Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5598177, 9 pages
https://doi.org/10.1155/2021/5598177

Research Article

Rapid Multimodal Image Registration Based on the Local Edge Histogram

Dong Zhao

State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment, China Nuclear Power Engineering Co., Ltd., Shenzhen, Guangdong 518172, China

Correspondence should be addressed to Dong Zhao; zhao.dong@cgnpc.com.cn

Received 1 March 2021; Revised 11 May 2021; Accepted 24 May 2021; Published 2 June 2021

Academic Editor: Chris Goodrich

Due to significant differences in imaging mechanisms between multimodal images, registration methods have difficulty in achieving the ideal effect in terms of time consumption and matching precision. Therefore, this paper puts forward a rapid and robust method for multimodal image registration by exploiting local edge information. The method is based on the framework of SURF and can simultaneously achieve real time and accuracy. Due to the unpredictability of multimodal images' textures, the local edge descriptor is built based on the edge histogram of neighborhood around keypoints. Moreover, in order to increase the robustness of the whole algorithm and maintain the SURF's fast characteristic, saliency assessment of keypoints and the concept of self-similar factor are presented and introduced. Experimental results show that the proposed method achieves higher precision and consumes less time than other multimodality registration methods. In addition, the robustness and stability of the method are also demonstrated in the presence of image blurring, rotation, noise, and luminance variations.

1. Introduction

Image registration maps two images of the same scene and similar viewpoints to the same coordinate system by looking for a certain spatial geometric transformation so that the points with the same spatial position have the same coordinates. The technology is a basic problem in the field of image processing. Multimodal images are multisensor images acquired by sensors with different imaging mechanisms. With the rapid development of sensor imaging technology, the application of multisensors is increasingly extensive. The information of a single image sensor cannot meet the needs of applications. For example, in the field of military terminal guidance, the fusion of synthetic-aperture radar (SAR) images, infrared images, and visible images can better judge and recognize the targets [1, 2]. In medicine, by analyzing computed tomography (CT) images sensitive to bone tissue and magnetic resonance imaging (MRI) images sensitive to soft tissue, diseases can be diagnosed more comprehensively [3]. Different multimodal images acquired by different sensors show different expression forms of pixels and provide more complex information. Different gray values may appear in the same position of the same object between different multimodal images.

With the attention of an increasing number of scholars on multimodal image registration technology, many methods have been proposed and have achieved success in a certain range. In [4], Zhao et al. introduced a novel multimodality robust line segment descriptor, which uses extracted highly equivalent corners and line segments of multimodal images. In [5], Ye et al. developed a local descriptor for remote sensing image registration, which is constructed by using a histogram of oriented gradients (HOGs) and local self-similarity (LSS). In [6], based on gradient reversals, Chen and Tian proposed a symmetric-SIFT descriptor suitable for multimodal image registration. In [7], the hybrid visual features were employed for visible and infrared image registration. This algorithm extracted many lines from the edge points, matched these lines of different images to roughly estimate the transformation parameters, and then used the estimated parameters to map the keypoints of one image to another image to adjust the

transformation parameters. In [8], a multispectral corner detection operator was introduced, which can improve the corner extraction performance in multimodal images. The algorithm can effectively match the near-infrared images with the visible images. In [9], Ye et al. proposed exploiting phase congruency (PC) as a generalization of the gradient information and developed a robust descriptor named the histogram of orientated phase congruency (HOPC). In [10], a feature detector named MMPC-Lap was proposed by using the minimum moment of PC for feature detection with an automatic scale location technique. In [11], Li et al. proposed a novel feature matching algorithm called RIFT which uses the maximum index map (MIM) instead of the gradient for feature description.

In some specific scenes, besides accuracy, time is one of the key factors for matching. However, most multimodal image registration algorithms pay much attention to the matching accuracy but ignore the time consumption of the algorithms. In fact, as long as enough correct matches are found, the accurate transformation matrix can be obtained, which is the purpose of matching. In monomodal image registration methods, the SURF operator [12] shows good performance in both computational speed and performances of repeatability, distinctiveness, and robustness. In order to reduce the computation time and simultaneously ensure the performances, SURF employs the integral images and simplified box filters to calculate the approximate Hessian matrix and Haar wavelet response. In multimodal images, although the gray values of corresponding points are susceptible to change, image boundaries tend to be preserved, which suggests that many of the same keypoints, at least those that do not depend on the texture, will tend to reappear in different images [13]. Therefore, utilizing the edge information around keypoints, we can build an effective descriptor.

Aiming at the problems of time consumption and precision in multimodal image registration, we propose a rapid and robust SURF-based multimodality registration method. The proposed method uses integral images and optimized box filters to speed up and describe the edge information of images. In the method, keypoints are obtained by using the SURF detector, and then the keypoints with lower significance are eliminated. The dominant orientation of each keypoint is identified by the structure degrees of the neighborhood. Then, a local region edge descriptor is constructed inspired by the normative edge histogram of MPEG-7 [14]. At last, in order to resolve ambiguities caused by the keypoints with similar local neighborhoods, the keypoints with high self-similarity factors are deleted. Experimental results show that the method has better performance in time and precision.

The rest of the paper is organized as follows. Section 2 details the proposed method. The experimental results are presented in Section 3. Finally, Section 4 concludes this paper.

