Image Registration of Infrared and Visible Image Based on Geometric Constraint

Juan Wang[1,2], Kaiwen Cheng[1], Cong Ke[1], Min Liu[1,*], Linkang Cai[1] and Hao Shi[1]

[1]Hubei Key Laboratory for High-Efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan 430068, People's Republic of China
[2]Post-doctoral Research Workstation, Wuhan Huaan Science and Technology Co., Ltd., Wuhan 430068, People's Republic of China

E-mail: happywj@hbut.edu.cn

Abstract. With the wide application of infrared sensors and visible sensors, the registration of two heterogeneous images has caused widespread concern. Multi-angle comprehensive analysis of multi-sensor imaging systems can obtain more accurate image information. However, the main challenge of image registration is the detection of feature points and the acquisition of the correct feature point matching pairs. A Geometric constraint method is proposed to obtain the correct feature matching pairs. Three feature point matching pairs are randomly selected in the two images, and the ratio of the corresponding line segment lengths is obtained according to the coordinates of the three feature points to find the most three similar feature point matching pairs. Compared with the RANSAC algorithm, the method can obtain the correct feature point matching pair better. The algorithm runs on the DSP platform for about 30s, the error is 1 pixel, and the correct rate reaches 80%.

1. Introduction

Image registration originated in the US military project in the 1970s. It is mainly used in the aircraft navigation assistance system, guidance system and photoelectric pod system, greatly improving the missile hit rate. At present, image registration technology has been widely used in 3D reconstruction [1], satellite remote sensing, medical image analysis, security monitoring and other industries. Image registration of infrared images and visible images belongs to heterogeneous image registration, that is, image registration of different sensors for imaging the same scene. Because of different imaging mechanisms, different resolutions, and different physical imaging conditions, imaging in two different modes has linear transformations such as rotation, translation, and scaling [2]. Image registration is to find an optimal geometric transformation to achieve geometric alignment of the two images.

Image registration algorithms can be roughly classified into the following categories: gray information-based method, image registration based on transform domain, feature-based image registration, and manual calibration registration algorithm [3]. The gray information-based method is used in the early registration field. The algorithm consists in selecting a similarity measure function and searching according to a certain path to find the geometric registration parameters when the maximum similarity is obtained [4]. Transform domain registration method is mainly to transform two-dimensional images from spatial domain to frequency domain. Common transform domain methods include wavelet transform, fast Fourier transform, discrete cosine transform, and so on [5]. Using the mathematical law of spatial domain transformation to the frequency domain, the registration

image can be subjected to frequency domain transformation, and then the phase and frequency mathematical relationship between the images to be registered is obtained according to the transformed result, and then inversely transformed into the spatial domain to become an image. The translation, rotation, and scaling operations result in two image geometric transformation model parameters [6]. Image registration method based on transform domain since the fast Fourier transform algorithm is quite mature, it has certain advantages in calculation. However, the disadvantage is that the image registration method based on the transform domain is also insufficient for complex geometric transformations [7].

Image features are the hotspot research direction in the field of image processing. The feature-based registration algorithm utilizes the geometric positional relationship of the image features to be registered to obtain geometric transformation parameters [8]. Image feature selection is related to the specific application scenario, and often determines the final effect of the entire registration algorithm. Selecting appropriate image features is necessary for image feature-based registration algorithms. Excellent features not only raise the registration effect but also improve the efficiency of the registration algorithm. Common image features generally have a few features, line features and contour features [9]. In addition to the above automatic registration algorithm, there is also an image registration method based on manual calibration. The image registration method based on manual calibration needs to artificially treat the same points in the registration image for coordinate calibration, and then combine the corresponding points in the two images into a pair of matching pairs. This manual calibration registration method [10] is simple to calculate and easy to operate. It is widely used in many low-end industrial binocular vision systems.

In this paper, a feature-based registration algorithm is used. The SURF algorithm [11] is used to extract visible image and infrared image feature points, and three feature points are respectively extracted to form a triangle template. The feature point matching uses the template matching constraint to eliminate the wrong match. The correct three pairs of feature point matching pairs are obtained to solve the affine transformation model parameters between the images with registration.

