2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

IR Remote Sensing Image Registration Based on Multi-scale Feature Extraction

Jun Kong, Min Jiang
Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education,
Jiangnan University
Wuxi 214122, China
Email: kongjun@jiangnan.edu.cn

Jun Kong
College of Electrical Engineering,
Xinjiang University,
Urumqi 830047, China

Jun Kong, Yi-Ning Sun
Institute of Intelligent Machines,
Chinese Academy of Sciences
Hefei 230031, China

***Abstract*—Infrared remote sensing image has poor contrast and lower SNR so that real-time and robustness are not superior in image registration. In order to solve it, a novel registration based on Multi-scale feature extraction is proposed in this paper. This algorithm is designed in two aspects. Firstly, Gaussian convolution template size adjusts adaptively with the increasing of scale factors. Then the Multi-space is reconstructed. Secondly, feature points bidirectional matching based on the City-block distance is introduced into image registration. So the real-time performance and robustness are enhanced further. Finally, the experimental results showed that by this improved algorithm the infrared remote sensing images are registered more quickly and accurately than by traditional SIFT algorithm.**

Keywords*—*IR remote sensing image; registration; Multi-scale; feature points

I. INTRODUCTION

Infrared (IR) remote sensing images are commonly used in the field of computer vision, resource exploration, target identification and tracking, and other military applications, etc[1, 2]. However, IR remote sensing image has some characteristics such as low contrast and low SNR, highly bright background[3], and more IR sensor noise interference so on. As a result, IR image registration and target recognition were more difficult than visible image[4]. Image registration is a work of establishing point-to-point correspondence between images of the same scene taken at different time[5], which is a fundamental task in visible light image processing [6-8] and IR remote sensing image system. A good image registration technique should be able to quickly and correctly identify the corresponding regions and determine the appropriate geometric transformation required to map the sensed image on the reference image, despite the presence of varying imaging conditions[9]. Therefore, how to realize the highly accurate image registration has become a critical hotpot problem in IR remote sensing image processing.

The paper is sponsored by National Natural Science Foundation of China (61362030, 61201429), Xinjiang Uygur Autonomous Regions Natural Science Foundation (201233146-6), Xinjiang Uygur Autonomous Regions University science and Research Key Project (XJEDU2012I08), the 111Project (B12018).

There are a plenty of research methods for the registration in recent years, which can be divided into two types generally: method based on feature and method based on image area [10]. The gloomy feature of IR remote sensing image is often inconsistent, so the method based on area can’t be used in this situation. Although the IR remote images have poor contrast and low SNR, some feature could easily be extracted from these images. And feature-based image registration algorithm is more suitable for the instance of rotation, illumination and zoom of IR remote sensing image. Aimed to it, a novel algorithm based on Multi-scale feature extraction[11] (MSFE) will be proposed. Firstly, by Gaussian template size adaptively adjustment, the Multi-scale space is reconstructed to decrease IR image registration time and to improve algorithm real-time. Secondly, by feature points bidirectional matching technology based on the City-block distance, we could augment the amount of feature points correctly matched (at the same time, the amount of feature points falsely matched was decrease) to improve algorithm’s accuracy/robustness.

II. ALGORITHM BASED ON MULTI-SCALE FEATURE EXTRACTION

In visible image registration, the performance always has a few disadvantages such as: target image’s rotation, affine transformation, scaling, and illumination variation so on. IR remote sensing image has not only those shortcomings mentioned above but also low contrast and poor visual effect. Therefore, with IR remote sensing image, the performance of registration has much more challenges. In this paper, algorithm based on MSFE has several aspects as follows. To adaptively reconstruct the Multi-scale space and to extract the feature points in this scale space. These feature points are still invariant in scale, rotation and position so on. The feature descriptors are calculated and they have stronger robustness. Bidirectional match based on the City-block distance will be introduced in registration of IR remote sensing images.

A. Reconstruction of Multi-scale Space

Multi-scale space $D(x, y, \sigma_n)$ is constructed by image Gaussian Pyramid[12] (GP) which is produced by the convolution of the second-order Gradient equation of Gaussian function with an input image $I(x, y)$. This function has variable scale factors σ_n, giving

$$D(x, y, \sigma_n) = [\sigma_n^2 \nabla^2 G(x, y, \sigma_n)] * I(x, y) \tag{1}$$

For the experiments in this paper, the factors σ_n vary as Equation (2), where $s = 3$. So every $D(x, y, \sigma_n)$ has 6 layers.