2. Proposed Method

In this section, we present our method in detail. The process of the proposed algorithm is as follows:

Step 1: given input images, use the SURF detector to extract their keypoints

Step 2: calculate each keypoint's saliency index, and select the keypoints with high saliency index

Step 3: identify the dominant orientation of each keypoint

Step 4: construct the local edge descriptor for each keypoint

Step 5: compute the self-similar factor of each keypoint, and delete the keypoints with high self-similar factor

Finally, the remained keypoints are used to match two images. The transformation model can be found through the RANSAC algorithm. Step 1 is not the idea of this paper, and this step can refer to the SURF method [12].

2.1. Saliency Assessment of Keypoints. SURF can detect many stable keypoints. Nevertheless, most of them do not contribute to the registration result and, instead, may affect the correct matches. Redundant keypoints not only reduce the matching precision but also greatly consume the matching time. To remove the redundant keypoints, we assessed the saliency of all the extracted keypoints.

The divergence of the textures between multimodal images is large, but the structure is relatively stable. Moreover, the subsequent operations of our method are based on neighborhood structure information of keypoints. Rich structure information can improve not only the repeatability of dominant orientation but also the robustness and distinctiveness of descriptors. Therefore, the structure information of keypoints' neighborhood is counted as the saliency index of the keypoints.

Structure in images exists in the form of edges. In [15], a method of computing the saliency index based on edge pixels is proposed. The saliency index is calculated within a local window based on the edge density and the distribution evenness of the edge pixels. The specific calculation is

$$I = d_e \cdot d_s, \tag{1}$$

where

$$d_e = \frac{n_e}{n_w}, \tag{2}$$

$$d_s = \frac{\min\ (n_i)}{\text{mean}\ (n_i)}, \quad i = 1, 2, 3, 4. \tag{3}$$

In (1), d_e denotes the edge density, and d_s denotes the distribution evenness. In (2), n_e and n_w are the number of edge pixels and all pixels in the local circle window, respectively. In (3), n_i is the number of edge pixels in each of the four quadrants of the local window, and d_s measures how evenly the edge pixels are distributed in the four quadrants. The evaluation criteria can effectively assess the keypoints' saliency. However, the calculation of the saliency index involves the number of edge pixels, which is vulnerable to the impact of the edge extraction algorithm. If the edge extracted by an edge extraction algorithm is thick, it may

improve the saliency indexes of keypoints, or vice versa. For simplicity, combining with the edge's idea, we propose using the local similar gradient to indicate the saliency index. In [16], the similar gradient is defined as the vector formed by the Haar wavelet responses in the horizontal and vertical directions. Combining with the integral images, the Haar wavelet responses can be calculated quickly. The saliency index is defined as follows:

$$\text{saliency} = \frac{1}{N(r)} \sum_{(i,j)\in L(r)} \sqrt{h_x^2(i,j) + h_y^2(i,j)}, \tag{4}$$

where

$$L(r) = \left\{(x,y) \mid (x - x_0) + (y - y_0)^2 \le r^2\right\}. \tag{5}$$

In (4), $N(r)$ is the pixel number in the circle of radius r, namely, the size of $L(r)$. $h_x(i,j)$ and $h_y(i,j)$ denote the Haar wavelet responses in the horizontal and vertical directions, respectively. In (5), (x_0, y_0) denotes the keypoint's coordinate. r is the radius of the local region. Here, we set the radius r as $10s$, with s the scale at which the keypoint was located to make all pixels of the region within the square region of descriptors. The image gradient reflects the gray variation in a certain direction, and its amplitude is the basis of edge detection algorithms. By computing the local gradient of keypoints' neighborhoods, the richness of keypoints' neighborhood structure information can be effectively expressed.

Thus, each keypoint has its own saliency index. We use the conventional method (strongest responses) to select a specific number of keypoints from each image. However, the numbers of extracted keypoints in various modal images are quite different, e.g., the number of keypoints extracted from visible light images tends to be larger than that of other modal images. Here, we determine the number we want to choose according to the total number of extracted keypoints. Obviously, removing the keypoints with lower significance will affect the repeatability of the features. However, if the saliency indexes of the same position keypoints in different modal images are different, which indicates their surrounding structure information is different, they are difficult to be matched successfully. We have verified through experiments that removing redundant keypoints can effectively improve the matching rate, which is the ratio of correct matches to total matches. In order to ensure the repeatability of keypoints and remove enough redundant points, we recommend that at least half of the keypoints should be retained in each image.

2.2. Orientation Assignment. In [17], Lee et al. proposed that edginess, instead of gradient amplitude, can show better performance in multimodal images. On the basis of this theory, we propose the concept of structural degree. The structural degrees of pixels can be calculated by structure tensor Q_σ, which is

$$Q_\sigma = \begin{bmatrix} G_\sigma * I_x^2 & G_\sigma * I_x I_y \\ G_\sigma * I_x I_y & G_\sigma * I_y^2 \end{bmatrix}, \tag{6}$$

where $*$ denotes the convolution operation, G_σ is the partial derivative on the σ-scale of the two-dimensional Gaussian function, and I_x and I_y denote the Haar wavelet responses in the horizontal and vertical directions, respectively. Through the matrix, two feature vectors w_1 and w_2 can be obtained, which, respectively, represent the directions of the maximum and minimum change in the gray level in the position, and the corresponding eigenvalues are μ_1 and μ_2 ($\mu_1 \ge \mu_2 \ge 0$). The structural degree C is defined as follows:

$$C = \begin{cases} \dfrac{(\mu_1 - \mu_2)}{\mu_1}, & \mu_1 \ne 0, \\ 0, & \mu_1 = 0. \end{cases} \tag{7}$$

For each keypoint, we first calculate the structural degree of each point within a circular neighborhood of radius $6s$ around the keypoint. Then, the structural degrees are weighted by Gaussian ($\sigma = 2s$) centered at the keypoints. Next, a fan-shaped region of $\pi/3$ angle is used to rotate in a specific step $\pi/18$ along the counterclockwise direction, and the sum of the structural degrees of all the smoothed points in the fan-shaped region is calculated in turn. We select the sector region whose structural degree summation is max as the orientation region, and its angle bisector is assigned as the dominant orientation of the keypoints.

