第33卷第1期
2014年2月

红外与毫米波学报
J. Infrared Millim. Waves

Vol. 33, No. 1
February, 2014

文章编号：1001－9014(2014)01－0090－08

DOI：10.3724/SP.J.1010.2014.00090

Visible and infrared automatic image registration based on SLER

DONG Xiao-Jie[1], LIU Er-Qi[2], YANG Jie[1], WU Qiang[3]

(1. Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China;
2. China Aerospace Science & Industry Corp, Beijing 100048, China;
3. School of Computing & Communications, University of Technology Sydney, Sydney 2007, Australia)

Abstract: A novel approach to the problem of visible and infrared automatic image registration was proposed. The registration is performed by extracting affine covariant regions through same level extremal region (SLER) detector on a gray gradient image. Then, hypergraph matching algorithm was employed to obtain identical key points. The approach is especially suitable for registering multi-sensor infrared images where the quality of images or the corresponding edge maps are worse than the counterparts on a common optical image. Experiments performed on several challenging real image pair show that our proposed method achieves better performance than other approaches.

Key words: multimodality image registration; Infrared image; maximally stable extremal region

PACS: 89.20.Bb

基于SLER检测的红外与可见光图像自动配准

董效杰[1]，刘尔琦[2]，杨 杰[1]，吴 强[3]

(1. 上海交通大学 图像处理与模式识别研究所，上海 200240；
2. 中国航天科工集团，北京 100048；
3. 计算通信学院，悉尼科技大学，澳大利亚 悉尼 2007)

摘要：针对可见光图像和红外图像配准问题，提出了一种新的自动配准方法. 该算法通过同级极值区域检测子在灰度梯度图像上提取仿射协变区域. 然后利用超图匹配算法确定匹配点对实现图像配准. 该方法尤其适合于红外图像的质量或者边缘比对应的可见图像质量或边缘差情况下的异模配准. 对一些具有挑战性的图像对进行试验，实验结果表明我们提出的方法比其他方法获得了更好的性能.

关 键 词：多模态图像配准；红外图像；最大稳定极值区域

中图分类号：TP391.41 **文献标识码**：A

Introduction

In recent years, multimodality imaging system has been used in a wide variety of fields, such as military, medic, urban monitoring. The integration of images from multi-sensor can provide complementary information, and therefore, increase the accuracy of the overall decision-making. A fundamental problem in multimodality image integration is that of aligning images taken under different conditions, like positions, viewpoints, times and illumination. This is known as image registration.

Finding reliable correspondences from multimodality image pair is a difficult and critical step. Once such correspondences have been found, a transformation matrix can be readily obtained. Based on the transformation matrix, these registered images can be transformed into the same reference. However, due to the

Received date: 2012-10-17, **revised date**: 2013-12-15 **收稿日期**：2012-10-17，**修回日期**：2013-12-15

Foundation items: Supported by National Natual Science Foundation of China (61273258).

Biography: DONG Xiao-Jie (1977-), male, Puyang, Henan Province, China, Ph. D. candidate. Research fields include image processing and image registration. E-mail: dxjzpc@163.com.

different properties between different sensors, the corresponding relationship between the intensities of matched pixels is usually unpredictable. The features presented in one image might only partially appear in the other image or do not appear at all. Multiple intensity values in one image may map to a single intensity value in the other image, and vice versa. Contrast reversal may occur between these images in some image regions but not in others. Therefore, multimodal registration is more complicated than monomodal registration. It is a challenging problem.

This paper focuses on the registration between infrared image and optical image. Due to the different properties of sensors, the intensities of infrared image are mainly influenced by object temperature and heat radiation in the scene. However, the intensities of optical image are mainly determined by the reflected light on the objects in the scene. Therefore, normally there is no direct relation in values of the pixel intensity between infrared image and optical image, which makes it difficult to match the identical feature points between these two kinds of images. A pair of infrared image and optical image is illustrated in Fig. 1 (a). Obviously, most clear texture features in optical image disappear in infrared image. Even the lines with strong contrast in optical image are weakened in infrared image. Existing feature extraction algorithms based on intensity fail. Interest features extracted by such detector (Canny, Harris) are demonstrated in Fig. 1. It can be seen that the challenging problem is to extract sufficient number of identical feature points from both images. Another problem encountered is feature description that ensures accurate feature matching across different images.