$$\sigma_n = \sigma \times 2^{n/s}, n = 0,1,\ldots, s+2, \tag{2}$$

Gaussian convolution template size is $N_n \times N_n$, where N_n can be computed as follow empirical formula (3).

$$N_n = \frac{9 \times \sigma_n}{1.2} \tag{3}$$

TABLE I. RELATIONSHIP BETWEEN σ_n AND N_n

$\sigma_n = \sigma \times 2^{n/s}$	$N_n = \frac{9 \times \sigma}{1.2}$	Size of Gaussian filter template $N_n \times N_n$
0.98	7.35	7×7
1.23	9.26	9×9
1.56	11.66	11×11
1.96	14.70	15×15
2.47	18.52	19×19
3.11	23.34	23×23

The relationship between scale factors σ_n and Gaussian filter template N_n is shown in TABLEI. In the paper, the initialization value of the scale factor is set to 0.98 empirically.

Theoretically, the template size of N_n turn larger with the increment of the scale factor value of σ_n. As a result, the image smoothness/fuzziness will become noticeable, but the convolution time will be increased too. The traditional SIFT matching algorithm, a registration based on feature, was proposed in 1999 and developed in 2004 by David G. Lowe[13, 14]. The SIFT algorithm proposed a method for extracting distinctive invariant features that are invariant to image scale and rotation, illumination, change in 3D viewpoint, affine transformation and so on. In the SIFT algorithm, the construction of scale space is in line with the principle that the template size will turn larger with the increment of the scale factor. Nonetheless, some experimental results showed that the feature points were not just increasing linearly with the increase of the scale factor. Beyond a certain critical point where $\sigma \in (1{,}20{,}1.70)$, the feature points will decrease. Namely, once template size exceeds a certain value, the feature points acquired are not more but fewer. Consequently, the computational time will be prolonged and real-time performance will be reduced deeply. From TABLEI, when the template size of the Gaussian filter is just 9×9 or 11×11 the amount of feature points will reach maximum. In MSFE algorithm, the template size will be auto-adaptively adjusted with the increasing of the scale factor in the course of reconstruction of image GP, where the initial value is 0.98. Namely, if $\sigma_n < 1.50$, the template of 9×9 is accepted; if $\sigma_n > 1.50$, the template size is 11×11. In this paper, the procedure of reconstruction of Multi-scale space may be summarized as follows. Given two images, one is registered image (I_r), the other is target image (I_t).

Step1: Take I_r as the first layer image of GP, and then convolve it with the Gaussian template of 9×9 to produce the second layer image. Similarly, the second layer image is sequentially convolved with the template of 9×9 to produce the third layer image of GP.
Step2: The third layer image is convolved with the template of 11×11 to produce the fourth layer image of GP. Similarly, the fifth and the sixth layer image are produced.
Step3: Based on **Step1** and **Step2**, this original image will produce the first image GP (first octave).
Step4: The initial image of the first octave is down-sampled by a factor 2 and it is taken as the initial image of the second image GP. Repeating **Step1**, produce the second image GP (second octave). Similarly, produce the third octave, …, the sixth octave.
Step5: In every octave, adjacent images are subtracted to produce the different-of-Gaussian Pyramid (DOGP). There will be six in all.
Step6: Repeat **Step1~Step5** for I_t.

B. Local Feature Points Detection

Multi-scale space $D(x, y, \sigma)$ will be expanded by the Taylor series[15] and its derivatives are measured at potential feature points. From calculating extreme, introducing 2×2 Hessian matrix, the final highly accurate feature points are acquired.

Detection of feature points is the fundamental stage in the proposed MSFE algorithm. It is implemented efficiently by using Multi-scale space to identify potential feature points that are invariant to scale and orientation. In order to detect them, each potential point is compared with its 26 neighbors in its $3 \times 3 \times 3$ cubic neighborhood. Only the point bigger/smaller than the other neighbors will be chosen as a feature point. In order to enhance algorithm robustness, those unstable points must be eliminated.

The Multi-scale space $D(x, y, \sigma)$ will be expanded by the Taylor series, giving

$$D(x) = D + \frac{\partial D^T}{\partial X} \Delta X + \frac{1}{2} \Delta X^T \frac{\partial^2 D}{\partial X^2} \Delta X \tag{4}$$

D and its derivatives are evaluated at the potential feature point. $X = (x, y, \sigma)^T$ is the offset from this point. Derivate (4) and order it to be zero, then the exact locations of the extreme $X_{\max}$ are obtained, giving

$$X_{\max} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X} \tag{5}$$

Equation (5) is substituted into (4), then (6) is obtained, which is used to reject unstable and low contrast points.

$$D(X_{\max}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} X_{\max} \tag{6}$$

For the experiments in this paper, if $|D(X_{\max})| \geq 0.03$ (after normalized), the potential feature points will be reserved, otherwise they will be discarded.