2.3. Local Edge Descriptor. In multimodal images, because of different imaging principles, the texture and color are unreliable. To some extent, the structures of multimodal images can maintain stability. The structure plays an important role in the multimodality registration. Considering the characteristic, many scholars carry out research on multimodal image registration based on the edge features.

In the MPEG-7 standard [14], the edge histogram descriptor is efficiently utilized for image description. The edge histogram descriptor defines five edge types. They are four directional edges and a nondirectional edge. Four directional edges include vertical, horizontal, 45-degree, and 135-degree diagonal edges. In this paper, we adopt the same edge types to build the descriptor, but the building process is different.

The first step of extracting the descriptor is constructing a $20s * 20s$ (s is the scale of the SURF detector) local square image region centered around a keypoint and oriented along the dominant orientation. For each sample point of the local region, five types of Haar wavelet responses are computed (filter size $2s$). h_x, h_y, h_{45}, h_{135}, and h_{no} are, respectively, used to denote the Haar wavelet response in a certain direction. The five Haar wavelet filters are shown in Figure 1. The directions here are defined in relation to the selected keypoint orientation. Then, we compare the absolute values of the five Haar wavelet responses. If the maximum value among the five responses' absolute values is greater than a threshold as in (8), the sample point is considered to be on the corresponding edge, otherwise considered a point of the nondirectional edge.

$$\max\left\{h_x, h_y, h_{45}, h_{135}, h_{no}\right\} > \text{Th}_{\text{edge}}. \tag{8}$$

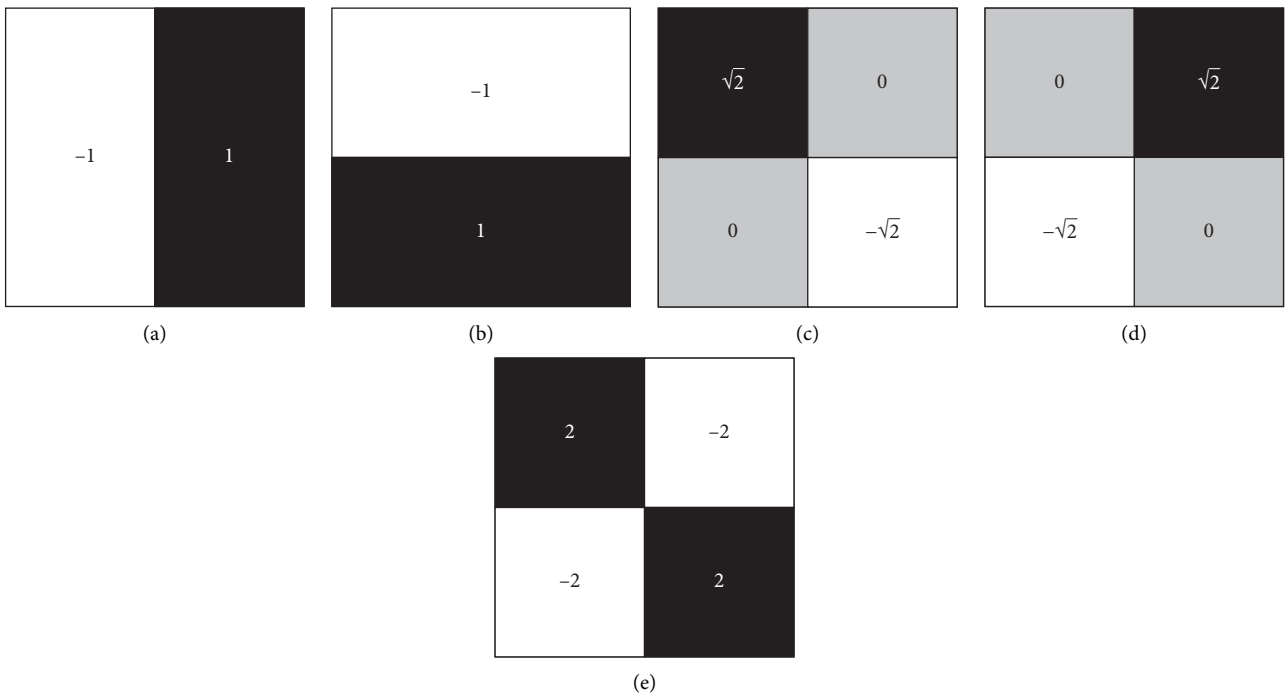


Figure 1: Five types of Haar wavelet filters to compute the responses. The numbers are the weights of the corresponding part. (a) Vertical filter. (b) Horizontal filter. (c) 45-degree filter. (d) 135-degree filter. (e) Nondirectional filter.