To address these two problems mentioned above, a novel detector SLER is proposed for extracting affine covariant regions of the structural image. The first moments of detected regions are used to build the high-order tensor. The registration algorithm based on SLER is presented as follows. The interested covariant regions are extracted by SLER detector in the gradient map at first. Then the enhanced high-order tensor is built by the integration of high-order tensor and low-level features through inner product. Thereafter, the soft distributable matrix is established by hypergraph

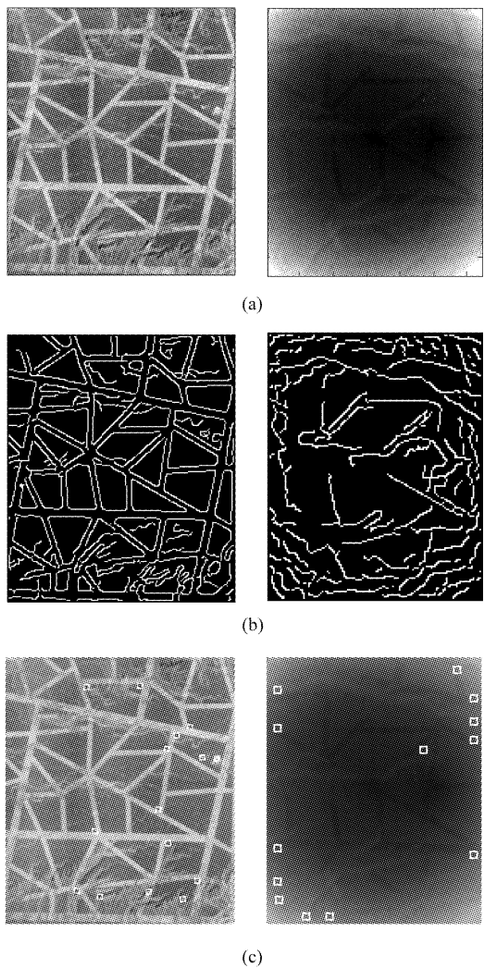

Fig. 1 Original image pair and interest features (a) Optical image and infrafed image. (b) Edge image extracted by Canny. (c) Interest points extracted by Harris detector

图 1 原始图像对和兴趣特征(a)光学图像和红外图像(b)Canny 算子提取的边缘图像(c)Harris 检测子提取的兴趣点

matching algorithm. At last, the random sample consensus (RANSAC) is employed to obtain the accurate transformation. This approach is capable of dealing with photometric and geometric variations, occlusions and view change, because the proposed method makes use of the local affine covariant regions extracted from the gray gradient image.

The main contribution of this paper is the introduction of the new affine covariant region detector SLER. Typically, affine covariant regions serve as measurement regions and tentative correspondences are determined by comparing invariants using a least-

squares approach. In this paper, the hypergraph matching algorithm is introduced to define the correspondences, which can substantially improve the registration accuracy.

1 Related work

Among multimodality image registration methods, a classic method using an invariant similarity measure is mutual information (MI)[1], which represents the statistical correlation of intensity values between two images. This method assumes that the statistical correlation between two images is globally stable. However the assumption is often violated. Statistical correlation between raw multimodal images tends to drop when spatial resolution drops[2]. Moreover, the approaches based on MI may suffer from a local maxima or an incorrect global maximum problem if intensity mapping relations in multimodal images are not spatially invariant or not globally statistical[3]. In order to alleviate the above problems, gradient information was combined with MI. In order to align locations with high gradient magnitudes and similar gradient orientations, the invariant similarity measure[4] is obtained by multiplying MI with a local gradient term between two images. Gradient vector flow (GVF) and intensity values are combined into a GVF-intensity (GVFI) map, and MI is calculated from GVFI map[5]. Considering that the gradient magnitudes may not provide reliable information in multimodal images, the edge orientation coincidence as well as pixel intensities in these two images is used to build a 3-D joint histogram, and then MI is calculated[6].