In order to further optimize the low contrast feature points and/or unstable edge ones, the principal curvatures are calculated by 2×2 Hessian matrix H. The eigenvalues of H are proportional to the principal curvatures of D.

$$H = \begin{vmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{vmatrix} \tag{7}$$

Let α and β respectively be the maximum and minimum eigenvalue of H. Let γ be the ratio, where $\gamma = \alpha / \beta$.

$$\begin{aligned} Tr(H) &= D_{xx} + D_{yy} = \alpha + \beta \\ Det(H) &= D_{xx} D_{yy} - D_{xy}^2 = \alpha\beta \end{aligned} \tag{8}$$

$$\begin{aligned} &Tr(H)^2 / Det(H)^2 = (\alpha + \beta)^2 / \alpha\beta \\ &= (\gamma\beta + \beta)^2 / \gamma\beta^2 = (\gamma + 1)^2 / \gamma \end{aligned} \tag{9}$$

If $Tr(H)^2 / Det(H)^2 < (\gamma + 1)^2 / \gamma$,the point will be reserved. Or it will be discarded. The paper proposes $\lambda = 10$.

Similarly, calculate feature points in the all scales of all Multi-scale space, the intersection of which are the final feature points.

C. Feature Descriptor Generation

In order to be invariant to rotation, scale change, illumination, feature descriptor is acquired as follows. Firstly, take a certain feature point as the center to form a circular neighborhood with radius 6σ (where σ is the scale at which the feature points were detected and the side length of wavelet is 4σ). Secondly, calculate the Harr wavelet responses in horizontal and vertical direction and calculate the weighted Harr wavelet responses with histogram. Finally, use a sliding window of $\pi/3$ to get the total response in horizontal and vertical direction respectively and select the direction of the longest vector as the main orientation.

With the feature point as the center, construct a square window region which size is 20σ , and the window region is divided 4×4 sub-regions. Compute the sum and absolute value of the response, thus in each sub-region vector will form the four-dimensional one as (10).

$$V_{sub} = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \tag{10}$$

dx denotes Harr response in horizontal direction and dy denotes response in vertical direction. For each feature point, the $(4 \times 4) \times 4 = 64$ dimensional description vector will be formed. With it being normalized, the feature description will have strong robustness to illumination, so that the algorithm is easily used in feature point extraction and matching for IR remote sensing image.

D. Feature Points Bidirectional Matching

According to traditional SIFT registration algorithm, taking two IR remote sensing images (one is I_r , the other is I_t), the feature match is by calculating all the Euclidean distance (spatial-distance) between a definite feature point in I_r and every feature points in I_t . If the ratio of the nearest distance to the second nearest distance is less than a certain threshold, the nearest matching points are got. But there are two disadvantages in the algorithm above.

The Euclidean distance is defined as a distance between the feature points X and Y in 64-dimension space, where X_{1i} denotes the ith feature descriptor of a certain feature point in one image, X_{2i} denotes the ith feature descriptor in the other image, giving

$$L_{Euclidean} = \sqrt{\sum_{i=1}^{64} (X_{1i} - X_{2i})^2} \tag{11}$$

The City-block and the Chessboard distance respectively are described as (11) and (12).

$$L_{City-block}(X_1, X_2) = \sum_{i=1}^{64} |X_{1i} - X_{2i}| \tag{12}$$

$$L_{Chessboard}(X_1, X_2) = \max_{1 \leq i \leq 64} \{ |X_{1i} - X_{2i}| \} \tag{13}$$

Comparing formula (11), (12) and (13), we can find that computation of the City-block and the Chessboard distance are smaller than the Euclidean distance. In addition to, among them the City-block distance has been accurate than the others. So in the paper, the City-block distance is used in image matching.

Furthermore, the traditional SIFT algorithm adopts single-direction matching so that two or more points in I_t can match simultaneously with the same point in I_r. Avoiding this, reverse matching is added to the traditional matching to produce bidirectional matching for improving algorithm accuracy and stability further. And error-matched points can be removed to a great extent.

IR remote sensing images have poor contrast and lower SNR which likely caused error-matching and missing-matching in registration. Therefore, the paper proposed a novel way based on the City-block distance and bidirectional matching. The procedure may be summarized as follows.

Step1: Compute the constraint of epipolar line so as to reduce ergodic areas in I_r and I_t.