2. Preliminaries

2.1. SURF algorithm

2.1.1. Hessian matrix.

The Hessian matrix is mainly used to detect feature points in an image. The Hessian matrix is mathematically defined as a square matrix of second-order partial derivatives of a multivariate function. Its formula is as follows

$$H(P,\delta) = \begin{vmatrix} L_{xx}(P,\delta) & L_{xy}(P,\delta) \\ L_{xy}(P,\delta) & L_{xy}(P,\delta) \end{vmatrix} \tag{1}$$

Where P is a pixel on the two-dimensional image, and $L_{xx}(P,\delta)$ represents a Gaussian second-order partial derivative of the image in the x direction at the pixel point P, and the rest are similar. The parameter δ represents the variance of the two-dimensional Gaussian function, which is called the scale in the field of image processing. The final calculation is the determinant of the Hessian matrix. Each pixel of the image has its own Hessian determinant value. The Hessian value of all pixels constitutes the Hessian matrix of the image. When the determinant obtains the region maximum value, the pixel point P considered as an image feature point.

The SURF algorithm uses a box filter to approximate the Gaussian second-order derivative filter. This method can not only solve the signal aliasing phenomenon of multiple Gaussian convolution but also use the integral image to improve the calculation speed of the Hessian matrix. The Hessian determinant using the approximated box filter can be approximated by the following formula:

$$H(P,\delta) = D_{xx} * D_{yy} - (0.9D_{xy})^2 \tag{2}$$

Where $D_{xx}D_{xy}D_{yy}$ represents the second-order partial derivative of the image in the x direction, the xy direction and the y direction after the pixel point P is approximated, and a weighting coefficient of 0.9 is multiplied by D_{xy} in order to balance the error.

2.1.2. Build scale space. The SURF algorithm forms the image order according to the size of the box filter. Different sizes of the three box filters are obtained to obtain Hessian matrix diagrams of different sizes. The image is calculated by four kinds of three kinds of box filters of different directions ($D_{xx}D_{xy}D_{yy}$ three types of templates) to obtain four Hessian matrix diagrams and form a first-order image. Due to the different sizes of the box filters, the scale of the filtered images is different, which constitutes the three-dimensional scale space of the SURF.

2.1.3. Feature point location.
Non-maximum suppression is performed on the region of the Hessian matrix in the three-dimensional scale space. Each pixel is not only compared with the surrounding 8 pixels at the same scale, but also with the upper and lower Hessian matrix of the same order 19 pixels are compared, and the feature point is recognized only when the Hessian value of the pixel is greater than the other 26 Hessian values in the 3*3*3 region. In order to obtain the feature points of sub-pixel point precision, a second-order Taylor expansion of the scale space function is needed to fit the feature point position [12]. At this point, the position and scale information (x, y, δ) of the feature points can be known.

2.1.4. Characteristic point main direction analysis.
In order to make the feature points have rotational invariance, in addition to the above position and scale information, the feature points also need to have direction information. In the field of image processing, the direction θ of a pixel can be calculated from its horizontal x direction and the first derivative of the vertical y direction. The definition of $θ$ is as follows:

$$θ = \frac{dy}{dx} \qquad\qquad （3）$$

The SURF algorithm is used to determine the direction of the feature points [0]. It is necessary to calculate the Harr wavelet response of the horizontal direction and the vertical direction of all the pixels in the area by using the position of the feature point as the center and the radius of the image area of 6δ. A sector with a central angle of π/3 is used to wrap around the 6δ circular region around the feature point, and the sum of the Harr wavelet responses of the x and y directions of all the pixels in the sector is calculated to obtain the x direction and y. The total response in both directions of the direction constitutes a vector, and the modulus length of the vector is obtained. Record all the vectors obtained by rotating the fan one round, then take the largest of them. This vector can be used as the direction of the feature points.

2.1.5. Feature point vector descriptor construction.
The descriptor [13] vector of the feature points allows the visual features of the SURF feature points to be specifically quantized as "digital features". With the feature point description [0] sub-vector, it is possible to compare the "similarities and differences" with different features to determine whether the feature is the same feature. When constructing the feature point to describe the sub-vector, the feature point is taken as the origin. The feature point direction obtained in the above section is the positive direction of the y-axis, and the coordinate system is established. The square area of 20δ*20δ centered on the coordinate origin is selected, and the square is obtained. The area is evenly divided into 4*4 sub-areas. Each sub-region uses a Harr wavelet template with a size of 2δ*2δ to obtain the total response of the region in the x-direction and the y-direction. The response of each pixel in the region is obtained by weighting the Gaussian distribution centered on the feature points. In addition to the total response dx, dy in the horizontal and vertical directions of each region, the absolute values $|dx|, |dy|$ are also calculated. So each region has 4 quantities $dx, dy, |dx|, |dy|$ to characterize the region. Since the feature points have 16 sub-regions, the SURF algorithm has a total of 64 numbers to characterize the feature points [0], which constitutes a 64-dimensional feature description sub-vector of the SURF feature.