Another category of methods based on the invariant image representation uses an invariant image representation that is invariant to intensity changes such as contrast change as well as contrast reversal. Some examples of invariant image representations are edges, contours, and interest points. Once relevant features are extracted in both images, and the correspondences of the features are correctly obtained via similarity measure, typically using Mahalanobis distance. For reliable registration in these methods, features should be distinct, efficiently detectable and spreading over the images.

Hybrid visual features[7] are employed in visible and infrared image registration in man-made environments in two stages: global transform approximation and local transform adaptation. An affine invariant shape descriptor[8] is introduced, only the shape of the detected maximally stable extremal region (MSER)[9] is used to compute the affine invariant feature descriptor. The feature consensus approach[10] does not require to establish the correspondences. The transformation is reparameterized into a sequence of elementary stages. At each stage, a single transformation parameter is estimated by using feature consensus mechanism wherein the parameter value that is maximally consistent with all possible feature pairs is determined.

The primary advantage of this type of method is that the transformation matrix is computed in a single step, and is accurate if the feature matching is reliable. The drawback is that the feature matching is required, which is difficult to accomplish in a multisensor context, and is computationally expensive, unless the two images are already approximately registered, or the number of features is small.

2 Proposed algorithm

Comparing with the optical image, object contours in the infrared image are more obscure and often broken, but the object contours, in contrast with the other texture features, are still preserved intact. If the affine covariant regions surrounded by object contours are extracted, then the problem of extracting identical interest points can be resolved. It was found through analysis of experimental data that the adjoined extremal regions, which would merge into one, can approximate the real regions surrounded by real objects, and their contours are similar in optical image and infrared image. These are the regions in which we are interested. Considering the difficulty of extracting interest regions in infrared image, the grayscale image is mapped into the gradient image to strengthen the object contours and weaken other intensities. The above analysis is based on the fact that boundaries associated with the contours of the objects in images are mostly preserved [11].

Figure 2 displays the overall structure of the proposed algorithm.

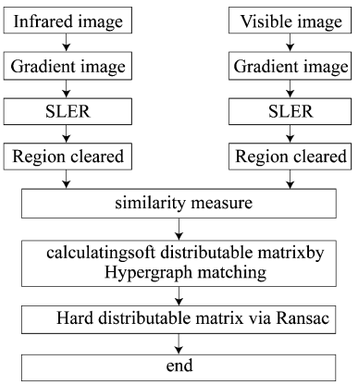

Fig. 2　Overall structure of the proposed algorithm
图 2　所提算法的总体结构

2.1　Gradient image

An image is convolved with a gradient operator (for instance, Prewitt and Sobel), which typically consists of horizontal and vertical gradient kernels. In this case, the gradient magnitude $|\nabla f(x,y)|$ shown below is obtained from a 2-D image:

$$|\nabla f(x,y)| = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad , \quad (1)$$

with

$$G_x(x,y) = f(x,y) * K_x(x,y) \quad , \quad (2)$$

$$G_y(x,y) = f(x,y) * K_y(x,y) \quad , \quad (3)$$

where $*$ denotes a 2-D convolution operation, $K_x(x,y)$ and $K_y(x,y)$ are the horizontal and vertical gradient kernels, respectively.

Then, the gradient magnitude is scaled into the range $[0, 255]$ as follows:

$$f(x,y) = \frac{f(x,y) - \min(f(x,y))}{\max(f(x,y)) - \min(f(x,y))} \quad . \quad (4)$$

The gradient images of the input images are used for registration in the following sections.