Step2: Calculate the City-block distance between a certain feature point F_1 in I_r and all feature points in I_t based on corresponding epipolar line. Search/find the nearest point F_2. If the distance is less than the threshold set, reserve the point F_2 then skip into **Step3**. Otherwise, discard F_1 and then repeat **Step2**.

Step3: Calculate the City-block distance between the feature point F_2 in I_t and all feature points in I_r based on corresponding epipolar line. Search and find the nearest point F_3.

Step4: If $F_1 = F_3$, match successfully. Otherwise, match unsuccessfully and repeat **Step2**.

III. Experimental Results and Discussions

To evaluate the performance of the proposed registration method, we use a pair of IR remote sensing images which were taken by a certain IR thermal imager as shown in Fig1. They have low contrast and much noise. The right image (I_t) of Fig1 is rotation in comparison with the left image (I_r), whose sizes are all 500×800. In this section, IR images are registered by the way of SIFT and MSFE.

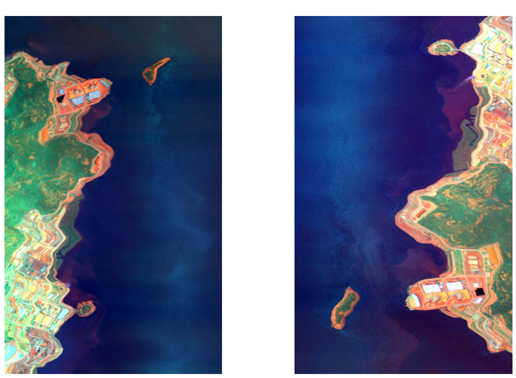

Fig. 1. Original IR remote sensing images

A. Experiment Based on SIFT Algorithm

By traditional SIFT algorithm, calculate all feature points in I_r and I_t. There are 2170 feature points in I_r and 2093 in I_t. Vectors of feature points are illustrated in Fig2. These vectors can describe the location, scale, and orientation of each feature point.

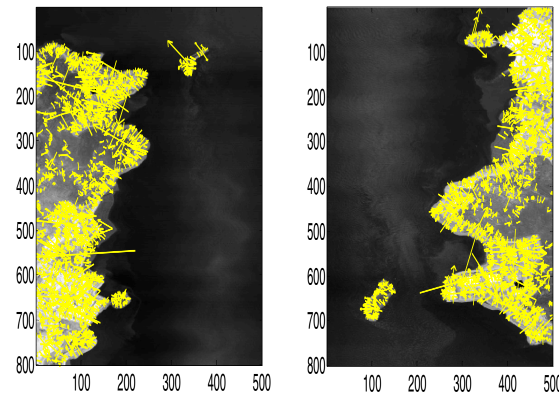

Fig. 2. Vectors of feature points in registered image and target one by SIFT

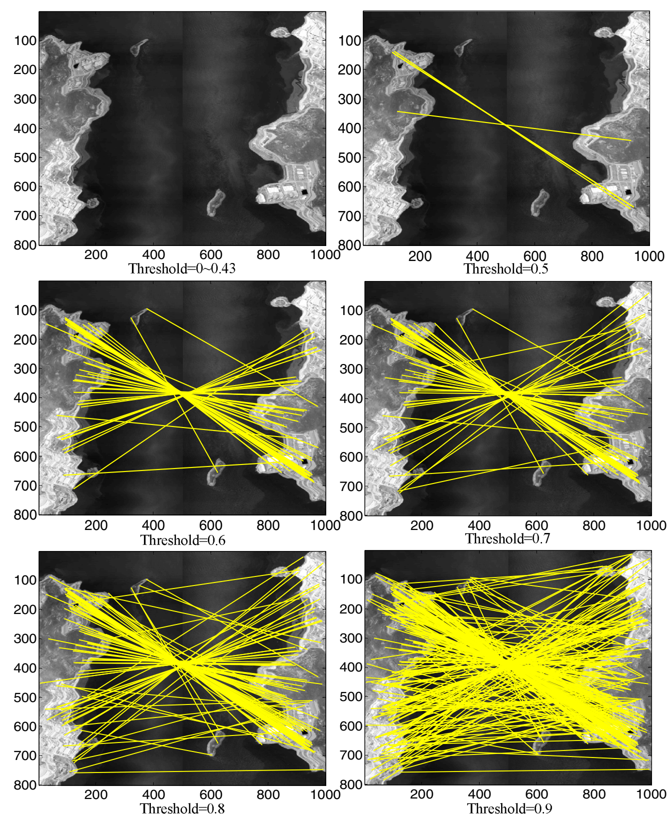

Fig. 3. IR images registration based on SIFT algorithm

Match I_r with I_t separately with different thresholds as shown in Fig3. The result of performance is shown in TABLEII. The conclusion can be drawn that: (1) the quantity of feature points has nothing with threshold; (2) if the threshold is too low, missing-matching occurs frequently; (3) the rates of error-matching and correct-matching are all increasing with the increasing of threshold; (4) matching time is about 9.135 seconds so algorithm's real-time is not high.