2.2. RANSAC algorithm eliminates mismatch matching

The RANSAC algorithm [14] eliminates mismatched feature point matching pairs. This algorithm for rejecting mismatched pairs mainly uses known data to estimate unknown mathematical model parameters. This method can eliminate obvious erroneous data. Firstly, a set of data is selected from the data points from the feature point rough matching to estimate the mathematical model parameters. After obtaining the mathematical model parameters, the specific model is obtained. Then use this model to verify other data in the dataset. If the verified data conforms to the model, then the confidence of the model is increased by one. After all the data is verified, the total confidence of the model is obtained. Then we use another set of data to get the model parameters, then find the overall confidence of the model. Repeat this step until you have traversed all the data in the data set. According to this iterative method, the confidence of all possible mathematical models can be obtained, and finally the mathematical model parameters with the highest confidence are selected as the final model.

3. Template matching

3.1 Template matching eliminates mismatch matching

In view of the shortcomings of the RANSAC algorithm, this section proposes a feature point matching pair extraction method based on template matching. Imaging of the imaging sensor after camera calibration can be considered as undistorted, eliminating the nonlinear transformation that may be caused by image distortion, which means that there are a large part of similar information in the two images of the same scene, and this part is only a linear geometric transformation relationship in the image. Then the geometrical positions of the image feature points extracted in the similar parts of the two images should also be similar. That is, if three feature points are extracted from the visible image to form a triangle template, the corresponding feature points in the infrared image can also be constructed. A similar triangle template with only translation, rotation, and scaling relationships between the two templates. If there are three pairs and above correct feature point matching pairs for the two images, then the two triangles formed by the pair of feature point matching pairs are also necessarily the most similar among all matching pairs. The constrained condition of the feature point matching pair plus the triangle similar template can be used to eliminate the mismatching pair and the correct three pairs of feature point matching pairs can be obtained, and the affine transformation model parameters between the registered images can be solved.

Whether the two template triangles are similar can be based on the ratio of the length of the line segments. How to find the optimal similarity in the algorithm, we can find a similar quantitative expression to facilitate programming. Suppose the side lengths of the first triangle are n1, n2, and n3, respectively. The side length of the second triangle is m1, m2, and m3, and the ratio of each side is

$$k_1 = \frac{n1}{m1}, \quad k_2 = \frac{n2}{m2}, \quad k_3 = \frac{n3}{m3} \qquad （4）$$

The length of each line segment is the length of the Euclidean distance [0] in the pixel coordinate system. In theory, the ratio of the corresponding triangle segments should be equal, but the triangles in the program cannot be judged to be similar under the condition of complete equality. So when designing the algorithm, you can set a threshold to ensure that the images are similar. The threshold is set as follows:

$$abs\left(1 - \frac{k_1}{k_2}\right) < 0.02 \&\& abs\left(1 - \frac{k_2}{k_3}\right) < 0.0 \qquad （5）$$

When both are less than the threshold 0.02, the above two conditions are satisfied to prove that the ratios of the three line segments are close to the same. If it is greater than the threshold, the triangles are judged to be dissimilar. Under similar conditions, set the dust value to select the most similar triangle. The definition of dist is as follows:

$$\text{dist} = abs\left(1 - \frac{k_1}{k_2}\right) + abs(1 - \frac{k_2}{k_3}) \qquad （6）$$