The construction process of the descriptor is depicted in Figure 2. In order to incorporate spatial information into the descriptor, the local region is split up regularly into smaller 4*4 square subregions. For each subregion, five histogram bins are defined to correspond to five different edges. We count the total number of edges for each edge type in each subregion. It is easy to figure out that the sum of the five histogram bins is a fixed value 25. The final descriptor is constructed by concatenating the histograms of all the subregions. This results in a descriptor vector of length 80. When the texture histogram is used as the local descriptor, the uniform strategy is more suitable for pixel weighting. That is to say, the weighting step can be omitted in our method. Finally, the descriptor is normalized to unit length. The new descriptor is denoted as the local edge descriptor.

2.4. Self-Similar Factor. At present, each keypoint possesses its own descriptor. If neighborhood structure information of two keypoints in the same image is similar, it is likely that their descriptors are also similar. When matched with keypoints of other images, the two keypoints may be matched to the same keypoint. If, in the same positions of another image, another two similar keypoints are extracted at the same time, the matches between them will be chaos. The more similar keypoints there are, the worse the situation is. Here, we introduce the concept of the self-similarity factor to maintain the distinctiveness of each descriptor.

For each keypoint, the similar factors of it with all the others in the same image are calculated. Considering the different contrast of images taken by multisensors, we adopt the cosine method to measure similar factors, which is defined as follows:

$$\mathrm{sim}(x, y) = \cos(x, y) = \frac{(x, y)}{\|x\| \cdot \|y\|} = \frac{\sum_{i=0}^{n-1} x_i \cdot y_i}{\sqrt{\sum_{i=0}^{n-1} x_i^2 \cdot \sum_{i=0}^{n-1} y_i^2}}. \tag{9}$$

In (9), sim(x, y) denotes similar factors, and x and y are two feature vectors in the same image. Obviously, a similar factor is in the range [−1, 1]. A larger value indicates that the vector angle is small and that the two vectors are more similar. If the value is 1, the two vectors are identical. After obtaining similar factors of the keypoint with other keypoints, the maximum similar factor is considered as the keypoint's self-similar factor. After that, we screen keypoints with high self-similar factors according to the method mentioned in the saliency assessment section.

The saliency assessment can remove redundant keypoints. The self-similar factor can filter the keypoints with similar descriptors and improve the distinctiveness of keypoints. They both increase the robustness of the whole algorithm and maintain the SURF's fast characteristics.

3. Experiments and Results

In this section, we conduct several experiments to verify the effectiveness of the proposed method. Considering the experiments involving the runtime, all algorithms are implemented with the same hardware development environment and platform. The development environment is an Intel Core 2 2.94 GHz CPU and 2 GB of memory. The operating system is 64-bit Windows 7. The development platforms are Visual Studio 2010 and OpenCV 2.3. The multimodal images that are used in experiments are from [13, 18, 19], and the

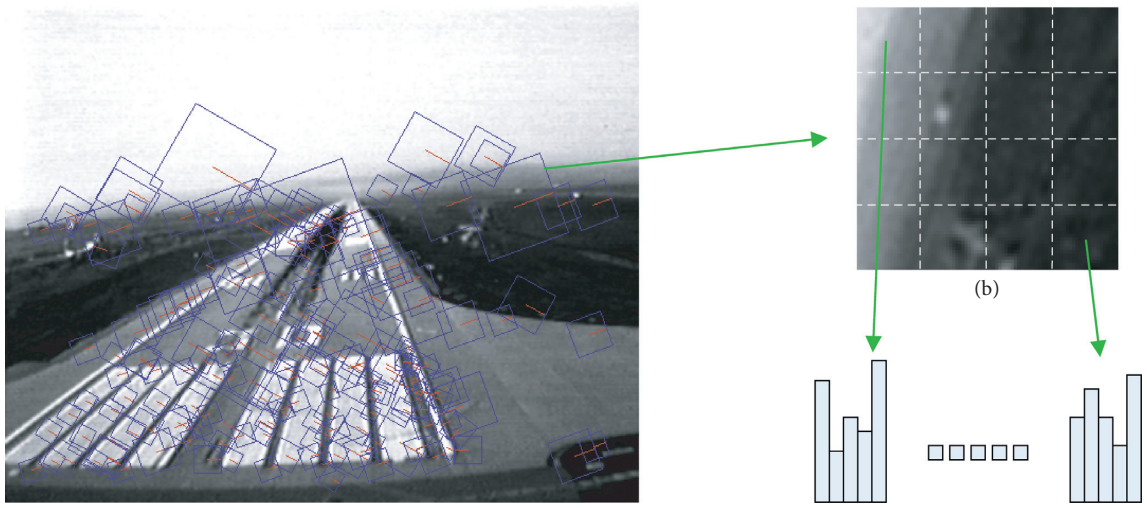


Figure 2: An illustration of constructing the local edge descriptor. (a) Local square regions around the keypoints detected by the SURF detector. (b) A normalized region with a Cartesian location grid after rotating along the dominant orientation. (c) The final local edge descriptor of the normalized region.

blurred, rotating, noise, and different luminance images of multimodal images are from [16].