TABLE II. RESULT OF IR REMOTE SENSING IMAGE REGISTRATION BASED ON SIFT ALGORITHM

Threshold	N_r and N_t	M	Registration time (s)
0~0.43	2170、2093	0	9.0974
0.5	2170、2093	5	9.1036
0.6	2170、2093	31	9.1633
0.7	2170、2093	65	9.1223
0.8	2170、2093	108	9.1739
0.9	2170、2093	201	9.1499

B. *Experiment Base on MSFE Algorithm*

By improved MSFE algorithm, calculate all feature points in I_r and I_t. There are 430 feature points in I_r and 426 in I_t. Feature points are labeled with "+", as shown in Fig4. The circles represent vectors of feature point, whose radius give location, scale, and orientation. Fig 4 just shows parts of vector descriptors (which having the mark of a circle).

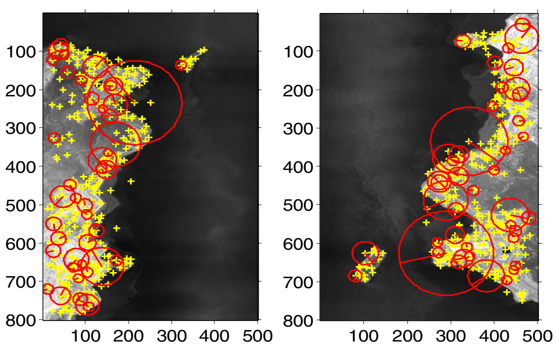

Fig. 4. Vectors of feature points in registered image and target one by MSFE

Match I_r with I_t separately with different thresholds as shown in Figure5, and the result of performance is shown in TABLEIII. The conclusion can be drawn that: (1) overcome the disadvantage of missing-matching when the threshold is more lower; (2) by bidirectional matching technology, error-matching rate decreases; (3) error-matching rate and correct-matching rate are all increasing with the increasing of threshold; (4) matching time is about 2.462 seconds so algorithm's real-time is improved significantly.

TABLE III. RESULT OF IR REMOTE SENSING IMAGE REGISTRATION BASED ON MSFE ALGORITHM

Threshold	N_r and N_t	M	Registration time (s)
0~0.16	430、426	0	2.4184
0.5	430、426	11	2.4727
0.6	430、426	23	2.4674
0.7	430、426	37	2.4571
0.8	430、426	73	2.4773
0.9	430、426	154	2.4778

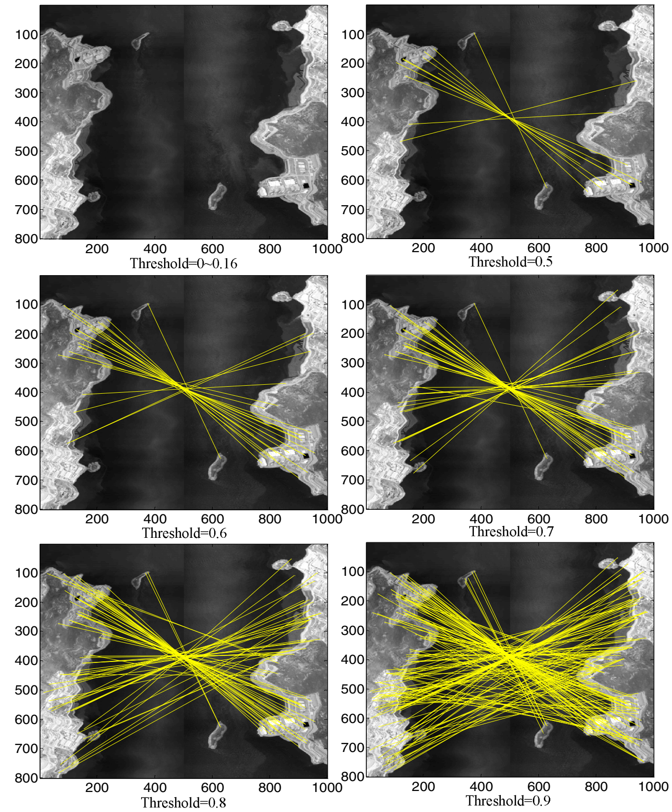

Fig. 5. IR images registration based on MSFE algorithm

C. *Improvement of matching accuracy*

In order to verify the effectiveness of the proposed algorithm, MR (Matching Rate) and FAR (False Acceptance Rate) are introduced in this paper. Let N_r and N_t be the quantity of feature points that are independently detected in I_r and I_t. Let M be the quantity of matched points which can be divided into the quantity of correctly matched feature points $M^{correct}$ and the quantity of falsely matched feature points M^{error}, giving