2.2　SLER algorithm

As introduced in the introduction section, most of the interest points detected using conventional algorithms from the multi-sensor image pair are not reproducible. The main reason is that all response values of the detector at each pixel position in an image are checked. Interest point is found at the pixel if the response value of this position is larger than a pre-defined threshold. In fact, this procedure is to select the top n pixels according to the rank of response value, where the number n is a controllable parameter. Since the infrared image has noticeable less texture than the optical image, it is unlikely that the n interest points extracted from optical image are identical with the n interest points extracted from infrared image. However, we have noticed that most of the contours indeed preserved in both visible and infrared images, since those pixels are usually on the boundaries of objects. Therefore, an interest region may be found at these locations surrounded by the contours.

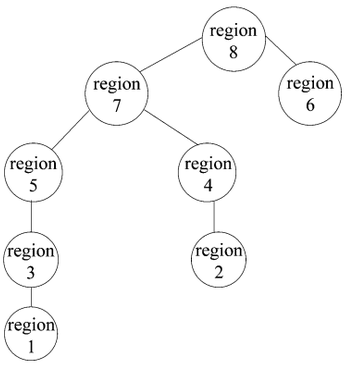

Fig. 3　Example of component tree
图 3　成份树示例

MSERs[9] denote a set of distinguished regions that are detected in a grayscale image. Due to the difference of intensities between the optical and infrared image, MSER cannot detect enough reproducible regions in both images. The detected regions are demonstrated in Fig. 4-6 (b). Our developed detector SLER, which is similar to MSER, extracts the same level and same father extremal regions in the component tree. For SLERs we only consider extremal regions which are defined by

$$\forall p \in R_i, \forall q \in \partial R_i \rightarrow I_{in}(p) < I_{in}(q) \quad . \quad (5)$$

The component tree is a rooted, connected tree and is built for gradient image with pixel values coming from an ordered set. These extremal regions, the nodes of the component tree, are identified as connected regions within binary image. Each node of the component tree is assigned one corresponding value g at which it is determined. The edges in the tree define an inclusion relationship among the connected regions. Thus, for a region R_j that is the father of a region R_i within the tree, the following inclusion relationship is fulfilled.

$$\forall p \in R_i \rightarrow p \in R_j \quad , \quad (6)$$

By moving in the component tree upwards, the corresponding value g of the extremal regions becomes higher, which leads to region sizes enlarging. The root of the component tree represents a region which includes all pixels of the input image. Fig. 3 shows part of the component tree.

For each connected region in the component tree, SLERs are defined as

$$\forall p \in R_i^g, \forall q \in R_j^g \rightarrow p, q \in R_k^{g+\Delta} \quad , \quad (7)$$

where R_i^g and R_j^g denotes an extremal region, Δ is a range parameter, and $R_k^{g+\Delta}$ is the extremal region which is obtained by moving upwards in the component tree from the region R_i^g and R_j^g. SLERs are the extremal regions in which the nodes are at the same level and have the same father in the tree. For instance, in the component tree shown in Fig. 3, some of the detected SLERs are the region 4 and region 5.

2.3 Hypergraph matching

Due to the low quality of the infrared image, only a few interest regions are extracted. In order to achieve accurate automatic registration, the spatial relationship among the features is employed. Not only the interest points but also the spatial relationships (for example, edges, angles) of the matched points are matched. A new high-order score (For simplicity, only third-order)[12] is defined:

$$score(X) = \sum_{i_1,i_2,j_1,j_2,k_1,k_2} H_{i_1,i_2,j_1,j_2,k_1,k_2} X_{i_1,i_2} X_{j_1,j_2} X_{k_1,k_2} \quad , \quad (8)$$

where H is a 6-D super-symmetric tensor, i. e., invariant under permutations of indices in $\{i_1, j_1, k_1\}$ or $\{i_2, j_2, k_2\}$.