Firstly, the proposed method is compared with the original SURF to verify whether the descriptor can effectively describe the structural information of the multimodal image. Ten pairs of multimodal images are tested. One pair is shown in Figure 3, in which the test images are the visible spectrum and IR images with different scales. Figure 3(a) shows the result of SURF, which generates no match. Our method generates 7 matches, and all of them are correct, as shown in Figure 3(b). The comparative results show that our method can be effectively used for multimodal image registration.

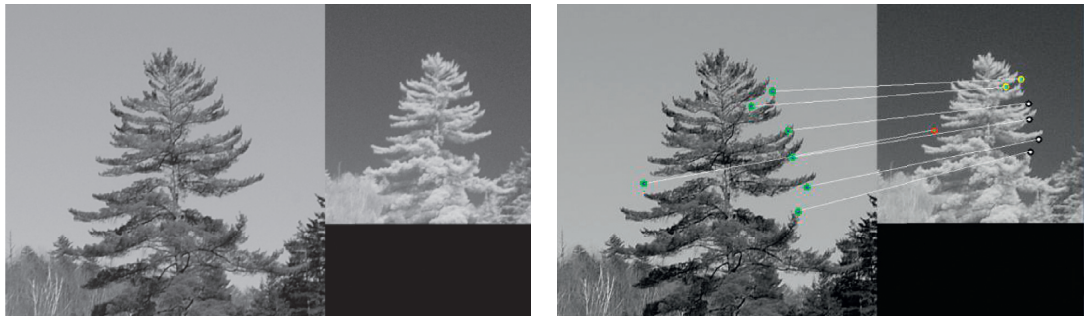
In the following experiments, we evaluate the performance of the proposed method from two dimensions: time and precision. The method in this paper is based on keypoints, so our method is compared with the same-type methods, RIFT [11], symmetric-SIFT [6], and the multispectral detector method [8], which adopts the multispectral Laplace detector. As mentioned above, if enough correct matches are found, the final transformation matrix can be obtained using the RANSAC algorithm. Therefore, we adopt the matching rate [20] and Correct@N [16] as the evaluation criteria. The matching rate is the ratio of correct matches to total matches, which can effectively reflect the efficiency of the algorithm. Correct@N is the number of correct matches of the first N in the ratio of the nearest descriptor to the next nearest descriptor. The two methods can evaluate the performance of the algorithm more comprehensively.

3.1. Time Evaluation. The time evaluation is a relative result [21], which only shows the tendency of time cost. The time is counted for complete processing, including keypoints' detection and matching. We take 8 pairs of multimodal images for experiments, as shown in Table 1, in which the runtimes are the time of finding the first 20 matches. Inheriting the merits of SURF, our method can dramatically reduce the computational time and is obviously faster than the other methods.

3.2. Precision Evaluation. In our method, the saliency assessment and calculation of self-similar factors can eliminate redundant keypoints, increase the number of correct matches, and improve the matching rate of experimental results. Since these two steps are not necessary for the registration process, we can obtain a simplified registration method by removing the two steps. We choose 4 pairs of multimodal images to compare our proposed method with the simplified method, and the same threshold is adopted in the experiments, as shown in Table 2. As the results show, the two steps of the saliency assessment and calculating self-similar factors can significantly improve the matching rate. In the streets, the matching rate is even improved by nearly 17%.

To evaluate the matching accuracy of algorithms, we use the Corrent@N method to compare our proposed method with the other three methods. Eight pairs of multimodal images are tested, and the Correct@N curves of two pairs are shown in Figure 4. Obviously, our method performs better than the other algorithms. In EO-IR-2, although the proposed method is not as effective as RIFT at the beginning, with the increase in N, the strength of the proposed method is gradually reflected.

Then, we take the transformed multimodal images of the tree branch for experiments to verify the robustness of the proposed method. The experimental images [16] are derived from a series of transformations of multimodal images of the tree branch, which is a visible spectrum and IR pair. The transformations are (1) 6 blurred images, convolving the visible image with different filters whose sizes are 3*3, 5*5, or 7*7 and sigmas are 10 or 20; (2) 6 rotating images, respectively, rotating the IR image 5, 10, 15, 20, 25, and 30 degrees; (3) 4 noise images, respectively, adding Gaussian



(a) (b)

Figure 3: Comparative results of registration between the visible spectrum and IR images with different scales. (a) Result of SURF. (b) Result of our method.

Table 1: Comparison of time of finding the first 20 matches (ms).

Methods	Mauna Loa	City	Tree branch	Brain3	EO-IR-2	Bay	Streets	EO-IR-1
Symmetric-SIFT	14990	3721	7819	6697	18243	10641	2271	2834
Multispectral detector	17603	4002	8427	7456	22306	10771	2538	2907
RIFT	20191	6750	10215	9850	25508	13201	4265	5214
Our method	3003	621	792	847	2011	2741	426	843

Table 2: The matching rates of our proposed method and the simplified method.

Specimen	Type of the image pair	Matching rates	
		The simplified method (%)	Our proposed method (%)
Brain3	MRI T1/T2	1.1	12.2
Streets	Digital camera/webcam with removed IR filter	1.7	18.4
Mauna Loa	Thermal/short-wave IR	0.14	2.7
Bay	Optical/SAR	0.63	9.17
City	Multispectrum/visible spectrum	4.6	17.6

noise, Poisson noise, pepper-salt noise, and speckle noise in which default values of MATLAB are adopted as the noise parameters; (4) 6 different brightness images, changing the luminance parameters from 0.4 to 1.6. The Correct@20 values are used to evaluate the experimental results, as shown in Figure 5. In the blurred and noisy images, our method can maintain good stability and is better than the other three algorithms. The proposed method can provide more reliable matching for the subsequent calculation of the transformation matrix. In the rotated images and different luminance images, although our method slightly decreases with the increase of rotation angle and luminance parameter, it still performs well and provides enough correct number of matches.