$$MR = \frac{M^{correct}}{\min(N_r, N_t)} \times 100\% \ , \qquad FAR = \frac{M^{error}}{\min(N_r, N_t)} \times 100\% \tag{14}$$

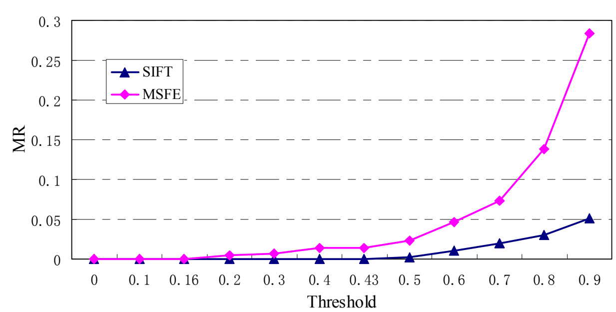

Fig. 6. Curve of relation between Threshold and MR

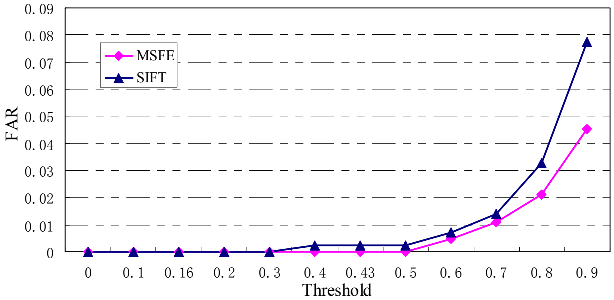

Fig. 7. Curve of relation between Threshold and FAR

Fig6 and Fig7 respectively show the relationship between SIFT and MSFE about *MR* and *FAR* . The conclusion can be drawn that at the same threshold, MSFE algorithm has much higher *MR* and lower *FAR* than SIFT algorithm. So the accuracy of MSFE algorithm is greatly improved.

D. Another Experiment Illuminated

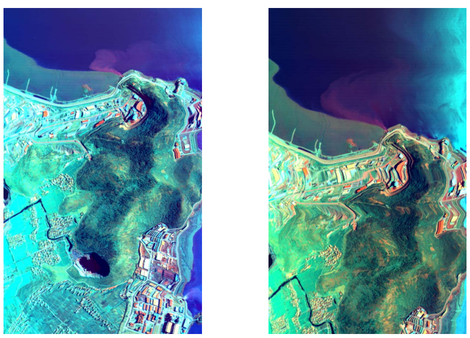

Fig. 8. Another original IR remote sensing images.

In order to further demonstrate the effectiveness of proposed MSFE algorithm, the more experiments on different images should be performed. Another IR remote sensing images were acquired whose size are all 500×800, having low contrast and much noise, as shown in Fig.8. The right image has zoom and deformation compare to the left image. In this section, the IR images are registered by the way of SIFT and MSFE.