In order to obtain the best performance, the weighted integration scheme of the different order tensors[12,13] is presented. Liu[14] pointed out that this scheme cannot improve the overall performance, because the distinguishability of the low-order tensor is always lower than that of the high-order tensor, and the weighted summation only reduces the overall discriminant. The formulation of the inner product is proposed[14] and the enhanced distinctive tensors are built as follows:

$$\overline{\mathbf{H}} = <\mathbf{H}, \mathbf{W}> \quad , \quad (9)$$

where $\mathbf{W}$ is a weighted tensor built with low-level features. Four ways[14] are explored to build $\mathbf{W}$ and $\overline{\mathbf{H}}$. In this paper, we follow the Priori constraints.

3 Experimental results

The performance of the proposed method is evaluated on real optical and infrared images.

For all experiments, the same set of parameters given by the authors is used for each detector, except the detected region area which is in [100, N/20], where N denotes all pixel numbers of the image. In our proposed method, the ratio of the convex hull of the region to the region area is used to exclude those regions with a value of the ratio larger than a pre-defined threshold, such as 2.5. Like the framework in Ref. [15] the detected regions are represented by ellipses.

In order to show a visual assessment and comparison of the detected results, Figs. 4-6 display the experimental results. For each figure, (a) shows the original infrared and optical images, respectively, (b) displays regions detected by MSER, and extracted regions by SLER are illustrated in (c). It is noted that the proposed detector extracts more affine covariant regions than MSER for all infrared images. Most of the regions detected by SLER are visible object regions, and their contours can approximate the real objects as well. However, MSER has achieved good performance in the optical image, and the detected regions evenly spread throughout the image. The number of the detected regions using SLER is more than MSER. Obviously, registration parameters cannot be obtained using the regions extracted by MSER.

As the ground truth was unknown for the sets of real image pairs, it was estimated by the given GPS information. In our experiments, the image pairs with mainly translation transformation are examined, and the number of the experimental image pairs is 100. There are 47 image pairs which the alignment error is not more than 1 pixel. If the alignment error in the pixel positions determined by both the estimated parameters and the ground truth is less than three pixels, we consider it a successful registration. The successful registration rate reaches 82%. The corresponding success rates are summarized with the alignment error range in Table 1. Three registration examples with the

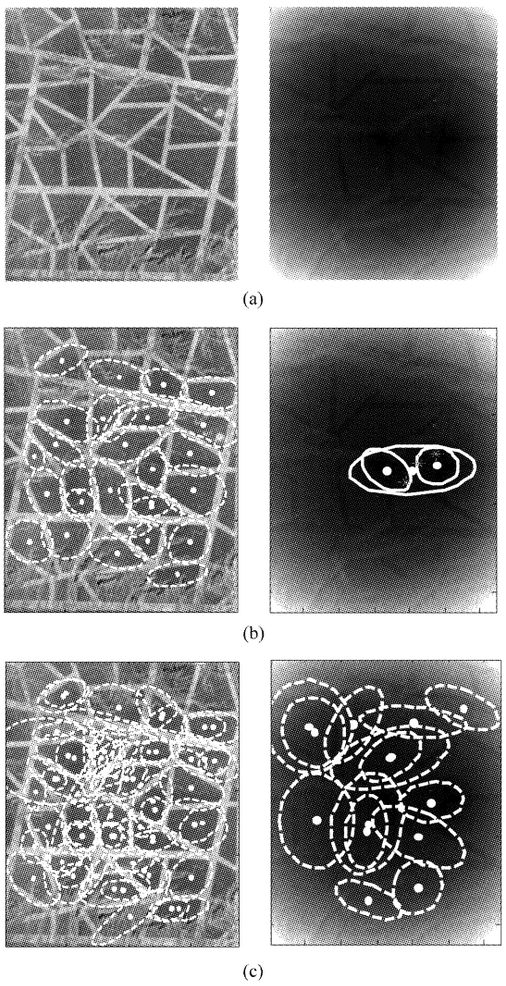

Fig. 4　Interest region - test image pair 1 (a) Optical image and infrared Image (b) Interest regions extracted by MSER (c) Interest regions extracted by SLER
图 4　兴趣区域 - 测试图像对 1(a)光学图像和红外图像(b)MSER 算子提取的兴趣区域(c)SLER 算子提取的兴趣区域