The smaller the dist. value, the more similar the two triangles are. In the feature point rough matching pair, three pairs of feature point matching pairs are arbitrarily selected, and the three pairs of feature point matching pairs will constitute two triangles. According to the coordinates [0] of the three feature points, the length of each line segment can be obtained, and then the ratio of the length of the corresponding line segment of the triangle is calculated, and three ratios are obtained. Finally, the three similar pairs of feature point matching pairs are found according to the above two conditions. The algorithm flow chart is shown in figure 1.
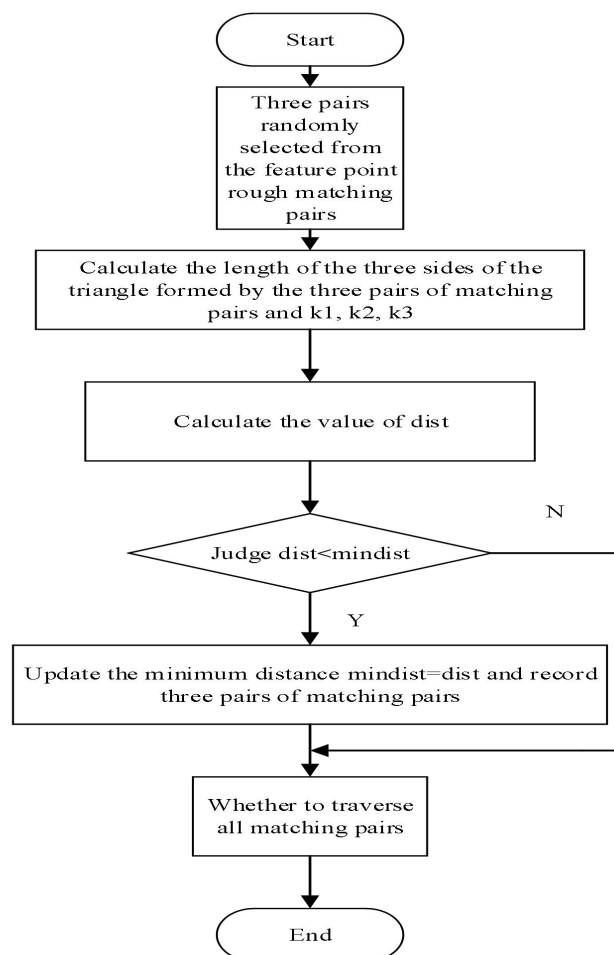

Figure 1. Geometric Constraint algorithm flow chart.

3.2 Edge coincidence degree-based registration evaluation method
This section mainly proposes a registration evaluation method based on edge coincidence degree. Firstly, the Canny algorithm[14] is used to extract the edge image of the registration image separately, and the edge of the image can be approximated to calculate the edge coincidence degree. The edges obtained by the Canny edge extraction of the visible light image or the infrared image are binary images.

Each pixel has an intensity value of 0 or 255, and the pixel with an intensity of 255 is an edge pixel. For the two edge images of the registration image, select the image with fewer edges in the two edge images as the reference image, traverse the pixel points in the reference image as edges, and then observe whether the corresponding pixel in the registration image is also the edge pixel. If both are edge pixels, the registration confidence is increased by one. By traversing all the edge pixels in the reference image as described above, the total confidence

of the registration can be obtained. The objective confidence rate is then obtained by dividing the total confidence by the number of edge pixels in the reference image. The steps of the registration evaluation method based on the edge coincidence degree are shown in figure2.
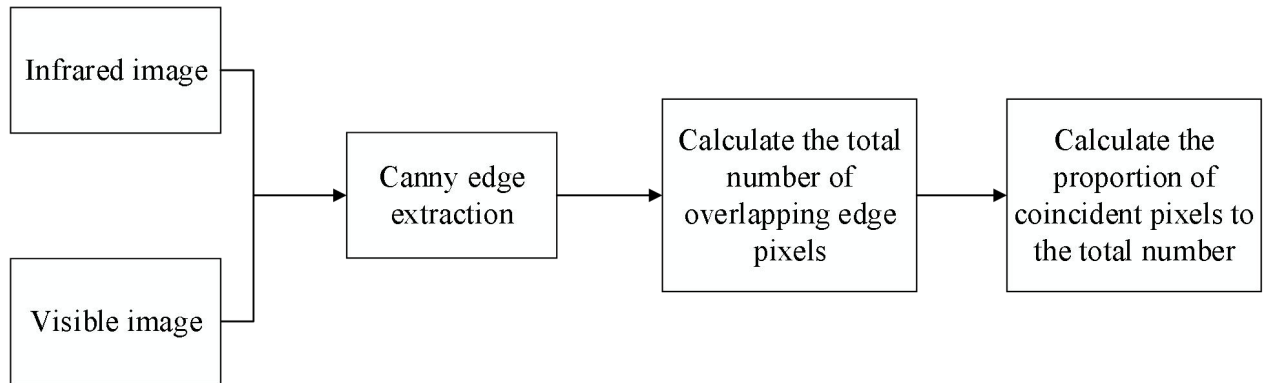

Figure 2. Registration evaluation method based on edge coincidence.