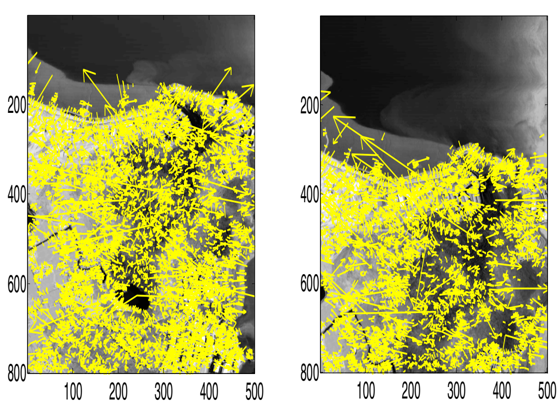

Fig. 9. Vectors of feature points in registered image and target one by SIFT

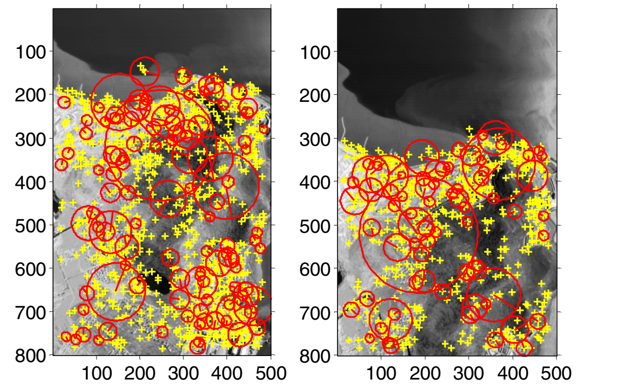

Fig. 10. Vectors of feature points in registered image and target one by MSFE

By SIFT, 6553 feature points were calculated in I_r and 4247 in I_t, as shown in Fig9. By MSFE, 1109 feature points were gotten in I_r and 722 in I_t, as shown in Fig10.

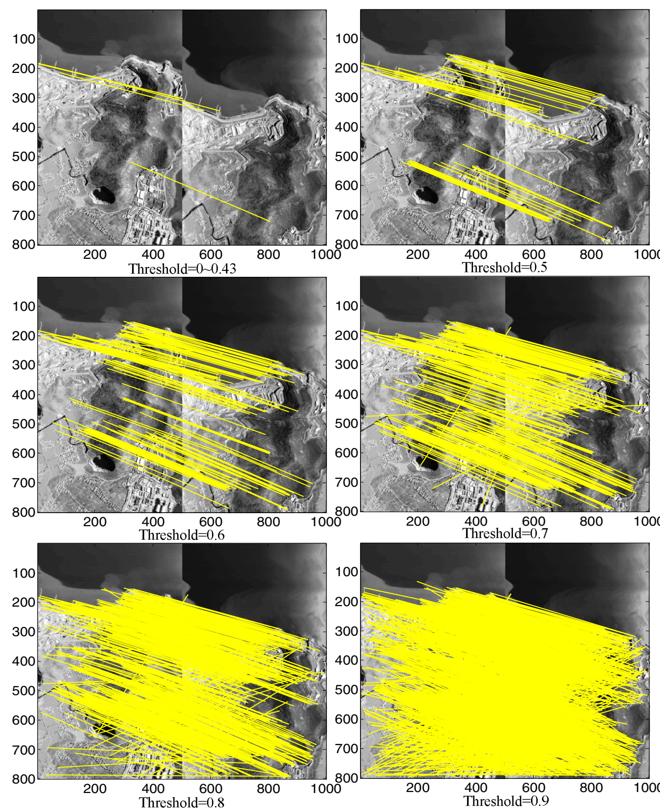

Fig. 11. Another IR remote sensing images registration based on SIFT algorithm

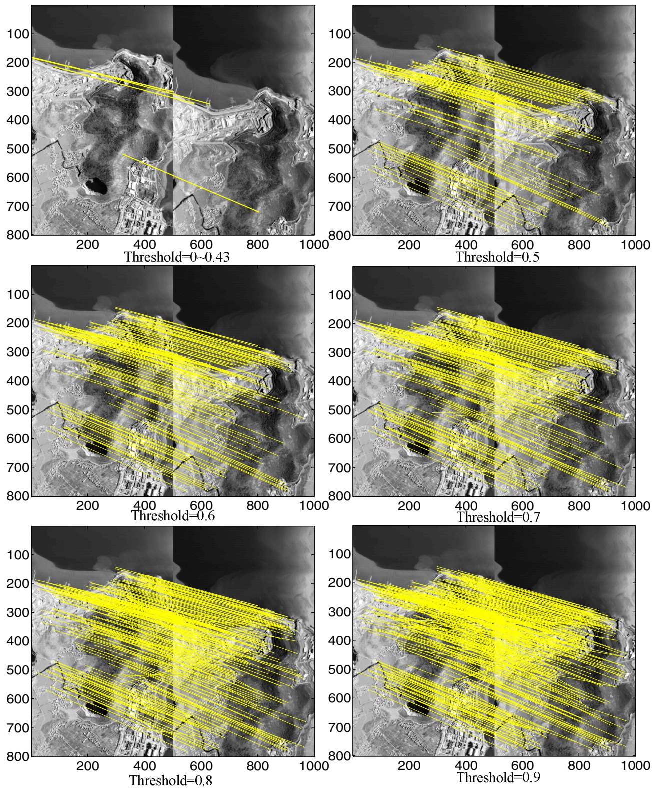

Fig. 12. Another IR remote sensing images registration based on MSFE algorithm

Fig11 and Fig12 present the registration results by SIFT and MSFE respectively. Compared with the Fig11, the Fig12 demonstrated that the MSFE algorithm is more accurate than the traditional SIFT algorithm. From the experiment, average matching time of MSFE is 2.625 seconds, and the SIFT processing time is 12.498 seconds. So in the MSFE algorithm, the robustness and real-time were all significantly improved performance. This experiment had further proved it.

