	0.0005	0.0597	(0,1)	8 (of 16)	0.0003	0.0512	(0,0)
Img1/Img2 (TM0/TM8)	1.000	–14.4423	(264.937, –54.7578)	5 (of 32)	0.0036	0.4151	(3,2)	12 (of 46)	0.0012	0.3728	(2,2)

References

Belongie, S., Malik, J., Puzicha, J., 2001. Matching Shapes. International Conference on Computer Vision (ICCV'01), Canada, **1**:454-462.

Brown, L.G., 1992. A survey of image registration techniques. *ACM Computing Surveys*, **24**(4):325-376. [doi:10.1145/146370.146374]

Dai, X., Khorram, S., 1999. A feature-based image registration algorithm using improved chain-code representation combined with invariant moments. *IEEE Trans. on Geoscience and Remote Sensing*, **37**(5):2351-2362. [doi:10.1109/36.789634]

Dudek, G., Tsotsos, J., 1997. Shape representation and recognition from multiscale curvature. *Computer Vision and Image Understanding*, **68**(2):170-189. [doi:10.1006/cviu.1997.0533]

Fonseca, L.M.G., Manjunath, B.S., 1996. Registration techniques for multisensor remotely sensed imagery. *Photogrammetry Engineering and Remote Sensing Journal*, **562**(9):1049-1056.

Haralick, R.M., Shapiro, L.G., 1993. Computer and Robot Vision. Addison-Wesley, MA, **1**:57-68.

Irani, M., Anandan, P., 1998. Robust Multi-Sensor Image Alignment. Sixth International Conference on Computer Vision, Bombay, India, p.959-966.

Li, H.H., Zhou, Y.T., 1996. Automatic visual/IR image registration. *Optical Engineering*, **35**(2):391-400. [doi:10.1117/1.600908]

Li, H., Manjunath, B.S., Mitra, S.K., 1995. A contour based approach to multisensor image registration. *IEEE Trans. on Image Processing*, **4**(3):320-334. [doi:10.1109/83.366480]

Li, Q., Zheng, N.N., Ma, L., Cheng, H., 2004. Principal Component Analysis Neural Network Based Probabilistic Tracking of Unpaved Road. International Symposium on Neural Networks, Dalian, China, p.792-797.

Rui, Y., He, L.W., Gupta, A., Liu, Q., 2001. Building an Intelligent Camera Management System. Proceedings of the Ninth ACM International Conference on Multimedia, Ottawa, Canada, p.2-11.

Rignot, E.J.M., Kowk, R., Curlander, J.C., Pang, S., 1991. Automated multisensor registration: Requirements and techniques. *Photogrammetry Engineering and Remote Sensing Journal*, **57**(8):1029-1038.

Reddy, B.S., Chatterji, B.N., 1996. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. on Image Processing*, **5**(8): 1266-1271. [doi:10.1109/83.506761]

Viola, P., Wells, W.M., 1997. Alignment by maximization of mutual information. *International Journal of Computer Vision*, **24**(2):137-154. [doi:10.1023/A:1007958904918]

Worring, M., 1993. Shape Analysis of Digital Curves. Ph.D Thesis. University of Amsterdam.