4. Experiment analysis

4.1 Feature point rough matching pair extraction

The feature points extracted by the SURF algorithm are then used to describe the nearest Euclidean distance of the sub-vectors according to the image feature points. The coarse matching of the feature points is as shown in figure 3.
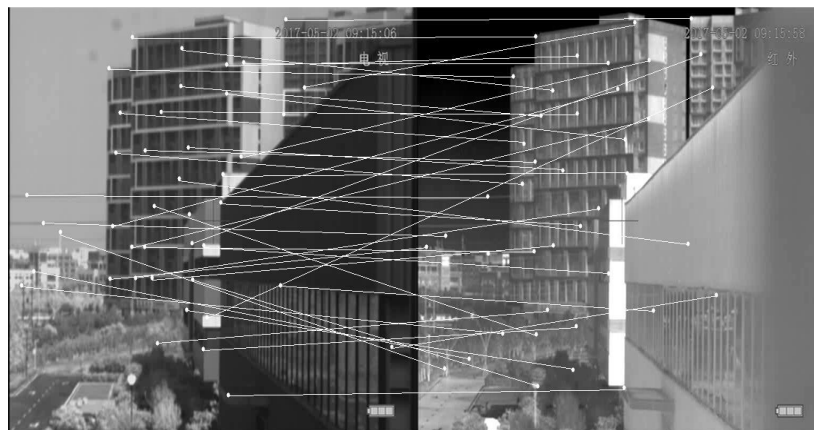

Figure 3. Pairs of visible and infrared image feature points.

It can be seen from Fig. 3 that there are about 40 pairs of feature point matching pairs formed, and the matching of the feature points formed by the infrared image and the visible image through the preliminary screening is mostly mismatched. The main reason for the large number of mismatching pairs here is the difference in the imaging mechanism of the imaging mechanism of the two heterogeneous image sensors. The visible light imaging sensor is imaged according to the brightness of the reflected light of the object, and the intensity values of the pixels in the infrared image sensor are Heat related to object radiation.

4.2 Template matching method for extracting feature point matching pairs

From the feature point rough matching pair, the template matching algorithm and the RANSAC algorithm are used to extract the correct feature point matching pairs respectively. The experimental results are shown in figure 4.

It can be seen from Fig. 4 that the feature point matching pair selected by the template matching method is obviously superior to the feature point matching pair selected by the RANSAC algorithm. It can be seen from the figure that there are 2 pairs of matching pairs in the feature point matching pair selected by the RANSAC algorithm belonging to the mismatch matching pair. It can be found that as long as the feature point matching pair has three pairs and above correct feature point matching pairs, the template matching method can be used to find the correct feature point matching pair.
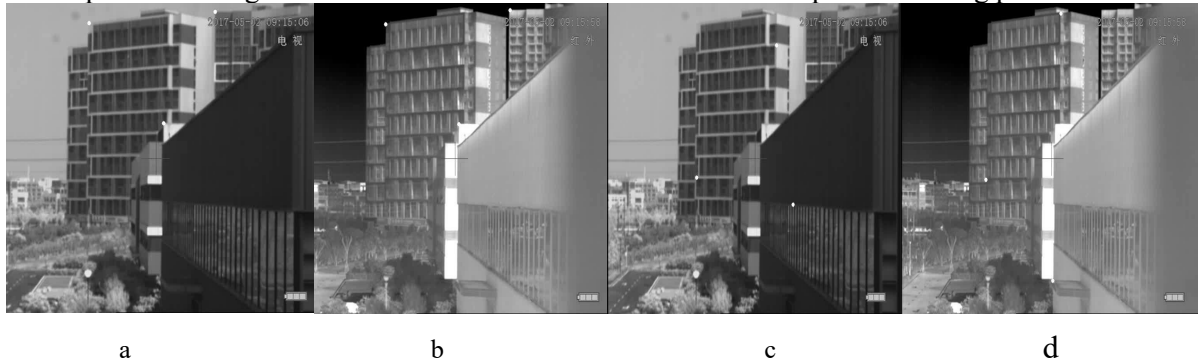

      a   b   c   d

Figure 4. (a,b) shows the feature points of the visible image and the infrared image obtained by the Geometric Constraint method, respectively. (c,d) respectively represent feature points of the visible image and the infrared image obtained by the RANSAC algorithm.