IV. CONCLUSIONS AND FUTURE WORKS

According to the IR remote sensing image's low contrast, low SNR, highly bright background, and IR sensor noise interference, the traditional SIFT algorithm used in registration was insufficient in real-time and robustness. Therefore, a novel algorithm based on MSFE was utilized in the IR remote sensing image registration. By the way of Gaussian convolution template size adaptive adjustment and by the way of feature points bidirectional matching based on the City-block distance, the proposed registration provides less computational time and more correctly matched feature points. Finally, the experimental results showed that with the proposed algorithm, the IR remote sensing images are registered more quickly and accurately. The algorithm was optimized in both aspects of real-time and accuracy. This research has good prospects in some areas such as IR image mosaic, IR target recognition and tracking etc. In future works, we will combine the proposed algorithm and RANSAC to improve the real-time performance and robustness.

REFERENCES

[1] J. Xie, B.-H. Li, W. Han, J.-H. Bao, F. Gu, F. Guo, "A tiny facet primitive remote sensing image registration method based on SIFT key points," in Instrumentation & Measurement, Sensor Network and Automation (IMSNA), IEEE International Symposium on, vol. 1, Sanya, 2012, pp. 138-141,.

[2] Q.-Q. Huang, J. Yang, C.-G. Wang, J.-B. Chen, Y. Meng, "Improved registration method for infrared and visible remote sensing image using NSCT and SIFT," in Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, Munich, 2012, pp. 2360-2363.

[3] C.-L. Lin, "An approach to adaptive infrared image enhancement for long-range surveillance," Infrared Physics & Technology, vol. 54, pp.84-91, 2011.

[4] A.O. Karali, O.E. Okman, T. Aytac, "Adaptive image enhancement based on clustering of wavelet coefficients for infrared sea surveillance systems," Infrared Physics & Technology, vol. 54, pp.382-394, 2011.

[5] C. Zuo, Q. Chen, G. Gu, X. Sui, "Registration method for infrared images under conditions of fixed-pattern noise," Optics Communications, vol. 285, pp.2293-2302, 2012.

[6] S. Dawn, V. Saxena, B. Sharma, "Remote Sensing Image Registration Techniques: A Survey," in 4th International Conference, ICISP, Trois-Rivières, QC, Canada, 2010. pp. 103-112.

[7] K. Rohr, W.J. Godinez, N. Harder, S. Yang, I. H. Kim, S. Wörz, R. Eils, "Tracking and Registration for Multidimensional Biomedical Image Analysis," in Progress in Industrial Mathematics at Ecmi 2008, vol. 15, 2010, pp.229-234.

[8] H.-X. Liu, W.-J. Yang, M. Huang, B. Wu, Y.-F. Xu, "Detection and Control Algorithm of Multi-Color Printing Registration Based on Computer Vision," in Image and Signal Processing, 2009. CISP '09. 2nd International Congress on. 2009. pp. 1-4.

[9] M.T. Hossain, G. Lv, S.W. Teng, G. Lu, M. Lackmann, "Improved Symmetric-SIFT for Multi-modal Image Registration," in Digital Image Computing Techniques and Applications (DICTA), IEEE International Conference on, Noosa, QLD, 2011, pp. 197-202.

[10] G. Hong, Y. Zhang, "Combination of feature-based and area-based image registration technique for high resolution remote sensing image," in Geoscience and Remote Sensing Symposium, IGARSS ,IEEE International, Barcelona, 2007, pp. 377-380.

[11] J. Kong, X.Y. Tang, M. Jiang, "Object location technique for moving target based on multi-scale feature extraction," Journal of Infrared and Millimeter Waves, vol. 30, pp. 21-26, 2011.

[12] Y.W. Wang, J. Lin, H. Liang, H. Yu, "X-ray image SIFT feature point detecting based on manifold learning," Advances in Information Sciences and Service Sciences, vol. 4, pp. 364-373, 2012.

[13] D.G. Lowe, "Object recognition from local scale-invariant features," in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 2, Kerkyra, 1999. pp. 1150-1157.

[14] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, pp. 91-110, 2004

[15] X.P. Wu, "Research on the palm print recognition algorithm research based on improved geometric invariance principle," Advances in Information Sciences and Service Sciences, vol. 4, pp. 97-103, 2012.