By using the integral image and optimized box filters, the proposed method can effectively weaken the influence of poor-quality images and perform better than the other methods. In the rotated images, the rotation leads to the change of Haar wavelet responses in the horizontal and vertical directions. And our method relies on the Haar wavelet to calculate the saliency indexes of keypoints. This results in some of the same keypoints being eliminated in the first step and makes the performance of this method decline sharply with the increase in the rotation angle. Utilizing edge information indeed makes our method perform well. However, with the change of brightness, the edge contrast becomes smaller, which leads to the gradual decline of the performance of the proposed method.

Through the above experiments, the effectiveness of the proposed method in the registration of multimodal images is verified. However, this method is not suitable for all cases. With the spread of COVID-19, thermal infrared images are often applied to real-time detection of human body temperature. We also register thermal infrared images and visible images, but the effect is not ideal. As shown in Figure 6, 601 keypoints are extracted in the visible light image and are mostly distributed in the scene with complex structures such as the window, lower end of the detector, and poster text. A total of 369 keypoints are extracted in the infrared image and are mostly distributed in the scene where the temperature changes, such as the detector top and human. Among the keypoints extracted from the two images, there are only 20 pairs of points in the same position. Different imaging mechanisms cause that the structure is presented in different regions. This affects the performance of the SURF detector and leads to the failure of the proposed

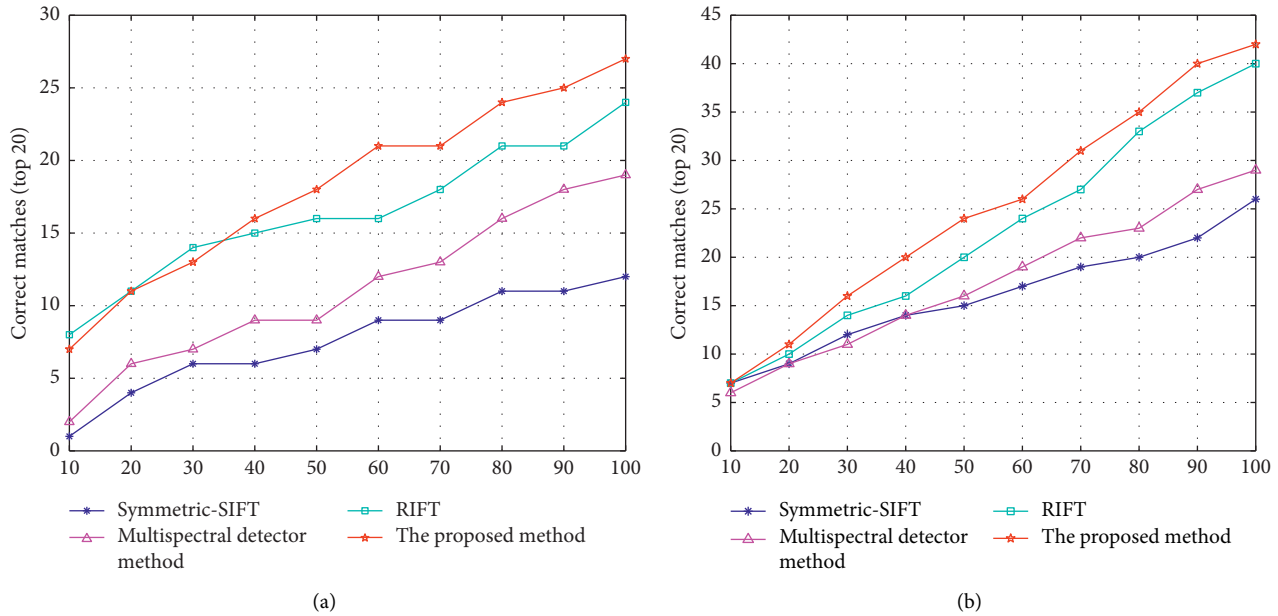


Figure 4: Correct@N curves of four methods in which N varies from 10 to 100. (a) EO-IR-2, a pair of electron optics and IR images. (b) Brain3, a pair of MRI T1 and T2 images.