4.3. Registration evaluation results based on edge coincidence

The registered infrared image edge and visible image edge obtained by Canny edge extraction algorithm are shown in figure 5.
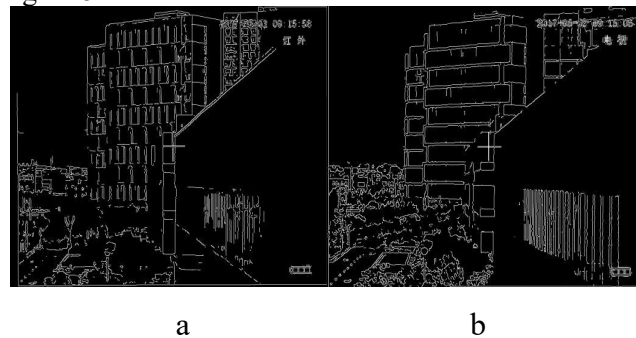

      a   b

Figure 5. (a) shows the edge of the infrared image (b) shows the edge of the visible image.

It can be seen from figure 5 that the two edges have both a similar portion and a plurality of different portions. For different portions of the edges, matching with other edge pixels of another image may increase the registration confidence without registration. This section mainly demonstrates from the two levels of translation and rotation that the registration rate obtained by this method has a maximum value attribute. For translation, consider the translation in the horizontal and vertical directions, translate one of the two edge images, and then recalculate the coincidence of the two images after translation. For rotation, the image center area is selected for edge coincidence calculation. The results of experimental are shown in Table 1 and 2.

Table 1. Ten-pixel translational registration rate.

vertical \ horizontal	0	1	2	3	4	5	6	7	8	9	10 (/pixel)
0	0.2919	0.2822	0.2841	0.2875	0.2823	0.2783	0.2761	0.2765	0.2690	0.2676	0.2662
1	0.2916	0.2827	0.2837	0.2883	0.2790	0.2734	0.2759	0.2748	0.2692	0.2663	0.2665
2	0.2897	0.2803	0.2831	0.2857	0.2795	0.2735	0.2746	0.2744	0.2656	0.261	0.2617
3	0.2873	0.2781	0.2782	0.2839	0.2748	0.2703	0.2705	0.2689	0.2659	0.2610	0.2605
4	0.2853	0.2778	0.2754	0.2802	0.2736	0.2684	0.2697	0.2669	0.2591	0.2581	0.2578
5	0.2795	0.2701	0.2728	0.2770	0.2709	0.2620	0.2641	0.2646	0.2561	0.2541	0.2560
6	0.2762	0.2673	0.2684	0.2737	0.2641	0.2621	0.2612	0.2600	0.2544	0.2526	0.2514
7	0.2719	0.2652	0.2661	0.2686	0.2635	0.2566	0.2563	0.2568	0.2515	0.2480	0.2505
8	0.2722	0.2615	0.2606	0.2665	0.2568	0.2549	0.2564	0.2541	0.2460	0.2467	0.2566
9	0.2646	0.2579	0.2551	0.2633	0.2553	0.2488	0.2517	0.2496	0.2470	0.2431	0.2447
10	0.2618	0.2538	0.2545	0.2568	0.2499	0.2481	0.2476	0.2468	0.2429	0.2384	0.2406

Table 2. Rotating registration rate within 30 degrees.

angle	0	1	2	3	4	5	6	7	8	9	10
rate	0.331	0.320	0.304	0.285	0.276	0.276	0.258	0.249	0.235	0.231	0.225
angle	11	12	13	14	15	16	17	18	19	20	21
rate	0.226	0.217	0.223	0.224	0.230	0.231	0.233	0.225	0.240	0.239	0.232
angle	22	23	24	25	26	27	28	29	30		
rate	0.230	0.237	0.236	0.236	0.237	0.235	0.234	0.224	0.231		

Table 1 shows that horizontal and vertical translation amount is 0, the registration ratio is at most 0.2919. According to the experimental data, it can be known that the registration rate obtained by this evaluation method can maintain the maximum value attribute for the translation transformation. Table 2 when the rotation angle is 0, the registration ratio of the two registration images without the rotation plane is up to 0.331. This registration evaluation method can maintain the maximum value attribute for the rotation transformation. The experimental data can prove that the registration evaluation method based on edge coincidence degree is reliable. In practical applications, the registration of the registered images can be manually calibrated, and then the edge coincidence registration ratio of the two images is calculated, and the registration rate is used as a priori condition to detect the final registration effect.