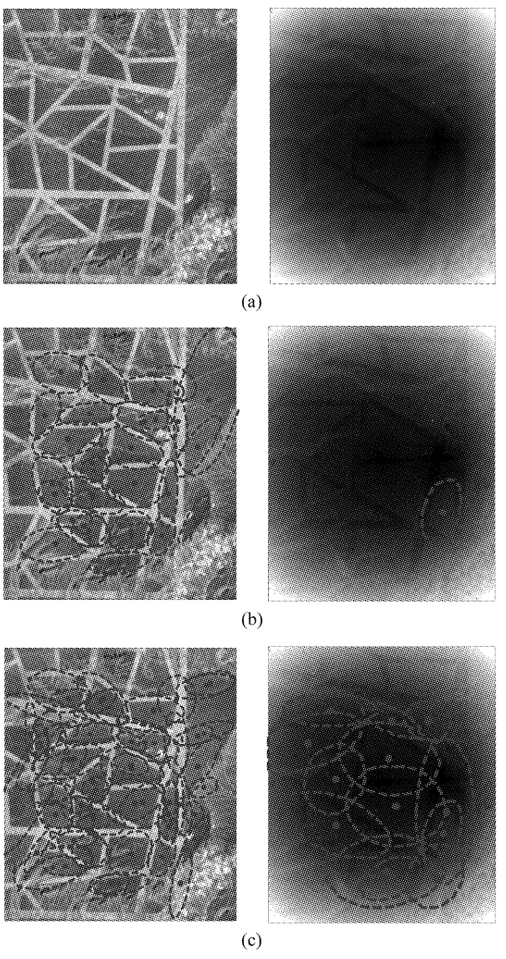

Fig. 5　Interest region - test image pair 2. (a) Optical image and infrared Image. (b) Interest regions extracted by MSER. (c) Interest regions extracted by SLER
图 5　兴趣区域 - 测试图像对 2(a)光学图像和红外图像(b)MSER 算子提取的兴趣区(c)SLER 算子提取的兴趣区域

proposed algorithm are shown in Figs. 7-9 separately.

Table 1　The relationship between the successful registration rate (SRR) and maximum error (ME)
表 1　成功配准率与最大误差之间的关系

ME (pixel)	[0,1)	[0,2)	[0,3)	[0,4)	[0,6)
SRR	47	70	82	89	95

By analyzing the experimental data with an alignment error exceeding one pixel, the main reason for low accuracy or failure is that the initialization of the tensor is random and the feature point location is not accurate. The tensor initialization method needs to be researched and the positioning accuracy of interest points should be improved for future study.

In the case of rotation and scaling on infrared image, the correspondences of the interest points are demonstrated in Figs. 10-11. The infrared image is rotated 5.25 degrees and scaled 1.15 times, respectively. In the case of large scale, the correspondences are hard to be found, because the priori knowledge of the same scale is used.

We also compared the registration results obtained from the proposed algorithm with the classical method based on SIFT. For each image pair, we extracted