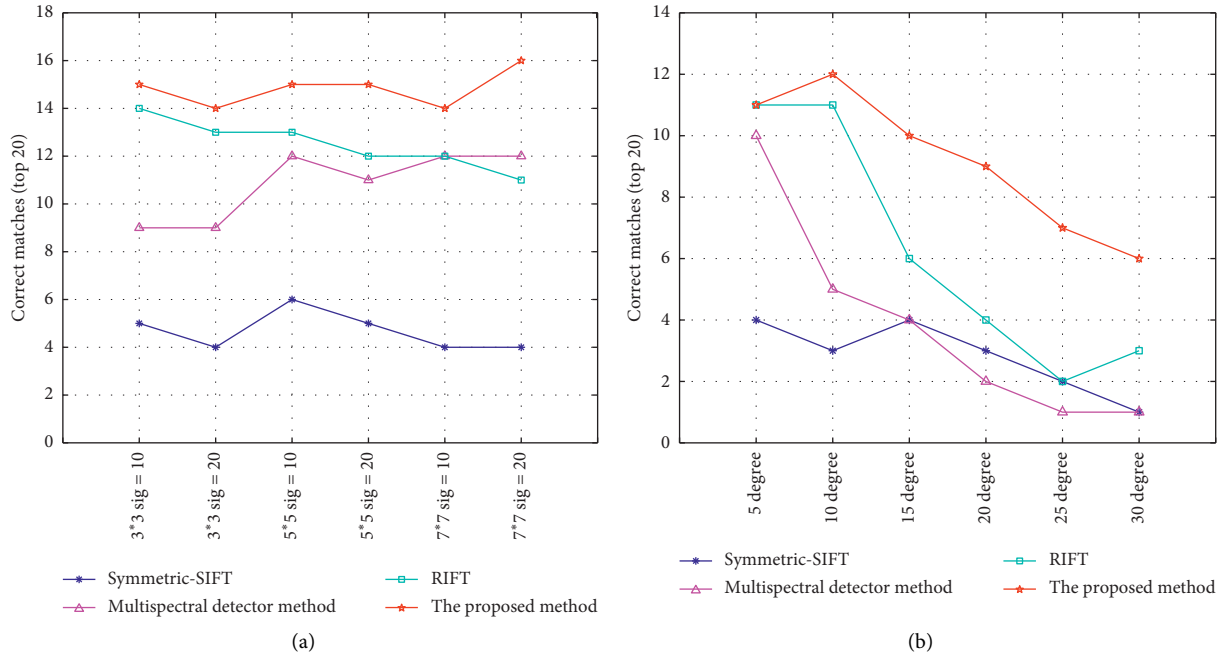


Figure 5: Continued.

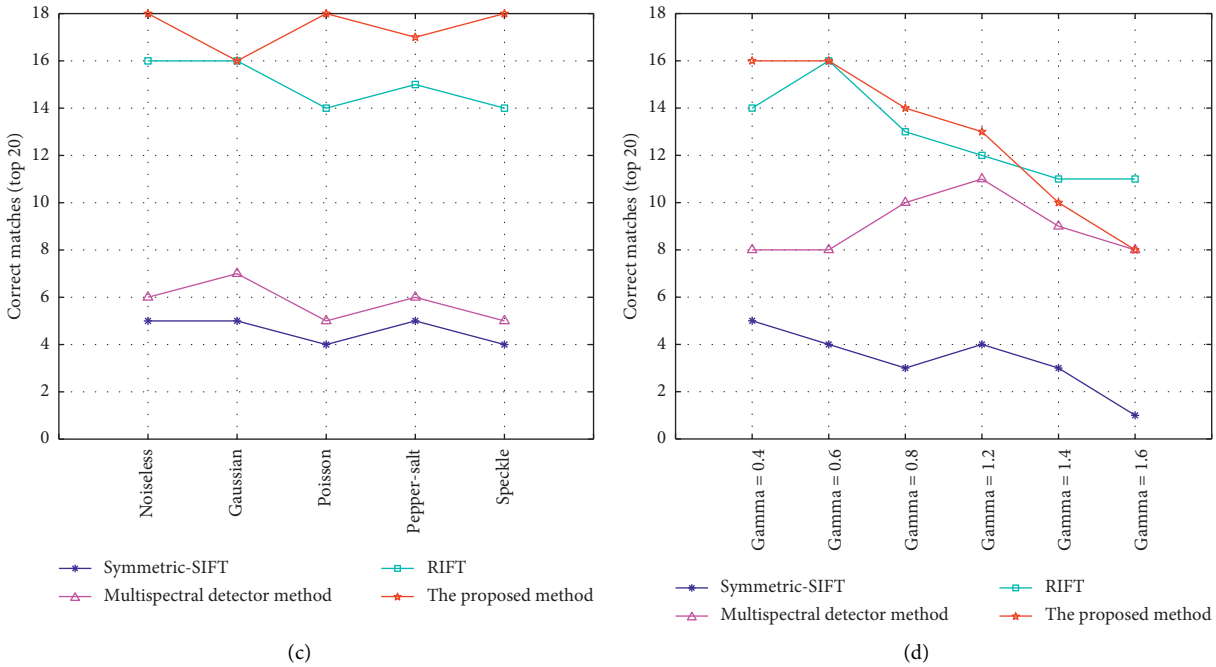


Figure 5: Correct@20 values of four methods for (a) image blurring, (b) rotation, (c) noise, and (d) brightness variations of the multimodal tree branch images.

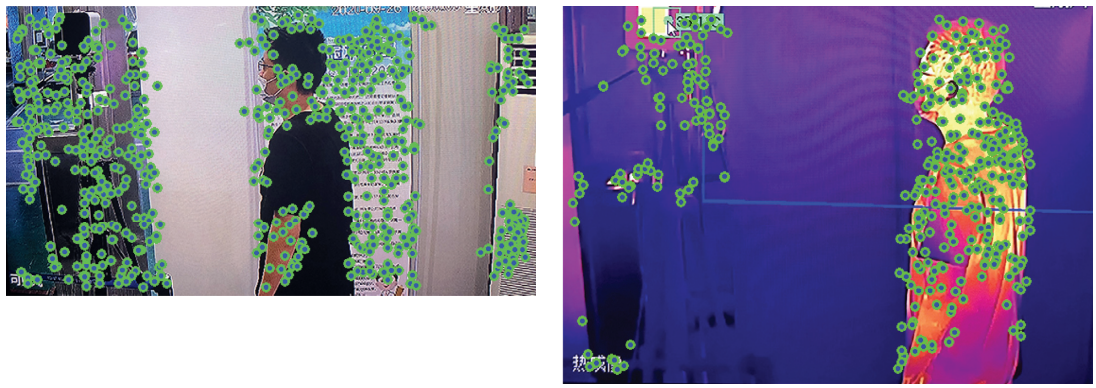


Figure 6: Distribution of keypoints in (a) visible and (b) thermal infrared images.

method. Therefore, our method is suitable for the registration of multimodal images in which enough shared keypoints can be extracted.