4.4. DSP platform algorithm implementation

TI's Da Vinci series TMS320DM6437 processor is a 32-bit fixed-point multimedia DSP processor with 128MB of Nand Flash ROM and 256MB of DDR2 RAM, clocked at up to 600MHz, peak processing capacity of 5600MIPS, low power consumption. Widely used in high-performance, low-cost vision development systems. Figure 6 is a graphical representation of the experimental results of the entire registration algorithm running on the DSP platform.

a   b

Figure 6. DSP platform test results.

The experimental results were taken by the mobile phone. As can be seen from the figure, the gray value is high, the lack of detail is the infrared image, the detail is fine, and the lower gray level is the visible image. Visible images and infrared images basically meet the registration requirements. The running time of the whole algorithm is about 30s, the registration error is 1 pixel, and the correct rate is about 80%.

5. Conclusions

This paper mainly presents a template matching method for infrared image and visible image registration to obtain the correct feature point matching pair. This method uses the triangle template by matching pairs of coarse feature points to obtain the correct feature point matching pairs based on their similarity. At the same time, this algorithm is also implemented on the DSP platform, but it can be found that the execution time of the algorithm on the DSP platform is relatively long, and the FPGA platform or GPU can be used later to reduce the processing time of the algorithm. The entire algorithm process also has a part that is worth optimizing. You can consider the optimization of the calculation to reduce the running time of the program and improve the performance of the program. This type of algorithm uses a fixed threshold, which means that this is not an adaptive algorithm. However, with the wide application of deep learning in the field of image, we can consider the application of deep learning algorithm [15] in the field of image registration.

References

[1] D. G. Lowe, "Local feature view clustering for 3D object recognition," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I.

[2] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration,"IEEE Transactions on Image Processing, vol. 5, no. 8, pp. 1266-1271, Aug. 1996.

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

[4] H Bay, A Ess, T Tuytelaars, et al. Speeded-Up Robust Features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3):346.

[5] E. Rublee, V. Rabaud,K. Konolige and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," 2011 International Conference on Computer Vision, Barcelona, 2011, pp. 2564

[6] D G .Lowe Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91.

[7] S. Leutenegger, M. Chli and R. Y. Siegwart, "BRISK: Binary Robust invariant scalable keypoints,"2011 International Conference on Computer Vision, Barcelona, 2011, pp. 2548-2555.

[8] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615-1630, Oct. 2005.

[9] J MMorel, GYu. ASIFT: A New Framework for Fully Affine Invariant Image Comparison[J]. SIAM Journal on Imaging Sciences, 2009, 2(2):438-469.

[10] E. Rosten, R. Porter and T. Drummond, "Faster and Better: A Machine Learning Approach to Corner Detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 1, pp. 105-119, Jan. 2010.

[11] Fischler, Martin & C. Bolles, Robert. (1981). RandomSample Consensus: A Paradigm for Model Fitting with Applications To Image Analysis and Automated Cartography. Communications of the ACM. 24. 381-395. 10.1145/358669.358692.

[12] M. C. Lin and J. F. Canny, "A fast algorithm for incremental distance calculation," Proceedings. 1991 IEEE International Conference on Robotics and Automation, Sacramento, CA, USA, 1991, pp. 1008-1014 vol.2.

[13] Flusser, Jan. (2011). Moment Invariants in Image Analysis. Transactions on engineering, Computing and Technology. VII.

[14] Canny, J. A computational approach to edge-detection. IEEE Trans. Pattern Anal. Mach. Intell 1986, 8, 679–698.

[15] Shuxiu Wang, Sheng Lei and Faliang Chang, "Image registration based on neural network," 2008 International Conference on Information Technology and Applications in Biomedicine, Shenzhen, 2008, pp.