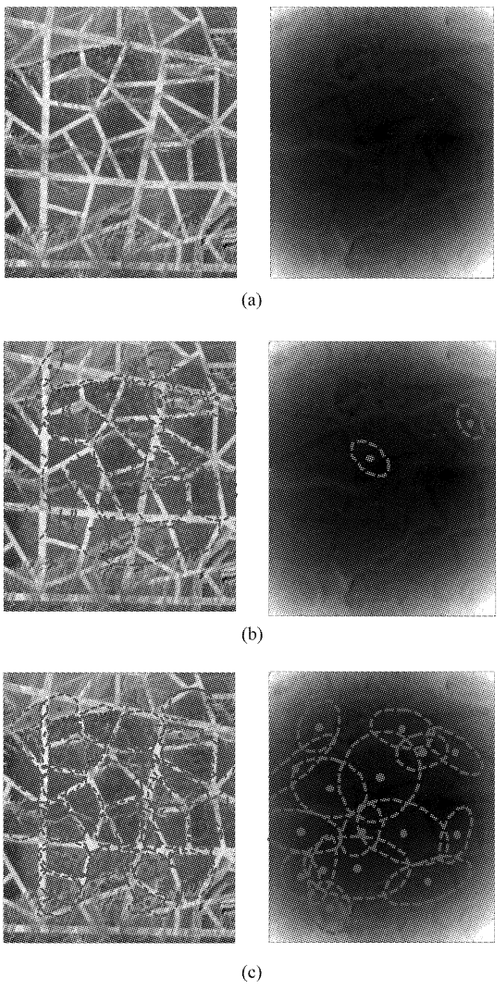

Fig. 6 Interest region - test image pair 3. (a) Optical image and infrared Image. (b) Interest regions extracted by MSER. (c) Interest regions extracted by SLER
图6 兴趣区域 - 测试图像对3(a)光学图像和红外图像(b)MSER算子提取的兴趣区域(c)SLER算子提取的兴趣区域

SIFT descriptors of the SIFT key points and matched them by the nearest distances of their descriptors. Because SIFT descriptor is calculated from a patch around the center of the interest point, and the pixel intensities in the patches of the identical key points both in optical and infrared image have no corresponding relationship, almost no correspondences can be obtained.

4 Conclusion

In this paper, the problem of multimodality image automatic registration is addressed in some applications. A novel detector SLER is suggested to extract the affine covariant regions of an image which includes structural objects within the scene. The detector can effectively extract stable regions. Thereafter, these interest regions are used to build a high-order distinctive tensor using an inner product. Finally, the hypergraph matching method is applied to obtain accurately the transformation parameters. The explored registration