4. Conclusion

This paper introduces a rapid and robust method by exploiting local edge information to register multimodal images. Inheriting the merits of SURF and utilizing the stable edge information, the proposed method consumes less runtime, simultaneously maintains high accuracy, and shows better robustness. The experiment results on multimodal image registration indicate that the proposed method has good performance in both time and accuracy, especially in the case of image blurring and noise. However, if the structural information of the multimodal images is greatly different and few shared keypoints can be extracted, the proposed method will not achieve the prospective result. Therefore, future work will attempt a robust method to improve the performance of keypoint detection.

Data Availability

The data used to support the findings of this study are available upon request to the author.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported in part by the Fundamental Research Funds of Guangdong Province (project no. 2019B1515120060).

References

[1] G. Ni and Q. Liu, "Analysis and prospect of multi-source image registration techniques," *Opto-Electronic Engineering*, vol. 31, no. 9, pp. 1–6, 2004.

[2] L. Liu and Y. Jiang, *Dual-Mode Homing Guidance Technology*, Publisher of PLA, Beijing, China, 2003.

[3] J. Tian, S. Bao, and M. Zhou, *Medical Image Processing and Analysis*, Publishing House of Electronics Industry, Beijing, China, 2003.

[4] C. Zhao, H. Zhao, J. Lv, S. Sun, and B. Li, "Multimodal image matching based on multimodality robust line segment descriptor," *Neurocomputing*, C, vol. 177, pp. 290–303, 2016.

[5] Y. Ye, L. Bruzzone, J. Shan et al., "Fast and robust matching for multimodal remote sensing image registration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 99, pp. 9059–9070, 2019.

[6] J. Chen and J. Tian, "Real-time multi-modal rigid registration based on a novel symmetric-SIFT descriptor," *Progress in Natural Science*, vol. 19, no. 5, pp. 643–651, 2009.

[7] J. Han, E. J. Pauwels, and P. De Zeeuw, "Visible and infrared image registration in man-made environments employing hybrid visual features," *Pattern Recognition Letters*, vol. 34, no. 1, pp. 42–51, 2013.

[8] D. Firmenichy, M. Brown, and S. Susstrunk, "Multispectral interest points for RGB-NIR image registration," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP)*, pp. 181–184, Brussels, Belgium, October 2011.

[9] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2941–2958, 2017.

[10] Y. Ye, J. Shan, S. Hao, L. Bruzzone, and Y. Qin, "A local phase based invariant feature for remote sensing image matching," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 142, no. 8, pp. 205–221, 2018.

[11] J. Li, Q. Hu, and M. Ai, "RIFT: multi-modal image matching based on radiation-invariant feature transform," 2018, https://arxiv.org/abs/1804.09493.

[12] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: speeded up robust features," *Computer Vision-ECCV 2006*, vol. 3951, pp. 404–417, 2006.

[13] A. Kelman, M. Sofka, and C. V. Stewart, "Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–7, Minneapolis, MN, USA, June 2007.

[14] S. J. Park, D. K. Park, and C. S. Won, *Core Experiments on MPEG-7 Edge Histogram Descriptor*, MPEG document M5984, Geneva, Switzerland, 2000.

[15] X. Hu, J. Shen, J. Shan, and L. Pan, "Local edge distributions for detection of salient structure textures and objects," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 466–470, 2013.

[16] D. Zhao, Y. Yang, Z. Ji et al., "Rapid Multimodality Registration Based on MM-SURF," *Neurocomputing*, vol. 131, no. 5, pp. 87–97, 2013.

[17] J. H. Jae Hak Lee, Y. S. Yong Sun Kim, D. Dong-Goo Kang, and fnm Jong Beom Ra, "Robust CCD and IR image registration using gradient-based statistical information," *IEEE Signal Processing Letters*, vol. 17, no. 4, pp. 347–350, 2010.

[18] Y. Keller and A. Averbuch, "Multisensor image registration via implicit similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 794–801, 2006.

[19] G. Yang, C. V. Stewart, M. Sofka, and C.-L. Tsai, "Alignment of challenging image pairs: Refinement and region growing starting from a single keypoint correspondence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1973–1989, 2007.

[20] D. Zhao, Q. Wang, H. Sun, and X. Hu, "A method of acceleration applied in symmetric-SIFT and SIFT," *Lecture Notes in Electrical Engineering*, vol. 64, pp. 575–582, 2013.

[21] L. Juan and O. Gwun, "A comparison of SIFT, PCA-SIFT and SURF," *International Journal of Image Processing (IJIP)*, vol. 3, no. 4, pp. 143–152, 2008.