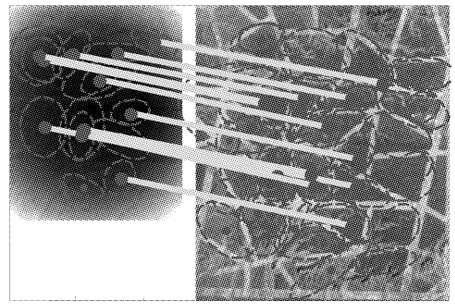

Fig. 7 Registration example 1
图7 配准示例1

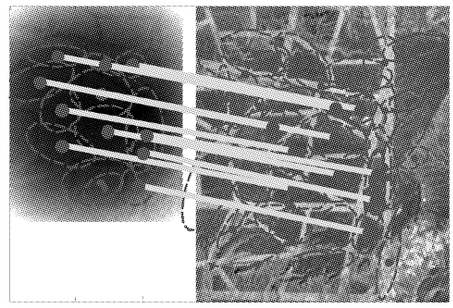

Fig. 8 Registration example 2
图8 配准示例2

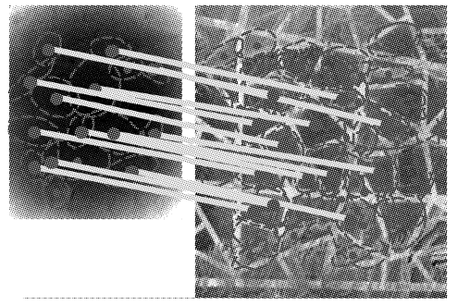

Fig. 9 Registration example 3
图9 配准示例3

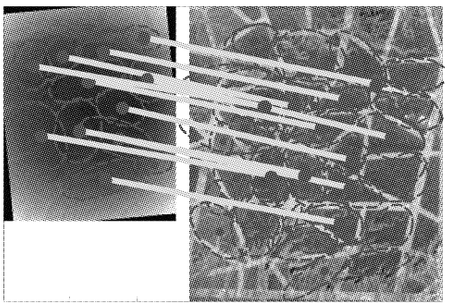

Fig. 10　Registration for the case of image rotation
图 10　图像旋转情况下的配准

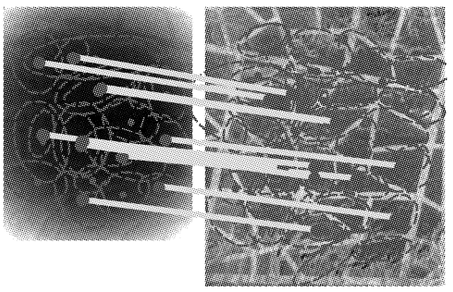

Fig. 11　Registration for the case of image scaling
图 11　图像缩放情况下的配准

scheme is robust to variations in sensor characteristics, imaging conditions, and fields of view thus can be used to register several classes of multimodality and multisensor images. The global registration achieved by the proposed method should be available for many applications.

REFERENCES

[1] Viola P, Wells III W M. Alignment by maximization of mutual information[J]. *International journal of computer vision*, 1997, **24**(2): 137-154.

[2] Irani M, Anandan P. Robust multi-sensor image alignment [C]. *Proceedings of the* 18*th International Conference on computer vision*, 1998: 959-966.

[3] Keller Y, Averbuch A. Multisensor image registration via implicit similarity[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(5): 794-801.

[4] Pluim J P, Maintz J A, Viergever M A. Image registration by maximization of combined mutual information and gradient information: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2000, 2000[C]. berlin Heidelberg: Springer-verlag, 2000,1935: 452-461.

[5] Guo Y, Lu C C. Multi-modality image registration using mutual information based on gradient vector flow[C]. *Proceedings of the* 18*th International Conference on Pattern Recognition*, 2006,3: 697-700.

[6] Kim Y S, Lee J H, Ra J B. Multi-sensor image registration based on intensity and edge orientation information[J]. *Pattern recognition*, 2008, **41**(11): 3356-3365.

[7] Han J, Pauwels E J, de Zeeuw P. Visible and Infrared Image Registration in Man-Made Environments Employing Hybrid Visual Features [J]. *Pattern Recognition Letters*, 2012.

[8] Forssen P E, Lowe D G. Shape descriptors for maximally stable extremal regions[C]. *Proceedings of the* 11*th International Conference on Computer Vision*, 2007: 1-8.

[9] Matas J, Chum O, Urban M, *et al*. Robust wide-baseline stereo from maximally stable extremal regions[J]. *Image and vision computing*, 2004, **22**(10): 761-767.

[10] Shekhar C, Govindu V, Chellappa R. Multisensor image registration by feature consensus[J]. *Pattern recognition*, 1999, **32**(1): 39-52.

[11] Ivarez N, Sanchiz J, Badenas J, *et al*. Contour-based image registration using mutual information[J]. *Pattern Recognition and Image Analysis*, 2005: 405-411.

[12] Duchenne O, Bach F, Kweon I-S, *et al*. A tensor-based algorithm for high-order graph matching[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(12): 2383-2395.

[13] Lee J, Cho M, Lee K M. Hyper-graph matching via reweighted random walks[C]. *Proceedings of the* 2011 *IEEE Conference on Computer Vision and Pattern Recognition*, 2011: 1633-1640.

[14] Liu Cong-Xin. Study on local invariant feature and high-order matching[D]. Shanghai: Shanghai Jiao Tong University (刘从新. 局部不变特征及高阶匹配技术研究. 上海交通大学), 2012.

[15] Mikolajczyk K, Tuytelaars T, Schmid C, *et al*. A comparison of affine region detectors[J]. *International journal of computer vision*, 2005, **65**(1): 43-72.